NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

Geometry of the Relativistic Structure of Space-time

Jean - Marie Claudius

A Thesis

in

The Department

of

Mathematics

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Science at
Concordia University
Montréal, Québec, Canada

June 1988

ISBN 0-315-44814-8

ABSTRACT

Geometry of the Relativistic Structure of Space-time

Jean - Marie Claudius

The motivation to structure space-time is examined through the nature of relativity. Differential geometry is then used to give a modern (fairly recent) treatment on the manifold structure of space-time in the absence of gravity. The axioms of physics which underline this development are shown to endow a geometry to space-time. The formulation of physical laws is obtained by assigning mathematical objects of the manifold to physical entities. Gravity is then included as a necessity to determine the metric. Einstein's field equation is then introduced first by motivating its form and then by showing that in weak gravitational fields it leads to Newtonian gravitation. Finally, a brief expository account of the evidence for the possible existence of black holes is given and attempts which have been made to explain the singularities of the theoretical formulas in terms of the theory of black holes are presented: this leads to a mathematical formulation based on the above structure. Important theorems are stated; both ends are areas of current research.

DEDICATION

This thesis is dedicated to the Infinitely Fascinating, without whose motivation this thesis would not have been written.

TABLE OF CONTENTS

INTRODUCTION

Since Einstein presented his General Theory of Relativity, there has been a considerable amount of research on the field of differential geometry and the structure of space-time, leading to a wide array of different areas of studies, from pure mathematics to physics (see (Torretti,1983) for example). The intent of this thesis is to analyse the physical structure of space-time from a mathematical point of view, i.e. by using tools developed in differential geometry; the goal is not to present new results in this latter field nor to study space-time physics. Starting with the underlying principles of relativity, space-time will be endowed with a manifold structure: though all the concepts involved are known, the presentation given in this work is not often met.

In the first chapter, a brief presentation of my point of view on the problem of obtaining temporal and spatial measurements is given. In chapter II, the notions of manifold, tensors, derivative operators, metric, linear connection and integral are presented together with some theorems judged important. In chapter III, the geometry of the manifold is characterized by the invariance of some part of the Poincaré group and the C^{∞} structure of the manifold is given through Lorentz charts. Since these charts capture the essence of the principles underlying relativity, a basis determined by Lorentz charts will be used to specify tensors, allowing us to express physical laws by tensor relations (some example are given). Then, non flat metrics are introduced to account for gravity, leading to a geodesic law of motion which depends on the gravitational field strengths. Einstein field equation is presented together with one particular solution: the weak field solution. In the last section of chapter III, evidence for the possible existence of black holes is given and singularities (breakdowns) are defined theoretically; mathematical definitions are given in order to state Hawking and Penrose's theorems: these determine sufficient conditions for singularities to occur.

CHAPTER I

Preliminaries

The physical laws that we include in our comprehensive system are always tested by observations realized by our conscious mind through experience and experiments. Any observation requires that events be ordered in a series, by a correspondence with the natural numbers, using the criterion of earlier and later. For a law to hold, it is necessary that the order of events be constant; as for the time elapsed between two given events, it varies from place to place, from one person to another. Even clocks are not reliable. As we implied, the measure of time is caused by the sensation of some variation in the surroundings, and as the observed object transforms (this could be one's own brain) the impression of time is felt. For example, a clock measures the time elapsed while the wheels of the mechanism turn, and since it is a law that for a constant torque the wheels will turn at a constant rate, the observer will feel an 'objective' measure of time due to the constancy of the applied torque. But this also requires the constancy of the matter fields, which is not the case; therefore time can never be trusted. In fact, since time brings every object to its end, one could say that time is a mortal disease transmitted through movement and using space as a medium.†

Every different infinitely small portion of space moves differently and is under the influence, through adjacent portions††, of different movements (or matter fields) so that the measure of time is different for every and in every different body, in every portion of this universe, as small portion as one wishes to imagine. For example, consider a pendulum on the the surface of the earth; as the earth voyages in space, two beats will never occur at the same place or should we say two beats will never occur in the same 'bath of waves', and so they will never have the same duration. Thus no law (ex. conservation of energy) can be included in a comprehensive system if there does not exist a mathematical way of transforming time and length measures from different positions in space. This is precisely one of the goals achieved by Einstein's theory (though only locally). In this way we will be able to express the laws of nature independently of the system of coordinates. We thus foresee

---

† in this chapter motion is 'motus' i.e. any kind of quantitative or qualitative change.

†† H.Poincaré defines a continuum as the set of bodies which can be joined by adjacent elements of this set without leaving the set. (Poincaré,1958).

the close relation between space and time.

To this date there are four known forces: those of strong and weak nuclei bonds, electromagnetism and gravity.† The first two have a very small radius of action while the last two both admit small and large radii of interaction. Some extreme examples are: the ratio of the radius of the electron to its gravitational field is of the order of 10^{20} to 10^{40}; as for electromagnetic fields, they play an important role in the formation of a galaxy since their strength and directions are directly related to the velocities of electrons travelling through neighbouring galactic clouds. This suggests that space is permeated by force fields (the last two being of special interest to us) which transfer energy to the encountered matter: these interchanges occur since the fields 'transport mass', and so are they called matter fields; the well known result, due to Einstein, that radiation conveys inertia from the emitting to the receiving body ($m = E/c^2$) is just an example of such exchanges (Einsein, Lorentz, ...,1924). Since these directly cause perturbations to the affected space portions, each event of a series of events will be perturbed and so will the time duration of the series. One consequence of this fact is that matter fields determine the rate of clocks in a one to one correspondence with space. Such a map

---

† mechanical forces are only effects of these four.

has only been found locally (i.e. only in the vicinity of a single source through the assignment of a metric) because of the difficulty to determine the effects of the numerous fields (n-body problem).

Thus on one hand humans subjectively feel time by observing movements in their environment, or in their brain, and on the other hand these movements are adjusted by the ever changing matter fields. This establishes many parallels between the large scale structure of space-time and our mental structure. For example, the human brain coordinates its reasoning by using a four dimensional continuum (space & time) necessary to define causality. In this continuum, the measures of length and time are purely dependent on the frame of reference; this makes us as well as space and time 'at the mercy' of the cosmos and its matter fields.† Hence we can state in a single sentence: all time is psychologic (the logic of the psyche) time, there are as many measures of time (and length) as there are positions in space, each position being permeated by a different flow. We shall call this the 'space-time principle'. The theory of Relativity can be seen

† the Russians are well advanced in the study of the influence of cosmos on life: they have, for example, closely related animal behaviour to solar activity and to the sun's revolution around the centre of the galaxy (for example see (L.Golovanov, 1981) or (G.Touchinski,1966)).

as a successful attempt to relate the different measures taken at different positions, and thus, as an attempt to overcome the fact that all time is psychologic as stated in the space-time principle; more particularly, Einstein's field equation can be seen as a mathematical bridge between the fields emitted by matter and the geometry of the Universe (a geometry based on light (Borel,1960)). To achieve this goal, and hence to structure space-time, we introduce, in the next chapter, some useful mathematical tools.

CHAPTER II

Some Differential Geometry

The following definitions and theorems are standard differential geometric notions; some possible references are: (Hawking & Ellis, 1973), (Spivak, 1965 and 1970), (Torretti, 1983) and (Wald, 1984); as for topology one can refer to (Fairchild & Tulcea, 1971).

We could define an n-dimensional manifold $\mathcal{M}$ to be a topological space for which every point has a neighbourhood homeomorphic to $\mathbb{R}^n$ ($\mathbb{R}$ will stand for the real numbers), but this definition, while describing the structure of a manifold, is not as explicit as the precise formulation:

Definition 2.1: an n-dimensional manifold $\mathcal{M}$ is a set M together with a C^k- atlas $\{\mathcal{O}_\alpha, \phi_\alpha\}$, i.e. a collection of charts $(\mathcal{O}_\alpha, \phi_\alpha)$ where $\mathcal{O}_\alpha$ are subset of M and ϕ_α are one to one correspondences from $\mathcal{O}_\alpha$ to open sets $\phi_\alpha(\mathcal{O}_\alpha)$ in $\mathbb{R}^n$, such that

a) $\forall p \in \mathcal{M}\ \exists \mathcal{O}_\alpha$ such that $p \in \mathcal{O}_\alpha$, i.e. $\{\mathcal{O}_\alpha\}$ cover $\mathcal{M}$

b) if $\mathcal{O}_\alpha \cap \mathcal{O}_\beta \neq \emptyset$ then

$$\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(\mathcal{O}_\alpha \cap \mathcal{O}_\beta) \rightarrow \phi_\beta(\mathcal{O}_\alpha \cap \mathcal{O}_\beta)$$

is a C^k map from an open subset of $\mathbb{R}^n$ to an open subset of $\mathbb{R}^n$.

The C^k- atlas $\{\mathcal{U}_\alpha, \psi_\alpha\}$ is said to be compatible with the C^k- atlas $\{\mathcal{O}_\alpha, \phi_\alpha\}$ if their intersection is also a C^k- atlas. The atlas made of all compatible atlases is called the complete atlas of a given manifold. If the open sets covering $\mathcal{M}$ are unions of open sets belonging to the complete atlas, we get the topology of $\mathcal{M}$, making each map ϕ_α into a homeomorphism. We likewise recover the first, more intuitive, definition of a manifold.

Let $1/2\mathbb{R}^n$ denote the region of $\mathbb{R}^n$ for which $x^1 \leq 0$; the boundary of $\mathcal{M}$ is

$$\partial\mathcal{M} = \{ p \in \mathcal{M} \mid \phi_\alpha(p) \in \text{boundary of } 1/2\mathbb{R}^n \} .$$

We could have defined a manifold with boundary by replacing $\mathbb{R}^n$ by $1/2\mathbb{R}^n$. Note that $\partial\mathcal{M}$ is an $(n-1)$-dim C^k manifold and $\partial(\partial\mathcal{M}) = \emptyset$.

Definition 2.2: a topological space $\mathcal{M}$ is said to be connected if for any non-empty open subsets A,B of $\mathcal{M}$ such that $A \cup B = \mathcal{M}$ then

$$A \cap B \neq \emptyset.$$

Definition 2.3 a) a function f on a C^k manifold is a map from

$\mathcal{M}$ to $\mathbb{R}$. If $f \circ \phi_\alpha^{-1}$, on any corresponding O_α, is a C^r map ($r \le k$) from $\mathbb{R}^n \to \mathbb{R}$ then f is said to be a C^r function.

b) a C^k curve $\lambda(t)$ is a C^k map from $\mathbb{R}$ (or an interval of $\mathbb{R}$) into $\mathcal{M}$; $t \in \mathbb{R}$ is called the parameter.

<u>Definition 2.4</u>: a vector $v = (\frac{\partial}{\partial t})_\lambda \Big|_{t_o}$ tangent to the C^1 curve $\lambda(t)$ at the point $p = \lambda(t_o)$ on $\mathcal{M}$ is the linear map

$$(\frac{\partial f}{\partial t})_\lambda \Big|_t = \lim_{h \to 0} \frac{1}{h} \{ f[\lambda(t+h)] - f[\lambda(t)] \},$$

from the set of functions at $\lambda(t_o)$ to the number $(\frac{\partial f}{\partial t})_\lambda \Big|_{t_o}$ in $\mathbb{R}$.

Note that 1. v can also be defined as $\frac{d(f \cdot \lambda)}{dt}\Big|_{t=t_o}$

2. For all functions f, g and $a, b \in \mathbb{R}$

$$v(af + bg) = a\,v(f) + b\,v(g)$$

and $$v(fg) = f(p)\,v(g) + g(p)\,v(f)$$

3. If f is constant (say $f(p) = c$, $\forall p \in \mathcal{M}$) then

$$\begin{aligned} c\,v(f) &= v(cf) = v(f^2) \\ &= f(p)\,v(f) + v(f)\,f(p) \\ &= 2c\,v(f) \end{aligned}$$

$$\Rightarrow v(f) = 0.$$

As defined in 2.4 the collection {v} at some point p does not form a vector space. Nevertheless, we can define a tangent vector in a more abstract way as the derivative of an equivalence class of curves through p; two curves being equivalent if they have the same derivatives in one, and hence all, coordinate systems at p.

More precisely, λ is said to be equivalent to λ' if

$$\frac{d(x^i \cdot \lambda)}{dt}\bigg|_{t=t_o} = \frac{d(x^i \cdot \lambda')}{dt}\bigg|_{t=t_o}$$

for some coordinate system $(x^1, \ldots, x^n)$ about p. This equivalence relation yields a classification of the curves at p. The first derivative of an equivalence class can be seen as a tangent vector of $\underset{\sim}{\mathcal{M}}$ at p: let V_p denote the collection of all tangent vectors at p. Except for the first part of the proof of Theorem 2.5, we will always think of V_p as a collection of representatives of classes.

<u>Theorem 2.5</u>: let $(x^1, \ldots, x^n)$ be the local coordinates in a neighbourhood of p. Then V_p is an n-dim vector space spanned by $(\partial/\partial x^1)|_p, \ldots, (\partial/\partial x^n)|_p$, the coordinate derivatives.

proof: first we show that V_p forms a vector space. Let $\lambda^j(t)$ run through all the curves of an equivalence class λ. Then for each k the mapping $\varphi^j(t) = \lambda^j(kt)$ is a curve and for a fixed k we obtain all the mappings of an equivalence class φ. Furthermore, for any coordinate system $(x^1,\ldots,x^n)$ we have

$$\frac{d(x^i \cdot \varphi^j)}{dt} = k\,\frac{d(x^i \cdot \lambda^j)}{dt}$$

Let $\lambda_1{}^j(t)$ & $\lambda_2{}^j(t)$ be curves in λ and $\varphi_1{}^j(t)$ & $\varphi_2{}^j(t)$ in φ. Then

$$\frac{d(x^i \cdot \lambda_1^j)}{dt}\Big|_{t=t_o} = \frac{d(x^i \cdot \lambda_2^j)}{dt}\Big|_{t=t_o}$$

and similarly for $\varphi_1{}^j(t)$ & $\varphi_2{}^j(t)$. It follows that

$$\frac{d}{dt}\left[x^i(\lambda_1^j + \varphi_1^j)\right]\Big|_{t=t_o} = \frac{d}{dt}\left[x^i(\lambda_2^j + \varphi_2^j)\right]\Big|_{t=t_o}$$

Hence as λ^j & φ^j vary in there respective classes λ and φ, the sum $\lambda^j + \varphi^j$ runs through a class we denote by $\lambda + \varphi$. We define the corresponding tangent vectors to be the sum of the tangent vectors corresponding to λ and φ. With this addition and multiplication on classes we get the desired structure of a vector space.

From here onwards we consider, as said above, representatives of classes, and thus we see V_p as a collection of maps from the space of functions into the real numbers: it will be understood that V_p only forms a vector space when seen as a collection of derivative of classes.

Now to show that this space is n-dimensional:

$$\left(\frac{\partial f}{\partial t}\right)_\lambda \Big|_{t_o} = \sum_{i=1}^{n} \frac{d}{dt} x^i(\lambda(t))\Big|_{t=t_o} \cdot \frac{\partial f}{\partial x^i}\Big|_{\lambda(t_o)}$$

$$= \sum_{i=1}^{n} \frac{dx^i}{dt} \frac{\partial f}{\partial x^i}\Big|_{\lambda(t_o)} .$$

Thus every vector at p can be expressed as the linear combination of the coordinate derivatives.

Conversely, consider the linear combination

$$\sum_{i=1}^{n} a^i \left(\frac{\partial}{\partial x^i}\right)\Big|_p \quad , \quad a^i \in \mathbb{R} .$$

and let $\lambda(t)$ be such that $x^i(\lambda(t)) = x^i(p) + t\,a^i$. Then the tangent vector to this curve at p is

$$\sum_{i=1}^{n} a^i \left(\frac{\partial}{\partial x^i}\right)\Big|_p .$$

We must show that the $(\partial/\partial x^i)\big|_p$ are independent: suppose that they are not, then there exist $a^i \in \mathbb{R}$, $i = 1,\ldots,n$, such that the latter sum is zero with $a^i \neq 0$ for at least one i; applying this operator (the sum) to the each coordinate x^j we get

$$\sum_{i=1}^{n} a^i \left(\frac{\partial x^j}{\partial x^i}\right) = a^j = 0 ,$$

a contradiction. Hence V_p is an n-dim vector space, under usual vector addition and scalar multiplication. •

V_p is called the tangent vector space to $\mathcal{M}$ at p and $(\partial/\partial x^1)|_p,\ldots,(\partial/\partial x^n)|_p$ the coordinate basis. This basis clearly depends on the chart (O_α , ϕ_α) ; suppose that we create a new basis by using the chart (U_α , ψ_α) , then the old basis could be expressed in terms of the new basis by (using the chain rule)

$$(\partial/\partial x^i)|_p = \sum_{j=1}^{n} \frac{\partial(\psi_\alpha \circ \phi_\alpha^{-1})^j}{\partial x^i}\Bigg|_{\phi_\alpha(p)} \frac{\partial}{\partial(\psi_\alpha \circ \phi_\alpha^{-1})^j}$$

Thus the components v'^j of v in the new basis would be

$$v'^j = \sum_{i=1}^{n} v^i \frac{\partial x'^j}{\partial x^i}$$

where x'^j is the j-th component of the map $\psi_\alpha \circ \phi_\alpha^{-1}$. This is known as the vector transformation law.

Definition 2.6: a tangent (vector) field, on a manifold $\mathcal{M}$, is the assignment of a tangent vector $v \in V_p$ at each point $p \in \mathcal{M}$.

Definition 2.7: V_p^*, the dual vector space to V_p, is the collection of linear maps $\omega: V_p \to \mathbb{R}$ at p; each ω is called a one-form.

If $\{e_i\}$ form a basis for V_p then the elements $\{e^j\} \in V_p^*$ for which $e^j(e_i) = \delta^j{}_i$ form a basis for V_p^* ($\delta^j{}_i = 1$ if i=j and 0 otherwise). This correspondence between $v^j \in V_p^*$ and $v_i \in V_p$ makes V_p^* isomorphic to V_p. The vectors of V_p are often called contravariant vectors and the one-forms of V_p^* called covariant vectors.

Each function f on $\mathcal{M}$ defines a dual vector df by the following rule: for any $v \in V_p$ let $df(v) = X^i f$, where X^i is the number to which v is mapped by a dual basis vector e^i (i.e. $X^i = e^i(v)$); df is called the differential of f. If

$(x^1, \ldots, x^n)$ are local coordinates, then the differentials $(dx^1, \ldots, dx^n)$ at p constitute the basis dual to the vector basis $(\partial/\partial x^1, \ldots, \partial/\partial x^n)$ at p, since

$$dx^i(\partial/\partial x^j) = \frac{\partial}{\partial x^j} x^i = \delta^i_j .$$

Thus $df = \sum_i \frac{\partial f}{\partial x^i} dx^i$, in terms of the coordinate basis.

<u>Definition 2.8</u>: a tensor of type (r,s) at p over V_p is a multilinear map

$$T : \underbrace{V_p^* \times \ldots \times V_p^*}_{r \text{ times}} \times \underbrace{V_p \times \ldots \times V_p}_{s \text{ times}} \to \mathbb{R} .$$

Thus a tensor of type (0,1) is a one-form, and a tensor of type (1,0) is an element of V_p^{**} and hence of V_p. An arbitrary tensor T is expressed in terms of any basis $\{e_a\}$ and $\{e^b\}$ for V_p^* and V_p (respectively) as

$$T = \sum_{a_1 \ldots b_s} T^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} \, e_{a_1} \otimes \cdots \otimes e_{a_r} \otimes e^{b_1} \otimes \cdots \otimes e^{b_s}$$

where $\{ e_{a_1} \otimes \cdots \otimes e^{b_s} \}$ is a basis for the space of all tensors of type (r,s) at p denoted by $\mathfrak{J}(r,s)$ and where $T^{a\ldots}{}_{b\ldots}$ are the components of the tensor T with respect to the basis $\{e_a\}$ and $\{e^b\}$. Relations in the tensor space $\mathfrak{J}(r,s)$ are usually expressed in terms of the tensor components; thus if T, T' $\in \mathfrak{J}(r,s)$ and $\alpha \in \mathbb{R}$ then

$$(\alpha T + T')^{a_1\ldots a_r}{}_{b_1\ldots b_s} = (\alpha T)^{a_1\ldots a_r}{}_{b_1\ldots b_s} + T'^{a_1\ldots a_r}{}_{b_1\ldots b_s}$$

$$(T \otimes T')^{a_1\ldots a_{r+r'}}{}_{b_1\ldots b_{s+s'}} = T^{a_1\ldots a_r}{}_{b_1\ldots b_s} T'^{a_{r+1}\ldots a_{r+r'}}{}_{b_{s+1}\ldots b_{s+s'}}$$

Another important operation on tensors is contraction:

<u>Definition 2.9</u>: contraction with respect to the i-th (dual vector) and j-th (vector) position is a map C: $\mathfrak{J}(r,s) \rightarrow \mathfrak{J}(r-1,s-1)$ such that

$$C(T) = \sum_{\sigma=1}^{n} T^{a_1\ldots\sigma\ldots a_r}{}_{b_1\ldots\sigma\ldots b_s}$$

where σ is at the i-th and j-th position in $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ respectively. It is a convention not to write the summation term. For example $T^{ab} T_{ac} = \Sigma_a T^{ab} T_{ac} = T'^{b}{}_{c}$.

As in the case of vectors, writing a tensor using components depends on the choice of basis $\{e^a\}$ and $\{e_b\}$ and thus on the different charts. As we had a rule for transforming vectors, the following gives the components of a tensor in the new basis $\{e^{a'}\}$ and $\{e_{b'}\}$ with respect to the old basis

$$T'^{a_1'\ldots a_r'}{}_{b_1'\ldots b_s'} = \sum_{a_1\ldots b_s} T^{a_1\ldots a_r}{}_{b_1\ldots b_s} \, E^{a_1'}{}_{a_1} \cdots E_{b_s'}{}^{b_s}$$

where $E^{a'}{}_a$ and $E_{b'}{}^b$ are $n \times n$ non-singular matrices such that $e^{a'} = E^{a'}{}_a e^a$ and $e_{b'} = E_{b'}{}^b e_b$ ($E_{b'}{}^b E^{b'}{}_a = \delta^b{}_a$).

The symmetric part of a (0,s)-tensor (similarly of a (r,0)-tensor) is defined to be

$$T_{(b_1\ldots b_s)} = \frac{1}{s!} \sum_{\pi} T_{b_1\ldots b_s}$$

and the antisymmetric part

$$T_{[b_1\ldots b_s]} = \frac{1}{s!} \sum_{\pi} \mathrm{sign}(\pi) \cdot T_{b_1\ldots b_s}$$

where π represents all possible permutations of $1,\ldots,s$.

Definition 2.10: a totally antisymmetric tensor field T of type (0,s) is a tensor for which

$$T_{b_1\ldots b_s} = T_{[b_1\ldots b_s]}$$

and is called a differential s-form. (As a vector field, a tensor field is obtained by assigning a tensor at every point of a manifold).

Definition 2.11: a metric $\mathbf{g}$, on a manifold $\mathcal{M}$, is a symmetric, non-degenerate tensor field of type (0,2).

The components of $\mathbf{g}$ with respect to a basis $\{e_a\}$ are

$$g_{ab} = \mathbf{g}(e_a, e_b).$$

If the coordinate basis $\{\partial/\partial x^a\}$ is used then

$$\mathbf{g} = g_{ab}\, dx^a \otimes dx^b.$$

Note that 1. Since $\mathbf{g}$ is symmetric then $g_{ab} = (g_{ab} + g_{ba})/2$

$$\Rightarrow \quad g_{ab} = g_{ba} .$$

2. The 'magnitude' of $v \in V_p$ at p is given by

$$|\mathbf{g}(v,v)|^{1/2} .$$

3. The 'cos angle' between u & $v \in V_p$ is

$$\frac{g(u,v)}{|g(u,u)\cdot g(v,v)|^{1/2}} \quad \text{where} \quad g(u,u)\cdot g(v,v) \neq 0 \ .$$

and where u and v are orthogonal if $g(u,v) = 0$.

4. Since (0,2)-tensors can be represented by $n \times n$ matrices,

$$g = \begin{bmatrix} \ldots, dx^a, \ldots \end{bmatrix} \begin{bmatrix} g_{11} & \cdots & \frac{1}{2}g_{1b} & \cdots & \frac{1}{2}g_{1n} \\ \vdots & & \vdots & & \vdots \\ \frac{1}{2}g_{a1} & \cdots & g_{ab} & & \vdots \\ \vdots & & & & \vdots \\ \frac{1}{2}g_{n1} & \cdots & \cdots & \cdots & g_{nn} \end{bmatrix} \begin{bmatrix} \vdots \\ \vdots \\ dx^b \\ \vdots \\ \vdots \end{bmatrix}$$

Associated with g_{ab} there exists a unique (2,0)-tensor g^{ab} which is defined by the relation

$$g^{ab} g_{bc} = \delta^a_c \ .$$

In the matrix representation $(g^{ab}) = (g_{ab})^{-1}$, and so these two tensors are used as isomorphisms between contravariant and covariant elements of $\mathfrak{I}(r,s)$ (this follows from the non-degeneracy condition). We often say that the metric raises and lowers indices since $v_a = g_{ab}v^b$ and $v^a = g^{ab}v_b$; for tensors we would have for example $T^a{}_b{}^c{}_d{}^e = g_{bf}\, g^{ch}\, g^{ei}\, T^{af}{}_{hdi}$: this association is unique.

<u>Definition 2.12</u>: the signature of $\mathbf{g}$ at p is the number of positive eigenvalues minus the number of negative eigenvalues of the matrix representing $\mathbf{g}$ at a given p. A metric of signature (n-2) for all p is called a Lorentz metric.

Lorentz metrics are used in relativity: they divide non-zero vectors at p into three classes; timelike, spacelike and null vectors whether $\mathbf{g}(v,v)$ is negative, positive or zero respectively. Null vectors form a cone at each p, separating timelike and spacelike vectors. (Events on the cone cannot be reached by a material particle but can be reached by a light signal. Events within the cone can be reached by material particles).

We will study three differential operators on manifolds: the exterior derivative, the Lie derivative and the covariant derivative. Only the latter will require an additional structure on the manifold, the affine connection.

Before we introduce these derivative operators the concepts of imbedding, immersion and tensor maps must be defined:

Definition 2.13: let $\mathcal{M}'$ be a manifold of dimension $n' \geq n$. If the map $\phi: \mathcal{M} \to \mathcal{M}'$ and its inverse are C^k maps $(k \geq 0)$, i.e. if $\forall p \in \mathcal{M}$ there exists an open neighbourhood 0 such that ϕ^{-1}: $\phi(0) \to \mathcal{M}$ is one to one and C^k, then $\phi(\mathcal{M})$ is said to be an n-dim immersed submanifold in $\mathcal{M}'$ and ϕ is called an immersion. If in addition ϕ is onto then $\phi(\mathcal{M})$ is said to be an imbedded submanifold of $\mathcal{M}'$. An imbedded submanifold of dimension (n-1) is called a hypersurface. If $n = n'$ the imbedding is referred to as a diffeomorphism from $\mathcal{M}$ to $\mathcal{M}'$.

The structure of an imbedded submanifold is naturally obtained through ϕ from the manifold itself. This can be seen as such: for any function $f: \mathcal{M}' \to \mathbb{R}$, define $\phi^* f$ on $\mathcal{M}$ to be the function which sends $p \in \mathcal{M}$ to the value of f at $\phi(p)$, i.e.

$$\phi^* f(p) = f(\phi(p)).$$

The tangent vector to the curve $\phi(\lambda(t))$ at $\phi(p)$ can be denoted $\phi_*(\frac{\partial}{\partial t})_\lambda \Big|_{\phi(p)}$, making ϕ_* a linear map from V_p of $\mathcal{M}$ to $V_{\phi(p)}$ of $\mathcal{M}'$. Thus for any $v \in V_p$ and any function f

$$v(\phi^* f)\Big|_p = \phi_* v(f)\Big|_{\phi(p)}.$$

In this way the vector mapping ϕ_* from $\mathcal{M}$ to $\mathcal{M}'$ defines a linear one-form mapping ϕ^* from $V^*_{\phi(p)}$ of $\mathcal{M}'$ to V^*_p of $\mathcal{M}$ by requiring that contractions of one-forms be preserved under the map.

The maps ϕ_* and ϕ^* are extended to maps of contravariant tensors $(\mathcal{M} \to \mathcal{M}')$ and of covariant tensors $(\mathcal{M}' \to \mathcal{M})$ respectively, i.e.

$$\phi_* : T \in \mathfrak{J}(r,0)\big|_p \;\to\; \phi_* T \in \mathfrak{J}(r,0)\big|_{\phi(p)} \quad ,$$

$$\phi^* : T \in \mathfrak{J}(0,s)\big|_{\phi(p)} \;\to\; \phi^* T \in \mathfrak{J}(0,s)\big|_p \quad .$$

Definition 2.14: the exterior derivative d is a map from q-form fields to (q+1)-form fields. If F is a q-form field then

$$dF = dF_{a_1 \ldots a_q}\, dx^{a_1} \wedge \ldots \wedge dx^{a_q} \quad ,$$

where $\wedge$ is the skew-symmetrized tensor product $\otimes$, i.e.

$$dx^{a_1} \wedge \ldots \wedge dx^{a_q} = dx^{[a_1} \otimes \cdots \otimes dx^{a_q]} \quad .$$

Theorem 2.15: the (q+1)-form field dF is independent of the choice of coordinates.

proof: let $\{x^a\}$ and $\{x^{a'}\}$ be two sets of coordinates. Then

$$F = F_{a_1'\dots a_q'}\, dx^{a_1'} \wedge \dots \wedge dx^{a_q'}$$

$$\text{where} \quad F_{a_1'\dots a_q'} = \frac{\partial x^{a_1}}{\partial x^{a_1'}} \cdots \frac{\partial x^{a_q}}{\partial x^{a_q'}} F_{a_1\dots a_q}$$

$$\text{Let} \quad X_i = \frac{\partial x^{a_i}}{\partial x^{a_i'}}, \quad i = 1,\dots,q.$$

$$\Rightarrow \quad dF = d\,(X_1 \cdots X_q\, F_{a_1\dots a_q}) \wedge dx^{a_1'} \wedge \dots \wedge dx^{a_q'}$$

$$= \frac{\partial^2 x^{a_1}}{\partial x^{a_1'}\,\partial x^{b'}}\, X_2 \cdots X_q\, F_{a_1\dots a_q}\, dx^{b'} \wedge dx^{a_1'} \wedge \dots \wedge dx^{a_q'}$$

$$+ \dots + \, X_1 \cdots X_{q-1}\, \frac{\partial^2 x^{a_q}}{\partial x^{a_q'}\,\partial x^{b'}}\, F_{a_1\dots a_q}\, dx^{b'} \wedge dx^{a_1'} \wedge \dots \wedge dx^{a_q'}$$

$$+ \, X_1 \cdots X_q\, dF_{a_1\dots a_q} \wedge dx^{a_1} \wedge \dots \wedge dx^{a_q}$$

but since the terms involving ∂^2 are symmetric in a' & b', and since $dx^{b'} \wedge dx^{a'}$ is skew-symmetric we get

$$dF = dF_{a_1 \ldots a_q} \; X_1 \, dx^{a_1'} \wedge \ldots \wedge X_q \, dx^{a_q'}$$

$$\Rightarrow dF = dF_{a_1 \ldots a_q} \; dx^{a_1} \wedge \ldots \wedge dx^{a_q} \; .$$

Note that 1. Since $d(fg) = g\,df + f\,dg$,

$$d(F \wedge G) = dF \wedge G + (-1)^q F \wedge dG$$

$$\forall \text{ q-form fields } F, G.$$

2. $df = (\partial f/\partial x^i)\, dx^i$

$$\Rightarrow d(df) = \frac{\partial^2 f}{\partial x^i \partial x^j} dx^i \wedge dx^j = 0$$

$\Rightarrow$ $d(df)$ is also zero.

A manifold $\mathcal{M}$ is said to be orientable if there exist an atlas $\{O_\alpha, \phi_\alpha\}$ such that the Jacobian $|\partial x^i/\partial x'^j|$ is positive in every non-empty $O_\alpha \cap O_\beta$; here $\{x^i\}$ and $\{x'^j\}$ are the coordinates in O_α and O_β respeotively.

<u>Definition 2.16</u>: let $\mathcal{M}$ be a compact orientable n-manifold with boundary $\partial\mathcal{M}$ and let $\{f_\alpha\}$ be a partition of unity for a finite oriented atlas $\{O_\alpha, \phi_\alpha\}$. Then for any n-form field F on $\mathcal{M}$, the integral of F over $\mathcal{M}$ is given by

$$\int_{\mathcal{M}} F = \frac{1}{n!} \sum_{\alpha} \int_{\phi_\alpha(\mathcal{O}_\alpha)} f_\alpha F_{1\ldots n} \, dx^1 \cdots dx^n \, ,$$

where $F_{1\ldots n}$ are the components of F with respect to $(x^1, \ldots, x^n)$ in $\mathcal{O}_\alpha$.

The second type of differentiation which does not require an additional structure on the manifold is the Lie derivative.

A one-parameter group of diffeomorphisms ϕ_t is a collection of C^k maps $(k \geq 1)$ from $\mathbb{R} \times \mathcal{M} \to \mathcal{M}$ such that for any $t \in \mathbb{R}$ $\phi_t : \mathcal{M} \to \mathcal{M}$ is a diffeomorphism and $\forall\, s,t \in \mathbb{R}$ $\phi_{t+s} = \phi_t \circ \phi_s = \phi_s \circ \phi_t$. For any vector field $\boldsymbol{v}$ we can generate ϕ_t as follows· by the fundamental theorem of ordinary differential equations the curve $\lambda(t)$ satisfying $dx^i/dt = v^i(x^1(t), \ldots, x^n(t))$ is the unique maximal curve passing through $p \in \mathcal{M}$ such that $\lambda(0) = p$ and such that the tangent at $\lambda(t)$ is $\boldsymbol{v}|_{\lambda(t)}$. This curve is called the integral curve of $\boldsymbol{v}$. We define $\phi_t(p)$ to be the point at distance t from p on the integral curve of $\boldsymbol{v}$. This one-parameter group of diffeomorphisms maps any tensor T at p into $\phi_{*t} T$ at $\phi_t(p)$.

<u>Definition 2.17</u>: the Lie derivative $\mathbf{L}_{\boldsymbol{v}} T$ of a tensor $T \in \mathfrak{I}(r,s)$ at p with respect to a tensor field $\boldsymbol{v}$ is defined to be:

$$\mathbf{L}_\nu T = \lim_{t\to 0} \frac{1}{t} \{ T - \phi_{*t} T \} .$$

It follows immediately from this definition and from the properties of ϕ_{*t} that

1. $\mathbf{L}_\nu : \mathfrak{I}(r,s) \to \mathfrak{I}(r,s)$, linearly.
2. If $S,T \in \mathfrak{I}(r,s)$ then

$$\mathbf{L}_\nu (S \otimes T) = (\mathbf{L}_\nu S) \otimes T + S \otimes (\mathbf{L}_\nu T)$$

3. For any function f, $\mathbf{L}_\nu f = \nu(f)$.

<u>Theorem 2.18</u>: if u and ν are vector fields then

$$(\mathbf{L}_\nu u)^i = \nu^j \frac{\partial u^i}{\partial x^j} - u^j \frac{\partial \nu^i}{\partial x^j} .$$

written as $\mathbf{L}_\nu u = [\nu, u]$ $(= -[u,\nu])$.

<u>proof</u>: let the coordinate system $(x^1, \ldots, x^n)$ be such that $\nu^i = (\partial/\partial x^1)^i$, i.e. such that t is chosen as x^1. Then ϕ_{-t} is just $x^1 \to x^1 + t$ with $x^2, \ldots, x^n$ unchanged. Thus for $T \in \mathfrak{I}(r,s)$

$$(\phi_{*-t} T)(x^1, \ldots, x^n) = T(x^1 + t, x^2, \ldots, x^n)$$

making $\mathbf{L}_\nu T = \partial T/\partial x^1$; $\Rightarrow \mathbf{L}_\nu u^i = \partial u^i/\partial x^1$ for any vector field u.

On the other hand any vector field $\boldsymbol{u}$ has for components

$$u^i = u^j(\partial/\partial x^j)^i$$

thus
$$[\boldsymbol{u},\boldsymbol{v}]^i = v^j \frac{\partial u^i}{\partial x^j} - u^j \frac{\partial v^i}{\partial x^j}$$
$$= \partial u^i/\partial x^1$$

$\Rightarrow$ $\quad L_v \boldsymbol{u} = [\boldsymbol{v},\boldsymbol{u}]$ $\quad$ since both sides are defined independently from the choice of coordinates. •

Note that because $\phi^*(df) = d(\phi^* f)$, $\quad d(L_v F) = L_v(dF)$ for any q-form field F. Also since $L_v(f) = v(f)$, we have

$$L_v(w_i u^i) = v(w_i u^i) \quad \text{for any form field } w;$$

$$\text{but} \quad L_v(w_i u^i) = u^i L_v w_i + w_i L_v u^i$$
$$= u^i L_v w_i + w_i [v,u]^i$$
$$\Rightarrow \quad u^i L_v w_i = v(w_i u^i) + w_i [u,v]^i$$

which determines L_v completely.

<u>Definition 2.19</u>: the covariant derivative ∇_a, on a manifold $\mathcal{M}$ is a map from C^∞ tensor fields of type (r,s) to C^∞ tensor fields of type (r,s+1) which satisfies:

a) linearity,

b) product rule (or Leibnitz rule),

c) commutativity with contraction,

d) tangent vectors are directional derivatives, i.e. for all functions f and $\forall v \in V_p \quad v(f) = v^a \nabla_a f$.

e) the torsion free condition $\nabla_a \nabla_b f = \nabla_b \nabla_a f$.

Lemma 2.20: let ∇_a and ∇'_a be two derivative operators and let w_a be a dual vector field. Then there exists a (1,2)-tensor field, say $D^c{}_{ab}$, such that

$$\nabla'_a w_b - \nabla_a w_b = D^c{}_{ab} w_c .$$

proof: for any scalar field (a function) f we have

$$\nabla'_a (f w_b) - \nabla_a (f w_b) = (\nabla'_a f) w_b + f(\nabla'_a w_b) - (\nabla_a f) w_b - f(\nabla_a w_b)$$
$$= f(\nabla'_a w_b - \nabla_a w_b).$$

Let w_b' be another dual vector field which is equal to w_b at p. Then we can find C^∞ functions $f_{(\alpha)}$ vanishing at p and C^∞ dual vector fields $w_b^{(\alpha)}$ such that

$$w_b' - w_b = \sum_{\alpha=1}^{n} f_{(\alpha)} w_b^{(\alpha)}$$

$$\Rightarrow \nabla'_a (w_b' - w_b) - \nabla_a (w_b' - w_b) = \sum_\alpha \left[\nabla'_a (f_{(\alpha)} w_b^{(\alpha)}) - \nabla_a (f_{(\alpha)} w_b^{(\alpha)}) \right]$$

$$= \sum_\alpha f_{(\alpha)} \left[\nabla'_a w_b^{(\alpha)} - \nabla_a w_b^{(\alpha)} \right] = 0$$

since each $f_{(\alpha)}$ vanishes at p.

$\Rightarrow$ $\nabla'_a w_b' - \nabla_a w_b' = \nabla'_a w_b - \nabla_a w_b$

$\Rightarrow$ $\nabla'_a w_b - \nabla_a w_b$ only depends on the value of w_b at p.

$\Rightarrow$ $\nabla'_a - \nabla_a$ defines a map from one-forms at p to (0,2)-tensors at p. By Definition 2.19 a) this map is linear

$\Rightarrow$ $\nabla'_a - \nabla_a$ is a (1,2)-tensor at p, say $D^c{}_{ab}$. •

Note that if we let $w_b = \nabla_b f = \nabla'_b f$ we find

$$\nabla'_a \nabla'_b f - \nabla_a \nabla_b f = D^c{}_{ab} \nabla_c f$$

$\Rightarrow$ $D^c{}_{ab} = D^c{}_{ba}$ by the torsion free condition.

Theorem 2.21: let ∇'_a and ∇_a be two covariant derivatives and let t^a be a vector field. Then there exists a (1,2)-tensor field, say $D^c{}_{ab}$, such that

$$\nabla_a t^b = \nabla'_a t^b + D^b{}_{ac} t^c .$$

<u>proof</u>: let w_a be a form field, then

$$(\nabla'_a - \nabla_a)(w_b t^b) = [(\nabla'_a - \nabla_a)w_b]t^b + w_b(\nabla'_a - \nabla_a)t^b$$

$$= (D^c{}_{ab}\, w_c)t^b + w_b(\nabla'_a - \nabla_a)t^b \quad \text{by Lemma 2.20}$$

but by Definition 2.19 d) $\quad (\nabla'_a - \nabla_a)(w_b t^b) = 0$

$$\Rightarrow \quad (D^c{}_{ab}\, w_c)t^b + w_b(\nabla'_a - \nabla_a)t^b = 0$$

$$\Rightarrow \quad w_b[(\nabla'_a - \nabla_a)t^b + D^b{}_{ac}\, t^c] = 0 \qquad \forall\, w_b$$

$$\Rightarrow \quad \nabla_a t^b = \nabla'_a t^b + D^b{}_{ac}\, t^c \qquad \bullet$$

This is a very important result since it characterizes the difference between two covariant derivatives by $D^c{}_{ab}$, i.e. by $n^2(n+1)/2$ independent components. Lemma 2.20 together with Theorem 2.21 can be generalized to any tensor $T \in \mathfrak{I}(r,s)$. The application of our interest is $\nabla'_a = \partial/\partial x^a$; then $D^c{}_{ab}$ is denoted $\Gamma^c{}_{ab}$ and is called the Christoffel symbol (or the affine connection). Thus we have $\nabla_a t^b = \partial t^b/\partial x^a + \Gamma^b{}_{ac} t^c$. (For simplicity of notation we sometimes write this particular $\nabla_a t^b$ as $t^b{}_{;a}$ and similarly $\partial t^b/\partial x^a$ is denoted by $\partial_a t^b$ or just $t^b{}_{,a}$).

<u>Definition 2.22</u>: given a curve $\lambda(t)$, its tangent v^a and a covariant derivative ∇_a , we say that the vector u^a is

parallelly transported if $v^a \nabla_a u^b = 0$ is satisfied as u^a moves along the curve.

By using the result of Theorem 2.21 applied to $\partial/\partial x^a$ we get

$$v^a \nabla_a u^b = v^a(\partial u^b/\partial x^a) + v^a \Gamma^b{}_{ac} u^c = 0$$

or in terms of the coordinate basis and the parameter t,

$$du^b/dt + v^a \Gamma^b{}_{ac} u^c = 0 .$$

<u>Definition 2.23</u>: given ∇_a, a geodesic is a curve whose tangent vector v^a is parallelly transported along itself,

$$\text{i.e.} \quad v^a \nabla_a v^b = 0 .$$

Thus the geodesic equation is

$$\frac{dv^a}{dt} + v^b \Gamma^a{}_{bc} v^c = 0$$

or just

$$\frac{d^2x^a}{dt^2} + \Gamma^a{}_{bc} \frac{dx^b}{dt} \frac{dx^c}{dt} = 0 \tag{2.1}$$

To understand the physical significance of (2.1) it is necessary to show that there exists a unique solution. By demanding that for any two vectors v^a and u^a the inner product $g_{ab}v^a u^b$ be constant when parallelly transported along any curve, it will be possible to show the uniqueness of $\Gamma^c{}_{ab}$ (or $D^c{}_{ab}$). This will be sufficient since Differential Equations garanties a unique solution to (2.1) when given a particular $\Gamma^c{}_{ab}$. As a bonus we will get a characterization of $\Gamma^c{}_{ab}$ (or $D^c{}_{ab}$) in terms of the metric $\mathbf{g}$ and the derivative operator $\partial/\partial x^a$ (or ∇'_a).

<u>Theorem 2.24</u>: let $\mathbf{g}$ be a metric. Then there exists a unique covariant derivative ∇_a satisfying $\nabla_a g_{bc} = 0$.

<u>proof</u>: since we want the inner product of any two vectors to remain unchanged when parallelly transported, we must have

$$t^a\nabla_a(g_{bc}v^bu^c) = 0 ,$$

as t^a moves along the curve and where v and u are two parallelly transported vectors. This is true if and only if.

$$t^a v^b u^c \nabla_a g_{bc} = 0$$

$$\Leftrightarrow \qquad \nabla_a g_{bc} = 0.$$

Hence $\quad 0 = \nabla_a g_{bc}$

$$= \nabla'_a g_{bc} - D^d{}_{ab} g_{dc} - D^d{}_{ac} g_{bd}$$

$$\Rightarrow \quad D_{cab} + D_{bac} = \nabla'_a g_{bc} \ .$$

By substitution of indices we also get

$$D_{cba} + D_{abc} = \nabla'_b g_{ac} \quad \text{(interchanging a \& b)}$$

$$\text{and} \quad D_{bca} + D_{acb} = \nabla'_c g_{ab} \ , \quad (c \to b, a \to c, b \to a) .$$

By adding the first two equations, subtracting the third, and then by using the symmetry property of $D^c{}_{ab}$ we find

$$2\ D_{cab} = \nabla'_a g_{bc} + \nabla'_b g_{ac} - \nabla'_c g_{ab}$$

$$\Rightarrow \quad D^c{}_{ab} = g^{cd} (\nabla'_a g_{bd} + \nabla'_b g_{ad} - \nabla'_d g_{ab}) \ / \ 2 .$$

Therefore a given derivative operator ∇'_a and a given $D^c{}_{ab}$ (through **g**) produce a unique ∇_a satisfying $\nabla_a g_{bc} = 0$.

Hence g_{ab} determines both the covariant derivative ∇_a and the Christoffel symbol which is now given by

$$\Gamma^a{}_{bc} = \frac{1}{2} g^{ad} \left\{ \frac{\partial g_{cd}}{\partial x^b} + \frac{\partial g_{bd}}{\partial x^c} - \frac{\partial g_{bc}}{\partial x^d} \right\} \qquad (2.2)$$

Consider the following: let w_a be a dual vector and f a scalar field, then

$$\nabla_a\nabla_b(fw_c) = \nabla_a(w_c\nabla_b f + f\nabla_b w_c)$$
$$= w_c\nabla_a\nabla_b f + \nabla_a w_c\nabla_b f + \nabla_a f\nabla_b w_c + f\ \nabla_a\nabla_b w_c$$

and by subtracting the equation obtained by interchanging a & b we get

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)(fw_c) = f(\nabla_a\nabla_b - \nabla_b\nabla_a)w_c\ .$$

Using the same method as in Lemma 2.20, it follows that $(\nabla_a\nabla_b - \nabla_b\nabla_a)w_c$ is a (1,3)-tensor depending only on the value of w_c at p, and thus $\nabla_a\nabla_b - \nabla_b\nabla_a$ defines a linear map from (0,1)-tensors at p to (0,3)-tensors at p. Hence we have shown:

Lemma 2.25: let w_a be a form field. Then there exists a tensor of type (1,3), say $R_{abc}{}^d$, such that

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)w_c = R_{abc}{}^d\ w_d\ .$$

$R_{abc}{}^d$ is called the Reimann curvature tensor.

Theorem 2.26: let t^a be a vector field. Then

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)t^c = -R_{abd}{}^c\ t^d\ .$$

proof: the proof is similar to that of Theorem 2.21:

$$
\begin{aligned}
0 &= (\nabla_a\nabla_b - \nabla_b\nabla_a)(t^c w_c) \\
&= \nabla_a(t^c\nabla_b w_c + w_c\nabla_b t^c) - \nabla_b(t^c\nabla_a w_c + w_c\nabla_a t^c) \\
&= t^c(\nabla_a\nabla_b - \nabla_b\nabla_a)w_c + w_c(\nabla_a\nabla_b - \nabla_b\nabla_a)t^c \\
&= t^c w_d R_{abc}{}^d + w_c(\nabla_a\nabla_b - \nabla_b\nabla_a)t^c
\end{aligned}
$$

Hence $\quad (\nabla_a\nabla_b - \nabla_b\nabla_a)t^c = -R_{abd}{}^c\, t^d$.

The same result can be obtained for any tensor field by using induction.

Corollary 2.27: 1) $R_{abc}{}^d = -R_{bac}{}^d$,

2) $R_{[abc]}{}^d = 0$

3) for ∇_a and g_{ab} satisfying $\nabla_a g_{bc} = 0$,

$$R_{abcd} = -R_{abdc}\,.$$

4) the Bianchi identity is satisfied:

$$\nabla_{[a}R_{bc]d}{}^e = 0 \qquad (\text{or } R^a{}_{b[cd;e]} = 0)\,.$$

proof: 1) holds trivially from Lemma 2.25

2) since by Theorem 2.21 $\nabla_a w_b = \partial w_b/\partial x^a + \Gamma^d{}_{ab} w_d$, we have

$$\nabla_{[a}\nabla_b w_{c]} = {}^1/_{3!}\ (\nabla_a\nabla_b w_c - \nabla_b\nabla_a w_c + \nabla_b\nabla_c w_a - \nabla_a\nabla_c w_b + \nabla_c\nabla_a w_b - \nabla_c\nabla_b w_a)$$

$$= {}^1/_{3!}\ [\partial_a\partial_b w_c + \partial_a\partial_b(\Gamma^d{}_{bc}w_d) - \partial_b\partial_a w_c - \partial_b\partial_a(\Gamma^d{}_{ac}w_d) + \partial_b\partial_c w_a + \partial_b\partial_c(\Gamma^d{}_{ca}w_d) - \partial_a\partial_c w_b - \partial_a\partial_c(\Gamma^d{}_{cb}w_d) + \partial_c\partial_a w_b + \partial_c\partial_a(\Gamma^d{}_{ab}w_d) - \partial_c\partial_b w_a - \partial_c\partial_b(\Gamma^d{}_{ba}w_d)]$$

$$= 0 \text{ by symmetry of } \partial_a \text{ and of } \Gamma^d{}_{**}$$

$$\Rightarrow\ 0 = 2\,\nabla_{[a}\nabla_b w_{c]} = \nabla_{[a}\nabla_b w_{c]} - \nabla_{[b}\nabla_a w_{c]}$$

$$\Rightarrow\ \forall\, w_d,\ R_{[abc]}{}^d w_d = 0.$$

3) by generalization of Theorem 2.26 to tensors,

$$0 = (\nabla_a\nabla_b - \nabla_b\nabla_a)g_{cd} = R_{abc}{}^e g_{ed} + R_{abd}{}^e g_{ce}$$

$$= R_{abcd} + R_{abdc}\ .$$

4) $(\nabla_a\nabla_b - \nabla_b\nabla_a)\nabla_c w_d = R_{abc}{}^e\,\nabla_e w_d + R_{abd}{}^f\,\nabla_c w_f$,

while $\nabla_a(\nabla_b\nabla_c w_d - \nabla_c\nabla_b w_d) = \nabla_a(R_{bcd}{}^e w_e)$

$$= w_e\nabla_a R_{bcd}{}^e + R_{bcd}{}^e\,\nabla_a w_e\ .$$

If we antisymmetrize in a,b & c these quantities, then the left hand sides become equal and we get

$$R_{[abc]}{}^e\,\nabla_e w_d + R_{[ab\underline{d}}{}^f\,\nabla_{c]}w_f = w_e\nabla_{[a}R_{bc]d}{}^e + R_{[bc\underline{d}}{}^e\,\nabla_{a]}w_e$$

where d is not antisymmetrized. By 2), the first term vanishes while the second terms on each side cancel out, leaving $\nabla_{[a}R_{bc]d}{}^{e} = 0$. •

Taking the trace of the Reimann tensor over its 2^{nd} and 4^{th} indices yields the Ricci tensor, R_{ac},

$$R_{ac} = R_{abc}{}^{d} ,$$

and the trace of the Ricci tensor

$$R = R_a{}^a$$

is called the scalar curvature.

Because of the symmetries of R_{abcd}, there are $n^2(n^2-1)/12$ independent components left. $n(n+1)/2$ of them can be found through the Ricci tensor; if $n > 3$, the remaining components can be found by using the Weyl tensor, C_{abcd},

$$C_{abcd} = R_{abcd} + 2(g_{a[c}R_{d]b} - g_{b[c}R_{d]a})/(n-2) - 2R\,g_{a[c}g_{d]b}/(n-1)(n-2).$$

(if $n=1$, $R_{abcd} = 0$, if $n=2$ the only independent component of R_{abcd} is R, and if $n=3$, $n^2(n^2-1)/12 = n(n+1)/2$).

<u>Definition 2.28</u>: the metrics **g** and **g'** are said to be conformal

if there exists a non-zero C^1 function Ω such that $\mathbf{g}' = \Omega^2 \mathbf{g}$. Note that the cone structure on V_p is preserved under conformal changes of metrics. Also, since $g'^{ab} = \Omega^{-2} g^{ab}$ and $g_{ab} = \Omega^{-2} g'_{ab}$, we have

$$\Gamma'^{a}{}_{bc} = \Gamma^{a}{}_{bc} + \Omega^{-1}(\delta^{a}{}_{b}\,\Omega_{,c} + \delta^{a}{}_{c}\,\Omega_{,b} - g_{bc}g^{ad}\,\Omega_{,d}).$$

Hence

$$R'^{ab}{}_{cd} = \Omega^{-2} R^{ab}{}_{cd} + \delta^{[a}{}_{[c}\,\Omega^{b]}{}_{d]}\ ;$$

and with some calculations $C'^{a}{}_{bcd} = C^{a}{}_{bcd}$, i.e. the Weyl tensor is conformally invariant.

<u>Definition 2.29</u>: an isometry is a diffeomorphism $\phi : \mathcal{M} \to \mathcal{M}$ for which $\phi_* \mathbf{g}$ is equal to $\mathbf{g}$, $\forall p \in \mathcal{M}$.

In such a case, the map $\phi_* : V_p \to V_{\phi(p)}$ preserves scalar products since

$$\mathbf{g}(u,v) = \phi_* \mathbf{g}(\phi_* u, \phi_* v)\,\big|_{\phi(p)} = \mathbf{g}(\phi_* u, \phi_* v)\,\big|_{\phi(p)}\ .$$

<u>Definition 2.30</u>: a vector field $\mathbf{\chi}$ is called a Killing vector field if the one-parameter group of diffeomorphisms ϕ_t generated by $\mathbf{\chi}$ forms a group of isometries.

<u>Theorem 2.31</u>: the vector field $\mathbf{\chi}$ is a Killing field if and

only if it satisfies the Killing equation

$$\mathcal{K}_{(a;b)} = 0 \qquad (\text{i.e. } \nabla_b \mathcal{K}_a + \nabla_a \mathcal{K}_b = 0).$$

<u>proof</u>: ($\Rightarrow$) for any function f

$$\begin{aligned}
[u,v](f) &= u[v(f)] - v[u(f)] \\
&= u^a \nabla_a (v^b \nabla_b f) - v^a \nabla_a (u^b \nabla_b f) \\
&= (u^a \nabla_a v^b - v^a \nabla_a u^b)\, \nabla_b f \\
\Rightarrow \quad [u,v]^b &= u^a \nabla_a v^b - v^a \nabla_a u^b \\
&= u^a \partial_a v^b - v^a \partial_a u^b \qquad \text{by Theorem 2.21} \\
&= (\mathbf{L}_v\, u)^b
\end{aligned}$$

therefore $\quad \mathbf{L}_{\mathcal{K}} g_{ab} = \mathcal{K}^c g_{ab;c} + \mathcal{K}^c{}_{;a}\, g_{cb} + \mathcal{K}^c{}_{;b}\, g_{ac}$

$$\Rightarrow \quad \mathbf{L}_{\mathcal{K}} g_{ab} = \mathcal{K}_{b;a} + \mathcal{K}_{a;b} \quad \text{since } g_{ab;c} = 0.$$

But since $\mathcal{K}$ is a Killing field,

$$0 = \lim_{t \to 0} (g - \phi_{*t} g) = \mathbf{L}_{\mathcal{K}} g$$

$$\Rightarrow \quad \mathcal{K}_{b;a} + \mathcal{K}_{a;b} = 0.$$

($\Leftarrow$) suppose that $\mathcal{K}_{b;a} + \mathcal{K}_{a;b} = 0$ for an arbitrary vector field $\mathcal{K}$.

$$\phi_{*t}\mathbf{g}\big|_p = \mathbf{g}\big|_p + \int_0^t d(\phi_{*s}\mathbf{g})\big|_p \qquad \text{(refer to discussion on maximal curves)}$$

$$= \mathbf{g}\big|_p + \int_0^t \frac{d}{ds}(\phi_{*s}\mathbf{g})\big|_p\, ds$$

$$= \mathbf{g}\big|_p + \int_0^t \frac{d}{dr}(\phi_{*s}\,\phi_{*r}\,\mathbf{g})\big|_{p,\ r=0}\ ds$$

$$= \mathbf{g}\big|_p + \int_0^t (\phi_{*s}\frac{d}{dr}\phi_{*r}\,\mathbf{g})\big|_{p,\ r=0}\ ds$$

$$= \mathbf{g}\big|_p - \int_0^t \phi_{*s}\left[(L_{\chi}\,\mathbf{g})\,(p)\big|_{\phi_{-s}}\right]\ ds$$

$$= \mathbf{g}\big|_p$$

$\Rightarrow$ χ is Killing vector field. •

To finish this chapter we state Stokes' Theorem and prove Gauss' Theorem.

The generalized Stokes' Theorem : if F is an (n-1)-form field on $\mathcal{M}$, then

$$\int_{\partial\mathcal{M}} F = \int_{\mathcal{M}} dF .$$

The proof can be found in (Spivak,1965).

Let $\{e^a\}$ be a basis for one-forms. Then we define the n-form

$$\zeta = n!\, g\, e^1 \wedge \ldots \wedge e^n,$$

where $g = \sqrt{|\det \mathbf{g}|}$. The components of ζ are

$$\zeta_{ab\ldots d} = n!\; g\; \delta^1{}_{[a}\delta^2{}_b \cdots \delta^n{}_{d]} \; .$$

Thus

$$\zeta^{ab\ldots d} = (-1)^{(n-s)/2} \quad n!\; g\; \delta^{[a}{}_1\delta^b{}_2 \cdots \delta^{d]}{}_n$$

where s is the signature of $\mathbf{g}$ and $(n-s)/2$ is the number of negative eigenvalues. Hence

$$\zeta^{ab\ldots d}\,\zeta_{ef\ldots h} = (-1)^{(n-s)/2} \quad n!\; \delta^{[a}{}_{[e}\delta^b{}_f \cdots \delta^{d]}{}_{h]}$$

and

$$\zeta^{ab\ldots d}{}_{;e} = 0 = \zeta_{ab\ldots d;e}$$

Define the volume of an n-dimensional submanifold $\mathcal{U}$ to be

$$\frac{1}{n!}\int_{\mathcal{U}} \zeta$$

with respect to the metric $\mathbf{g}$. If f is a function on $\mathcal{M}$ then the integral of f over $\mathcal{U}$ is

$$\int_{\mathcal{U}} f \, dv = \frac{1}{n!} \int_{\mathcal{U}} f \zeta = \int_{\mathcal{U}} f \, g \, dx^1 \cdots dx^n \quad ,$$

where dv is a volume measure on $\mathcal{M}$, not the exterior derivative. Note that by Theorem 2.15, the definition of this integral is independent of the choice of coordinates.

Let $\mathbf{X}$ be a vector field on $\mathcal{M}$. The contraction of $\mathbf{X}$ with ζ is

$$(\mathbf{X} \cdot \zeta)_{b \ldots d} = X \zeta_{ab \ldots d} \quad ,$$

an $(n-1)$-form field. The integral of this form field over an $(n-1)$-dimensional orientable submanifold such as the boundary $\partial\mathcal{M}$ is given by

$$\int_{\partial\mathcal{M}} X^a \, d\sigma_a = \frac{1}{(n-1)!} \int_{\partial\mathcal{M}} \mathbf{X} \, \zeta \quad ,$$

where ζ defines the measure-valued norm $d\sigma_a$. By Stokes' Theorem we thus have

$$\int_{\partial\mathcal{U}} X^a \, d\sigma_a = \frac{1}{(n-1)!} \int_{\mathcal{U}} d(\mathbf{X} \cdot \zeta) \quad .$$

But as

$$dF = F_{a\ldots c;d}\, dx^d \wedge dx^a \wedge \ldots \wedge dx^c \Leftrightarrow (dF)_{a\ldots cd} = (-1)^q F_{[a\ldots c;d]}$$

for any q-form F, we get

$$\begin{aligned} d(\mathbf{X}\cdot\zeta) &= (-1)^{n-1} (\mathbf{X}^f \zeta_{f[a\ldots d})_{;e]} \\ &= (-1)^{n-1} \delta^s{}_{[a}\cdots\delta^t{}_d\delta^u{}_{e]}\, \zeta_{fs\ldots t}\, \mathbf{X}^f{}_{;u} \\ &= (-1)^{(n-1)+(n-s)/2} ({}^1/_{n!})\, \zeta^{s\ldots tu}\, \zeta_{a\ldots de}\, \zeta_{fs\ldots t}\, \mathbf{X}^f{}_{;u} \\ &= \zeta_{a\ldots de}\, \delta^s{}_{[s}\cdots\delta^t{}_t\delta^u{}_{f]}\, \mathbf{X}^f{}_{;u} \\ &= ({}^1/_n)\, \zeta_{a\ldots de}\, \mathbf{X}^f{}_{;f}\,. \end{aligned}$$

therefore

$$\int_{\partial u} \mathbf{X}^a\, d\sigma_a = \int_u \mathbf{X}^b{}_{;b}\, d\upsilon$$

for any vector field **X** ; this is Gauss' Theorem.

CHAPTER III

1. Minkowski's space-time

As it was pointed out in chapter I, our mental representation of physical objects is underlined by some kind of four dimensional continuity. Furthermore, ever since we were young, we learnt that the spacio-temporal images obtained in the brain depend very much on our own point of view (Piaget,1966). For example there can be an infinite number of pictures of the same room for a given time; similarly one could draw numerous different maps for an atlas from the same terrestrial globe. It was not before Minkowski that such intuitive concepts were given an exact representation. Indeed, it does not require much idealization to conceive the space-time continuum as a four-dimensional differentiable manifold since the latter can be covered by open sets which are mapped to $\mathbb{R}^4$ by homeomorphisms and which have the important property of sending intersections of open sets to homeomorphic parts of $\mathbb{R}^4$. This conception is in accordance with our mental structure if we allow that physical locations and durations be idealized to points and instants respectively. On the manifold, such points are called events or worldpoints as Minkowski named them.

Many different charts can be used to construct a 'rigid grid' necessary to obtain measurements such as distances and durations. The choice of a coordinate system is critical because relativity demands that the laws governing nature be independent of the transformations applied on a chosen chart. We define an inertial frame of reference to be a Cartesian coordinate system that meets the two following conditions:

I. Three free particles emitted non-collinearly from a point of an open set of a 4-dimensional manifold describe straight lines in this set.

II. A light ray transmitted in vacuo from a point of an open set describes a straight line in this set.

And to define an inertial time frame of reference:

III. A free particle moving in an inertial frame (I & II) has constant speed.

IV. A light ray transmitted in vacuo from a source at rest in an inertial frame has constant speed.

The last two conditions settle the definition of time and of simultaneity. Since they are not the only ones possible, they are sometimes named after their establisher, Einstein time and Einstein simultaneity (for more details on these definitions see (Einstein,1950) for example).

Let (O_x, x) be a chart for a 4-dim manifold $\mathcal{M}$. For $p \in O_x$, $x(p) = (x^0(p), x^1(p), x^2(p), x^3(p)) \in \mathbb{R}^4$, where $x^0 =$ time t. If O_x is an inertial frame (I & II) and if the real-valued function x^0 assigns an Einstein time (III & IV) to every $p \in \mathcal{M}$, then x is said to be a Lorentz chart. Lorentz charts are linear bijections of $\mathcal{M}$ onto $\mathbb{R}^4$.

Let x and y be two Lorentz charts (omitting to write their respective domains). The bijections $x \cdot y^{-1}$ and $y \cdot x^{-1}$ from $\mathbb{R}^4$ into $\mathbb{R}^4$ are called Lorentz coordinate transformations; they act like permutations of $\mathbb{R}^4$. It is not hard to see that all such transformations form a group, called the Poincaré group. The elements of this group are defined by the system of equations

$$y^i = P^i{}_j x^j + k^i$$

where $P^i{}_j$ is the ij-th component of the Jacobian matrix $(\partial y^i / \partial x^j)$, and $k^i \in \mathbb{R}$. Let P denote this Jacobian. P is non-singular since every element $x \cdot y^{-1}$ has for inverse $y \cdot x^{-1}$. If $k^i = 0 \;\; \forall\, i$, then $y^i = P^i{}_j x^j$ defines the homogeneous Lorentz group L(4).

In this context the relativity principle has the precise formulation:

> the states of physical systems described by the laws of nature are not affected by the application

of elements of the Poincaré group, i.e. it does not make any difference if these states are referred to the one or the other Lorentz chart.

By using this principle together with the light principle, we can obtain the familiar Lorentz transformation (as said in chapter I). This transformation describes the three-parameter class of velocity changes in the Poincaré group: let x and y be two Lorentz charts and let 0_y move with speed v in 0_x parallelly to x^1 (it could have been x^2 or x^3), then the transformation is:

$$y^0 = \frac{x^0 - vx^1/c^2}{\sqrt{1 - v^2/c^2}} , \qquad y^1 = \frac{x^1 - vx^0}{\sqrt{1 - v^2/c^2}}$$

$$y^2 = x^2 \qquad\qquad y^3 = x^3 .$$

The Jacobian of this transformation is:

$$\left(\frac{\partial y^i}{\partial x^j}\right) = \begin{bmatrix} \beta & -\beta v/c^2 & 0 & 0 \\ -\beta v & \beta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where $\beta = (1 - v^2/c^2)^{-1/2}$.

The Poincaré group is a 10-parameter Lie group; three velocity parameters, three parameters for space rotations and four parameters for space and time translations. Since the Poincaré group is generated by L(4) plus the group of translations, it will be sufficient to characterize L(4); we will use the natural Jacobian representation of 4×4 non-singular matrices.

Consider the matrix

$$Q = \begin{bmatrix} -c^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

with components Q_{ij} .

By using the Lorentz transformation we have,

$$Q_{ij}\, y^i y^j = -c^2 (y^0)^2 + (y^1)^2 + (y^2)^2 + (y^3)^2$$

$$= -c^2 (x^0 - vx^1/c^2)^2 \,/\, (1 - v^2/c^2) + (x^1 - vx^0)^2 \,/\, (1 - v^2/c^2) + (x^2)^2 + (x^3)^2$$

$$= [-c^2(x^0)^2 - v^2(x^1)^2 / c^2] / (1 - v^2/c^2)$$

$$+ [(x^1)^2 - v^2(x^0)^2] / (1 - v^2/c^2) + (x^2)^2 + (x^3)^2$$

$$= [(x^0)^2 (v^2 - c^2) + (x^1)^2 (1 - v^2/c^2)] / (1 - v^2/c^2)$$

$$+ (x^2)^2 + (x^3)^2$$

$$= -c^2(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2$$

$$= Q_{ij}\, x^i x^j \qquad \text{(for velocity transformations).}$$

Furthermore, since $y^i = P^i{}_j\, x^j$ defines an isometry of hyperplanes when P represents a space rotation, we have

$$Q_{\alpha\beta}\, y^\alpha y^\beta = Q_{\alpha\beta}\, x^\alpha x^\beta \quad , \quad \alpha,\beta = 1,2,3.$$

(for rotational transformations).

Hence the quadratic form $Q_{ij}\, x^i x^j$ is invariant under L(4). This invariance will provide the desired characterization of L(4) in this way: for any $P \in L(4)$

$$Q_{ij}\, x^i x^j = Q_{ij}\, y^i y^j = Q_{ij}\; P^i{}_k\, x^k \;\; P^j{}_l\, x^l$$

$$\Leftrightarrow \quad Q_{kl} = P^i{}_k\; Q_{ij}\; P^j{}_l$$

$$\text{i.e.} \quad \Leftrightarrow \quad Q = P^T Q P .$$

Therefore L(4) can be characterized as the subgroup of $GL(4,\mathbb{R})$ represented by all 4×4 non-singular matrices P which satisfy $Q = P^T Q P$. Since the metric will have the form $Q_{ij}\,dx^i dx^j$ (in Minkowski's space-time), this characterization endows to the manifold a geometry that is completely determined by L(4)-invariance.

The Lie group L(4) can now be decomposed:

$$\det Q = (\det P^T)(\det Q)(\det P)$$
$$\Rightarrow \quad (\det P)^2 = 1$$
$$\Rightarrow \quad \det P = \pm 1 .$$

Also, $Q_{00} = P^i{}_0 \; Q_{ij} \; P^j{}_0$

$$\Rightarrow \quad -c^2 = -c^2 (P^0{}_0)^2 + \Sigma_\alpha (P^\alpha{}_0)^2$$
$$\Rightarrow \quad (P^0{}_0)^2 - \Sigma_\alpha (P^\alpha{}_0 / c)^2 = 1 \qquad \text{for } \alpha = 1,2,3$$
$$\Rightarrow \quad (P^0{}_0)^2 \geq 1 .$$

Hence there are four topologically disconnected parts in L(4):

i) $\{ P \in L(4) \mid \det P = 1 \;\;\&\;\; P^0{}_0 \geq 1 \}$

ii) $\{ P \in L(4) \mid \det P = -1 \;\;\&\;\; P^0{}_0 \geq 1 \}$

iii) $\{ P \in L(4) \mid \det P = 1 \;\;\&\;\; P^0{}_0 \leq 1 \}$

iv) $\{ P \in L(4) \mid \det P = -1 \;\;\&\;\; P^0{}_0 \leq 1 \}$

The set i) constitutes a Lie subgroup of L(4) denoted by $\mathcal{L}_0$. It acts on L(4) by $(L,P) \to LP$ for $L \in \mathcal{L}_0$ and $P \in L(4)$.

The sets ii), iii) & iv) are, under this left action by $\mathcal{L}_0$, orbits† of time reversals ($P^0{}_0 \leq 1$) and reflections of hyperplanes about the origin (det P = -1). The union of i) and ii) is called the orthochronous homogeneous Lorentz group. $\mathcal{L}_0$ is called the proper orthochronous homogeneous Lorentz group; proper means that the cartesian systems of x and y are both right-handed or both left-handed, and orthochronous means that the time coordinates of both charts increase together, i.e. $\partial x^0 / \partial y^0 > 0$. $\mathcal{L}_0$ together with space-time translations (inhomogeneous) is denoted by $\mathcal{L}$.

The geometry of the manifold will thus be characterized by $\mathcal{L}$-invariance. This can be seen by simply defining the Minkowski interval between two events $p, q \in \mathcal{M}$ in this way: let x be a Lorentz chart and let $A_{\mathcal{L}}$ be the set of all Lorentz charts related to x by a transformation of the group $\mathcal{L}$. (Since all such transformations are C^∞, $A_{\mathcal{L}}$ gives to the set of all events the structure of a C^∞ 4-dim manifold in the sense of Definition 2.1). If $x \in A_{\mathcal{L}}$ then we define

$$I(p,q) = \eta_{ij}(x^i(p) - x^i(q))(x^j(p) - x^j(q))$$

† if a group G acts on a set S, the orbit of $s \in S$ is the set $\{gs \mid g \in G\}$

where η_{ij} is Q_{ij} with time dimensions such that $c = 1$.

$I(p,q)$ is clearly independent of the choice of elements of $A_{\mathcal{L}}$ and is thus invariant under the action of $\mathcal{L}$ on $\mathcal{M}$. By this action the geometry of $\mathcal{M}$, endowed by $\mathcal{L}$-invariance, is identical to the geometry of $\mathbb{R}^4$ endowed by the action of $\mathcal{L}$ as a group of coordinate transformations. Space-time, or in this context Minkowski space-time, is seen as the collection of all events structured into a C^∞ 4-dim manifold by $A_{\mathcal{L}}$ together with a geometry endowed by $\mathcal{L}$-invariance. The physical significance of this definition is grounded on the physical hypotheses behind the definition of Lorentz charts. If these hypotheses turn out to be true only locally or approximatively, as is the case in the general context of relativity, the importance of the matters discussed hitherto will not be diminished; they will only become the limit of a theory in which gravity is taken into account, just as in Calculus a tangent line approximates the curve near the point common to both; the straighter the curve the better the approximation. This analogy will in fact be quite helpful.

Up to this point we have studied and characterized the structure of the set of possible events and shown that the manifold structure responded accurately to the demands of the principle established in chapter I. We will now use this set up to construct a mathematical frame in order to express relations between physical quantities. This will further show the necessity to employ C^∞ 4-dim manifolds.

All physics experiments yield real numbers and thus all relation between physical quantities must eventually be reducible to reals. Moreover, one should include in a theory all matter fields which can be experimentally detected. For these reasons, maps from vectors (and dual vectors) to reals are required to express the laws of nature. Tensors, while being the only relations defined by the manifold structure, encompass a general class of such multilinear maps (analytic maps can be Taylor expanded); thus tensors will be used to express equations between the fields that are emitted and received by the matter content of space-time.

By Definition 2.8, tensors can be expressed in any preferred basis, and thus they depend on the different choices of charts. (We also gave the tensor transformation law in terms of any basis). It thus becomes natural to construct a basis by using Lorentz charts so that our tensors agree with the Minkowski space-time. Recall the definition of a tangent vector (Definition 2.4); there was no preferred choice of curve $\lambda(t)$. Let x be a Lorentz chart defined at $p \in \mathcal{M}$; the i-th parametric line of x through p, λ^i, is a map from $\mathbb{R}$ into $\mathcal{M}$ defined by

$$t \to \lambda^i(t) := \{q \in \operatorname{dom} x \mid x^i(q) = t,\ t \in \mathbb{R}$$
$$\& \text{ for } j \neq i \ \ x^j(q) = x^j(p)\}$$

and thus each coordinate function x^i determines a curve λ^i. Hence by Definition 2.4 the tangent vector to λ^i at p is the linear map

$$\lim_{h\to 0} \frac{1}{h} \{ f[\lambda^i(t+h)] - f[\lambda^i(t)] \}$$

$$= \lim_{h\to 0} \frac{1}{h} \{ f \cdot x^{-1}[x^1(p),\ldots,x^i(p) + h,\ldots,x^n(p)]$$
$$- f \cdot x^{-1}[x^1(p),\ldots,x^i(p),\ldots,x^n(p)] \},$$

$$= (\partial f/\partial x^i)\,|_p \ .$$

Therefore by Theorem 2.5 the set of vectors $(\partial/\partial x^1,\ldots,\partial/\partial x^n)\,|_p$ spans V_p with the difference that this time the coordinate derivatives are determined by the Lorentz chart x. If $X = (\partial/\partial x^1,\ldots,\partial/\partial x^n)$, where the $\partial/\partial x^i$'s are vector fields, we can now refer to the components of a tensor relative to X as its components relative to the chart x.

An important example of a tensor is the metric defined in 2.11. This (0,2)-tensor is completely determined on $\operatorname{dom} x$ by the $n(n+1)/2$ C^∞ functions $g(\partial/\partial x^i, \partial/\partial x^j)$, known as the components of g relative to the chart x. Since g is symmetric, one can view these components as the lower (or upper) triangle of the $n \times n$ matrix which represents $\mathbf{g}$ ($n + n-1 + n-2 + \ldots + 1 = 10$ components for $n=4$). Thus we say that the set of all events is a pair $(\mathcal{M}, \mathbf{g})$ where $\mathcal{M}$ is a 4-dim C^∞ manifold (with structure and geometry as before) and where $\mathbf{g}$ can be specified by the values of the ten vector fields $\partial/\partial \mathbf{x}^i$, whose integral curves are the parametric lines of the chart $\mathbf{x} \in A_L$.

The metric is only an example of tensor specification by a basis determined by Lorentz charts. In general an (r,s)-tensor field can be uniquely associated with a mapping from the set of basis $\{\partial/\partial \mathbf{x}^i\}$ into the set of functions to the power n^{r+s}; these maps assign to each $\{\partial/\partial \mathbf{x}^i\}$ the list of tensor components relative to $\{\partial/\partial \mathbf{x}^i\}$. In the case where the tensor is $\mathbf{g}$ and the manifold is 4-dimensional we have: $\{\partial/\partial \mathbf{x}^i\} = (\partial/\partial \mathbf{x}^0, \partial/\partial \mathbf{x}^1, \partial/\partial \mathbf{x}^2, \partial/\partial \mathbf{x}^3)$, $r + s = 2$, and thus the mapping is from $(\partial/\partial \mathbf{x}^0, \partial/\partial \mathbf{x}^1, \partial/\partial \mathbf{x}^2, \partial/\partial \mathbf{x}^3)$ into $(\text{functions})^{16}$, a 4×4 matrix of functions of which only ten components are independent. This matrix is just (g_{ij}) in

$$ds^2 = g_{ij}\, d\mathbf{x}^i\, d\mathbf{x}^j ,$$

where ds is the line element expressed in terms of the coordinate differentials associated with the chart $\mathbf{x}$.

This being said, another point must be discussed before expressing physical laws by tensor relations. These physical laws must determine the variations of the matter fields from worldpoints to worldpoints. Since tensors are multilinear maps which have for domain products of vector fields, the problem reduces to comparing vectors at different points on the manifold. Definition 2.4 corresponds to the directional derivative of a vector since $(\partial f/\partial t)|_{\lambda}$ is the derivative of f in the direction of λ with respect to the parameter t. Our interest is thus to compare the direction of distant vectors, i.e. their relative parallelism. When the manifold is parallelizable, i.e. when there exists a family of n vector fields $\mathbf{v}^1,\ldots,\mathbf{v}^n$ such that their values at each $p \in \mathcal{M}$ form a basis for the tangent space V_p, then the vectors at different points can be partitioned into equivalent classes of parallel vectors. A trivial example is Euclidean 3-space. On that manifold, a vector v at p is parallel to a vector u at q if the translation map ϕ sending p to q is such that $u = \phi_*(v)$. Similarly if $\mathcal{M}$ is developable, like a cylinder or a cone, then $v|_p$ is parallel to $u|_q$ if $\phi_{*p}(v)$ is parallel to $\phi_{*q}(u)$, where ϕ is an isometry of a region of $\mathcal{M}$ into a plane.

The above cases refer to so called absolute parallelism. When $\mathcal{M}$ is not parallelizable, it is possible to talk of relative parallelism, relative to the path (curve) joining the points p and q. For example let $\mathcal{M}$ be a sphere and λ a curve on $\mathcal{M}$ from p to q. Let a plane touch $\mathcal{M}$ at p. Since this plane is tangent, a vector v at p corresponds to a directed segment v' on the plane. Now allow the sphere to roll on the plane, without slipping, in such a way that the point of contact is always a point in the range of λ. When the point of contact is q, the vector u at q correponds to a directed segment u' on the plane. v is said to be parallel to μ along λ if and only if v' is parallel to u'. This concept of path-dependent parallelism can be extended to a wide variety of manifolds through the notion of linear connection, and more specifically the affine connection $\Gamma^a{}_{bc}$ (2.2) obtained by Theorem 2.24 in terms of the metric (Γ is sometimes called the Levi-Civitá connection). Many different connections could be used but this one seems to be the preferred one. We do not intend to develop the general concept of parallelism for it would take several sections. The main idea is that the choice of a linear connection on the manifold determines a linear isomorphism $\phi^\lambda{}_{pq}$ from V_p to V_q with λ joining p and q on $\mathcal{M}$. Isomorphisms between tensor spaces are then induced by this map. $\phi^\lambda{}_{pq}$ is said to parallel transport V_p to V_q along λ.

Nevertheless the notions developed in chapter II about the Levi-Civitá connection are sufficient to examine parallel transport of a vector through an infinitesimal distance. This will be enough to fulfill the task of observing variations of tensor fields at neighbouring points. Consider the following set up (Einstein,1974). Let v^a be the coordinates of a vector at $p \in \mathcal{M}$ with respect to the chart x. Let q be infinitely close to p and let $v^a + \delta v^a$ denote the coordinates of v^a displaced to q along the infinitely small distance dx^a. v^a is then said to be parallelly displaced from p to q. If the coordinates of v^a at q with respect to the chart x are $v^a + dv^a$ then we expect the difference between $v^a + dv^a$ and $v^a + \delta v^a$ to be given by the 'slope' of v^a multiplied by the distance dx^a. In other words we expect that,

$$d v^b - \delta v^b = \nabla_a v^b \, dx^a \quad ,$$

in terms of the general concept of derivative ∇_a. By Theorem 2.21 there exists a tensor $\Gamma^b{}_{ac}$ such that

$$\begin{aligned} d v^b - \delta v^b &= (\partial v^b/\partial x^a + \Gamma^b{}_{ac} v^c) \, dx^a \\ &= (\partial v^b/\partial x^a) \, dx^a + \Gamma^b{}_{ac} v^c \, dx^a \\ &= dv^b + \Gamma^b{}_{ac} v^c \, dx^a \\ \Rightarrow \quad \delta v^b &= -\Gamma^b{}_{ac} v^c \, dx^a \\ \Rightarrow \quad v^b + \delta v^b &= v^b - \Gamma^b{}_{ac} v^c \, dx^a \ . \end{aligned}$$

Since the same result can be obtained for a dual vector through Lemma 2.20, and since $\Gamma^b{}_{ac}$ is completely determined by the metric and its partial derivatives, this calculation describes explicitly the variation of a tensor field upon displacement dx^a between neighbouring points of the manifold.

Note that if the components of the metric are constants, as in the case of Euclidean 3-space or the case of Minkowski space-time, then all the components $\Gamma^b{}_{ac}$ are zero and thus $\delta v^b = 0$, meaning that v^a stays parallel to itself when transported on these manifolds along infinitesimal distances. In these cases the geodesic curve (2.1) is just $d^2x^a/dt^2 = 0$, a straight line (i.e constant speed as in the case of Lorentz transformation).

It is now possible to take a look at some space-time physics on Minkowski's manifold. Although the theory of manifolds and the theory of tensors were developed in the last century, it was Einstein's theory of General Relativity that encouraged mathematical research in these fields and so, the scheme developed in this section was created recently. Indeed, Minkowski referred to objects of space-time as space-time vectors composed by four real numbers together with a rule for transforming these numbers whenever an element of $\mathcal{L}_0$ was applied. For us, a vector field is the assignment of a four-vector at each point of the manifold; this vector field varies smoothly (C^∞) from one point to another. This can be done by replacing the four real numbers of a space-time vector by four C^∞ functions. Here is how: let $(\partial/\partial x^0, \partial/\partial x^1, \partial/\partial x^2, \partial/\partial x^3)$ be the basis which makes Minkowski space-time parallelizable. Then any vector v element of a vector field $\boldsymbol{v}$

can be, at a given $p \in \mathcal{M}$, expressed as the linear combination $v^i (\partial/\partial x^i)|_p$, $v^i \in \mathbb{R}$. Thus any vector can be described as the set (v^0, v^1, v^2, v^3), relative to the chart x, together with a rule for calculating its new components whenever an element of $\mathcal{L}_0$ is applied. In this way v can be seen as a 4-vector.

Consider the motion of a particle π in an inertial frame represented by a curve $\lambda(t)$, parametrized by time t. Let $(\partial/\partial t)_\lambda$ be the tangent at any point $\lambda(t) = p$, $t \in \mathbb{R}$ and p an element of Minkowski space-time $(\mathcal{M},\eta)$, where $\eta = \eta_{ij}\, dx^i\, dx^j$. λ is spacelike, timelike or null if for all t in dom λ, $(\partial/\partial t)_\lambda$ is spacelike, timelike or null respectively. Since it is assumed that nothing travels faster than light, λ will always be timelike (tachion universes not being of concern here). In Newtonian mechanics, the velocity of π at time t is a vector $\underline{u}^\alpha = (\partial/\partial t)_\lambda$ with components relative to a Cartesian chart x †

$$\underline{u}^\alpha = dx^\alpha/dt \quad , \quad \alpha = 1,2,3.$$

The achievement of special relativity is to explain the meaning of time t. It states that the timelike curve λ should

---

† the cartesian chart x is an injective mapping from $\mathcal{O}_x$ to $\mathbb{R}^3$ such that $|x^\alpha(p)|$ is the cartesian distance from p to the planes $OX_\beta X_\gamma$, $\alpha,\beta,\gamma = 1,2,3$. $O, X_1, X_2, X_3 \in \mathcal{O}_x$ and OX_1, OX_2, OX_3 are mutually perpendicular.

be parametrized by the proper time τ defined by

$$\tau = \int (-\eta_{ij} T^i T^j)^{1/2} \, dt'$$

where t' is an arbitrary parametrization of λ and T^i is the tangent to λ under this parametrization. The 4-velocity of π is given by

$$u^i = dx^i/d\tau \quad , \quad i = 0,1,2,3.$$

where x is a Lorentz chart. Because of time dilation (a consequence of Lorentz transformation) we have:

$$dt = \beta \, d\tau \ ,$$

where $\beta = (1 - |u|^2)^{-1/2}$, $|u|$ being the speed of π in dom x.

Hence

$$u^0 = \frac{dx^0}{dt}\frac{dt}{d\tau} = \beta, \qquad u^\alpha = \frac{dx^\alpha}{dt}\frac{dt}{d\tau} = \beta \underline{u}^\alpha , \qquad \alpha = 1,2,3.$$

The 4-acceleration of π relative to x has thus for components

$$a^i = du^i/d\tau .$$

If π is unaffected by external forces, then it has for equation of motion

$$u^i \frac{\partial}{\partial x^i} u^j = 0 ,$$

which is π's geodesic in an inertial frame of reference.

Let m_0 be the rest mass of the particle. The 4-momentum is then

$$p^i = m_0 u^i .$$

Note that $p^0 = m_0 u^0 = \beta m_0 = m$ (the true mass depending on the speed of the particle), but since our time units are such that $c = 1$, p^0 is really mc^2: the energy of π as a function of the speed. Thus the principles of energy and momentum conservation are both included in the absolute principle of 4-momentum conservation. ~~The~~ rate of change of the dynamical state of the particle is called the 4-force; its components are given by

$$K^i = dp^i/d\tau ,$$

where K^0 is the rate of change of π's energy and K^α $(\alpha = 1,2,3)$ is the rate of change of π's momentum $d(mu^\alpha)/d\tau$. Since we constructed a structure in such a way that the laws of nature be unaffected by coordinate transformations and since $K^0 = 0$ & $K^\alpha = m_0(du^\alpha/dt)$ when dom x is an inertial frame, K^i is in effect a frame independent representation of the force on a particle, determined by the Newtonian force when restricted to the particle's own inertial frame.

2. Gravity

How can gravity be measured ? How fast does gravity travel ? Since the 'micro-mechanical' explanation of gravity has not been settled yet, we can only measure the effect of gravity, i.e. the acceleration of bodies in gravitational fields. As for the second question, the same lack of quantum understanding makes it only possible to expect that the action attributed to gravity does not propagate faster than light, although some recent theories conject that the phase velocity of gravitational waves is greater than the transversal speed of light. Natural philosophers of Newton's epoch and Newton himself (not to say that Newton was not a natural philosopher) thought that attraction could act on matter only if some kind of particle dragged the attracted body in its flow by colliding with it, the same way a swimmer is 'attracted' down flow by the colliding molecules of water. For Newton the flow of ether was the micro-mechanical explanation to gravity, but he never could, at such an early time, detect ether, though he tried many times with pendula (Dobbs,1975). Later, when Michelson and Morley unsuccessfully tried to detect ether, the whole idea was dropped. Many discoveries were, before this century, related to ether; even Maxwell summarized his electrodynamic philosophy as a determination of aethereal motion set by electric currents and magnets. In 1864 he said before the Royal Society:

> "It appears therefore that certain phenomena in electricity and magnetism lead to the same conclusion as those in optics, namely that there is an aethereal medium pervading all bodies, and modified only in degree by their presence; that the parts of this medium are capable of being set in motion by electric currents and magnets;..." †

Lorentz, who was one of the first to examine the possibility that the interferometer null result was due to the shrinking of the arm in the direction of motion, also based much of his research on this aethereal medium. Many others were motivated by this medium.

And with regard to gravity, Newton explains in a tract written in 1679 entitled De aere et aethere †† that gravity is the pressure of a descending aethereal shower; in the past mechanical philosophies required that motions be caused by impact phenomena, as we previously said.

In any case, our concern here is not to try to explain

† (Torretti,1983)

†† De aere et aethere is available in Newton, Unpublished papers, pp. 214-28. See (Dobbs,1975).

'microgravity' but only to understand the effects of 'macrogravity' measured in metres per second squared. In fact, the equivalence principle can be viewed as the assertion that this measure of gravity is exact, i.e. that the effects of a gravitational field of strength γ are the same as the effects of a uniform acceleration $-\gamma$. Or in other words: let Σ_1 be a system accelerated in the direction of the y-axis and let γ be the constant magnitude of its acceleration. Let Σ_2 be a system in a uniform (homogeneous) gravitational field which imparts an acceleration of $-\gamma$ to all objects. Then all the laws of physics are the same in Σ_1 and Σ_2 . In the light of what has been said in the begining of this paragraph, the equivalence principle does not seem to be relevant; but its importance is revealed in Einstein's theory of general relativity.

In the last section we have only considered the structure of space-time relative to changes of coordinate systems moving with constant speeds, i.e. which were not under the influence of gravity. Thus, as we did before, we should try to construct inertial frames of reference. We could be tempted to attribute to free fall the name of inertial frame and then do the following: let F be a free falling body in a gravitational field, on whose path lies a point p. Let $\mathbf{O}_p$ be a space-time region around p in which gravity is homogeneous. A Lorentz chart $\boldsymbol{x}$ in then construced on $\mathbf{O}_p$, as usually, by means of clocks and rods at rest in F. Then, some would say: By the equivalence principle, the laws of nature should take the same form in terms of this local Lorentz chart as they would when expressed in terms of a Lorentz chart in absence of gravity.

This point of view is questionable: gravity-free regions are non-existant, as far as we can tell; similarly free falling laboratories are also quite rare and somewhat fancy. But even if we wanted to imagine such regions of space-time, we would face another difficulty: the homogeneity of the gravitational field in 0_p can only be achieved if 0_p is infinitely small, in fact so small that there may not even exist such a region. Therefore we may have some justified doubts as to whether the domains of the local Lorentz charts could cover all of space-time. And even if we agreed on some margin of accuracy within which the field would be homogeneous, and in this way cover all of space-time, the laws of nature, as referred to local charts would only encompass some very small portion of the universe. The collection of such portions, in each of which special relativity is approximatively true, could only be viewed as being a picture of the cosmos in the light of truly universal theory, which would unite the many nearby homogeneous fields into a single continuous highly inhomogeneous gravitational field. Einstein's General Theory of Relativity is a successful attempt (even accepted by those who were sceptical at first) to this universal theory, and as we will see, it does not use free falling frames of reference.

First we observe some consequences of the equivalence principle. Consider a rotating disk of radius r and of angular velocity ω, with a clock at the centre and a clock at r. The time t measured at the centre is related to the time t' at r by

$$t = \frac{t'}{\sqrt{1 - v^2/c^2}} = \frac{t'}{\sqrt{1 - \omega^2 r^2/c^2}} .$$

By the equivalence of the inertial and gravitational mass and by Newtonian gravitational theory we have

$$- \operatorname{grad} \phi = \omega^2 \vec{r}$$

where ϕ is the gravitational potential. Note that ϕ need not homogeneous for this application of the equivalence principle since the magnitude of the acceleration on the disk depends on r, contrarily to the constant acceleration of Σ_1 in the original statement of the equivalence principle. This equation represents the equality of gravitational acceleration and of acceleration due to centrifugal forces. Thus

$$\partial\phi/\partial r = \omega^2 r$$

$$\Rightarrow \quad \phi = -(\omega^2 r^2)/2 \quad , \quad \text{for } \phi = 0 \text{ at } r = 0.$$

$$\Rightarrow \quad t = t'(1 + 2\phi/c^2)^{-1/2} \ .$$

Hence, as we said in the space-time principle of chapter I, gravitation determines proper time.

For a massive star of radius r,

$$\phi = -GM / r \quad , \quad \text{for } \phi = 0 \text{ at large distances}$$

$$\Rightarrow \quad t'(\text{surface}) \approx t(\text{large distance}) \ (1 - GM/rc^2)$$

where G is Newton's gravitational constant. From this result we can obtain the gravitational red shift

$$v' \approx v\,(1 - GM/rc^2) \ .$$

Futhermore, if c_ϕ is the speed of light at a point where the gravitational potential is ϕ then

$$c_\phi = c(1 + \phi/c_\phi^{\,2}) \ .$$

From these results, a question naturally arises: how can gravity be included in the geometrical context of the theory of relativity ? Unlike electromagnetic forces, gravity cannot be 'transformed away'. It is impossible to construct an inertial frame in which the observer is at rest. This is probably due to the fact that gravity is always positive, the more mass is piled up, the stonger the gravitational field. For example, the worldpath of a particle on earth could not practically be observed in a frame exempt from the effects of gravity. In some way, this leads to a concept which is imposed on us by these facts together with the results of the previous paragraph; we leave it to the discoverer to formulate this concept:

> "The metrically real is given only through the combination of the space-time coordinates and the mathematical quantities which describe the gravitational field." †

Let $\mathcal{M}$ be a surface in Euclidean space, diffeomorphic to $\mathbb{R}^2$. At any given point $p \in \mathcal{M}$ all the tangent vectors form a plane, the tangent plane V_p . If $\mathcal{M}$ is some portion of a cone or of a cylinder, then there exists an isomorphism which maps $\mathcal{M}$ onto a neighbourhood of p in V_p . All such surfaces are said to be flat. In general it is not possible to 'fit' $\mathcal{M}$ in a region of the plane without having to 'stretch and shrink' some parts. The Gaussian curvature measures, in that jargon, the quantity of stretching and shrinking required to fit $\mathcal{M}$ into a plane; one can easily see that according to this intuitive definition, the curvature of a flat surface is

† Einstein, A. (1934), "Mein Weltbild", (Querido Verlag, Amsterdam). See (Torretti,1983).

everywhere zero, and is constant on a sphere. If $\mathbf{g}$ is the metric on a two-dimensional manifold and x^1, x^2 are the coordinate functions of a chart x, and if $E = \mathbf{g}(\partial/\partial x^1, \partial/\partial x^1)$, $F = \mathbf{g}(\partial/\partial x^1, \partial/\partial x^2)$, $G = \mathbf{g}(\partial/\partial x^2, \partial/\partial x^2)$, then the line element on $\mathcal{M}$ is given by

$$ds^2 = E\ (dx^1)^2 + 2F\ dx^1 dx^2 + G\ (dx^1)^2$$

and the surface element on $\mathcal{M}$ is $dS = (EF - G^2)^{1/2}\, dx^1 dx^2$. If $H = (EF - G^2)^{1/2}$ then the Gaussian curvature K satisfies

$$KH = \frac{\partial}{\partial x^1}\left[\frac{F\ \partial E/\partial x^1 - E\ \partial G/\partial x^2}{2EH}\right] + \frac{\partial}{\partial x^2}\left[\frac{2E\ \partial F/\partial x^1 - F\ \partial E/\partial x^1 - E\ \partial E/\partial x^2}{2EH}\right]$$

This formula, found by Gauss, is independent of the position of the surface in space.

We now extend this formula to higher dimensions. Let $\mathcal{M}$ be an n-dim manifold with metric $\mathbf{g}$. Any point $p \in \mathcal{M}$ has a neighbourhood $\mathcal{O}$, diffeomorphic to $\mathbb{R}^n$, such that $\forall q \in \mathcal{O}$ there is a unique geodesic curve λ between p and q lying entirely in $\mathcal{O}$. Moreover there exits a parametrization of λ for which $\lambda(0) = p$ and $\lambda(1) = q$. Let f be a diffeomorphism that sends $\mathcal{O}$ into a neighbourhood of the zero vector in V_p and which is such that $f(q)$ is equal to the tangent vector of λ at p. If u and v are two linearly independent vectors in V_p spanning a two-dimensional subspace of V_p , $V_p(u,v)$, then the intersection of $f(\mathbf{O})$ and $V_p(u,v)$ is the image of a two-dimensional submanifold of $\mathbf{O}$ denoted by $\mathcal{M}(u,v)$. If $\mathbf{g}$ is either positive or negative definite then $\mathcal{M}(u,v)$ is, with the metric induced by $\mathbf{g}$, isometric to a smooth surface in Euclidean space and has therefore Gaussian curvature $K_{u,v}$ at p. $K_{u,v}$ is called the sectional curvature of $\mathcal{M}$ at p. If, for all points of $\mathcal{M}$, $K_{u,v}$ is constant then $\mathcal{M}$ is a Reimann manifold of constant curvature. When that constant is zero, $\mathcal{M}$ and its metric $\mathbf{g}$ are said to be flat.

In 1861 Reimann found a (0,4)-tensor field, $\mathbf{R}$, which determines and is determined by $K_{u,v}$. Though Reimann did not introduce his tensor field in the above manner, they are related by

$$K_{u,v} = \alpha^2 \, \mathbf{R}(u,v,u,v)\,\big|_p$$

where $1/\alpha$ is the area of the parallelogram spanned by u and v.† The Reimann tensor possesses the symmetries proven in Corollary 2.27; it is given by

$$g_{hl}\, R^l{}_{ijk} \;=\; g_{hl}\, (\Gamma^l{}_{ki,j} - \Gamma^l{}_{ji,k} + \Gamma^s{}_{ki}\, \Gamma^l{}_{js} + \Gamma^s{}_{ji}\, \Gamma^l{}_{ks})$$

---

† for proof see Spivak, M. (1970), A comprehensive introduction to differential geometry, Vol.II, (Publish or Perish Inc., Boston), pp. 196 *f*.

Therefore the components of R_{hijk} are determined by the metric **g**. We see that if the components of **g** are constants then the R_{hijk} are all zero and the manifold is flat, as in the case of Minkowski space-time.

Einstein decided to build his theory on the fact that space-time is generally curved by gravity, which curvature is determined by the metric. In this way did he achieve the metrically real. Hereafter space-time is a connected 4-dim Reimann manifold with Lorentz metric.† The only requirement is that **g** be approximated on small neighbourhoods by the local flat metric η. For this reason **g** must have, as η, Lorentz signature. We also require that the local η has the same value as **g** at the origin of the local Lorentz chart.

Let λ be the worldline of a particle π, parametrized by proper time τ. The variation principle

$$\delta \int d\tau = 0$$

implying that λ's length, $\int d\tau$, is constant under variations of λ, must be obeyed if π is to travel on a geodesic. This invariance is called the geodesic law of motion, a natural extension of the equivalence principle to arbitrary

† connected implies that $\mathcal{M}$ is path-connected, i.e. any two points can be joined by a curve.

gravitational fields. The line element occurring in this law is

$$d\tau^2 = g_{ij}\, dx^i\, dx^j$$

for a Lorentz chart x. In dom x, λ must satisfy

$$\frac{d^2x^i \cdot \lambda}{d\tau^2} = -\Gamma^i_{jk} \cdot \lambda \frac{dx^j \cdot \lambda}{d\tau} \frac{dx^k \cdot \lambda}{d\tau} .$$

When g is equal to η the right-hand side is zero, but since a coordinate transformation cannot reduce g to η, in a inhomogeneous gravitational flow, space-time is not flat. Therefore we cannot directly interpret physical results given by coordinate systems that chart the world into the flat $\mathbb{R}^4$. The metrically real also depends on gravitational fields determined by the 'gravitational field strengths' Γ^i_{jk} themselves derived from the 'potentials' g_{ij}. Thus, where before we required that the laws of nature be expressed independently of coordinate transformations (special covariance), we now require that these laws be expressed only in terms of covariant derivatives (general covariance) because they solely depend on the 'potentials'. Remember that this view was forced on us by the impossibility to construct inertial frames with respect to gravity and thus the inability to decide 'a priori' the structure of space-time through the significance of coordinate systems.

In this context the new relations between vectors and tensors are obtained simply by changing all η_{ij} for g_{ij} and all ∂_i for ∇_i in the old relations. For example: if x is a local Lorentz chart, the 4-acceleration is now

$$
\begin{aligned}
a^i &= u^j \nabla_j u^i \\
&= u^j \left(\frac{\partial u^i}{\partial x^j} + \Gamma^i_{kj} u^k \right) \\
&= \frac{\partial u^i}{\partial x^j} \frac{dx^j}{d\tau} + \Gamma^i_{kj} u^k u^j \\
&= \frac{d^2 x^i}{d\tau^2} + \Gamma^i_{kj} \frac{dx^k}{d\tau} \frac{dx^j}{d\tau}
\end{aligned}
$$

This example is instructive because it shows that a particle describes a geodesic if and only if its 4-acceleration is always zero.

Amazingly, the derivative is now determined by gravity, and thus it depends on the distribution of matter.

3. Einstein's field equation

Hitherto we have constructed a geometrical frame for measurements to be obtained, as close to reality as possible. The next few pages will be used to study the energetic content of these frames. Let $\Psi_{(i)}{}^{a...b}{}_{c...d}$ denote the matter fields of space-time; here (i) denotes the number of fields included in the theory. Analysing the contents of space-time will bring us closer to the determination of the geometrical structure since all the elements of the theory will be determined by the physically observable.

In the last section, we have reached the conclusion that to incorporate gravity in our model, the manifold must differ from $\mathbb{R}^4$ and the metric must not be flat. If we refer to the space-time principle of chapter I, which stated that there are as many measures of time (and length) as there are positions in space, we see that in special relativity it led to Lorentz length contraction and time dilation, and in general relativity it led to the determination of the metrically real by the gravitational forces and thus by the distribution of matter. The mathematical formulation of this relation is given in this section; it is called the Einstein field equation.

The following two postulates, being local, are common to special and general relativity; they describe the nature of the equations obeyed by the matter fields.

A) Local causality.

The equations governing $\Psi_{(i)}{}^{a..b}{}_{c...d}$ must be such that if p and q are elements of any neighbourhood $\mathcal{O}$, then it is possible to send a signal from p to q (or q to p) if and only if p and q can be joined by a timelike or a null C^1 curve which lies entirely in $\mathcal{O}$.

This postulate enables us to measure, by observation, the metric up to a conformal factor. Let u and v be respectively timelike and spacelike vectors of V_p, $p \in \mathcal{O}$. Then the equation

$$\begin{aligned} \mathbf{g}(u + \alpha v, u + \alpha v) &= \alpha^2 \mathbf{g}(v,v) + 2\alpha\, \mathbf{g}(u,v) + \mathbf{g}(u,u) \\ &= 0 \end{aligned}$$

has two real roots since $\mathbf{g}(u,u) < 0$ and $\mathbf{g}(v,v) > 0$. They are

$$\alpha_1,\ \alpha_2 = \frac{-2\,\mathbf{g}(u,v) \pm \sqrt{4\,\mathbf{g}(u,v)^2 - 4\,\mathbf{g}(u,u)\,\mathbf{g}(v,v)}}{2\,\mathbf{g}(v,v)}$$

If $\{x^i\}$ are the coordinate functions of a (Lorentz) chart at $p \in \mathcal{O}$ then all the points $q \in \mathcal{O}$ which can be reached from p by signals travelling on non-spacelike C^1 curves are those points with coordinates satisfying

$$-(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2 \leq 0$$

Thus, since the boundary of these points is the null cone N_p at p, one can determine by observation the set of points which can communicate with p and in this way find N_p in V_p. When N_p is known, α_1 and α_2 may be calculated from the fact that $u + \alpha v$ lies on the cone. But

$$\alpha_1 \alpha_2 = \mathbf{g}(u,u)/\mathbf{g}(v,v)$$

and thus the ratio of a spacelike and a timelike vector can be found by observing the null cone. Let $w, z \in N_p$; since the magnitude of a null vector can be compared with the magnitude of u and v, and since

$$\mathbf{g}(w,z) = \{ \mathbf{g}(w,w) + \mathbf{g}(z,z) - \mathbf{g}(w+z,w+z) \} / 2 ,$$

$\mathbf{g}(w,z)/\mathbf{g}(u,u)$ can be determined. Therefore the metric can be measured up to a conformal factor, by observing consequences of local causality. In practice, the conformal factor can be calculated by using a large number of similar systems (like electronic states of atoms). This can be done by isolating the systems from external fields so that the motions follow geodesics and, assuming that these motions are equally curved in space-time for similar systems, the arc-length between given events can be measured. This measure, which only depends on the metric, must be the same for each pair of successive events in similar systems; from this fact, the conformal factor may be determined at any point of space-time.

B) Local conservation of energy and momentum.

The equations governing $\Psi_{(i)}{}^{a...b}{}_{c...d}$ must be such that there exists a (2,0) symmetric tensor T^{ab}, called the energy-momentum tensor, which only depends on the the metric, the fields, and their covariant derivatives. Futhermore it must satisfy the following:

i) T^{ab} is identically zero on $\mathcal{O}$ if all $\Psi_{(i)}{}^{a...b}{}_{c...d}$ vanish on the open neighbourhood $\mathcal{O}$.

ii) The covariant divergence of T^{ab} vanishes

$$\text{i.e.} \quad T^{ab}{}_{;b} = 0 \ .$$

The principle that all fields have energy is represented by i). Note that 'only if' is omitted because T^{ab} can vanish when two equal and opposite non-zero fields cancel out.

To show that ii) represents the principle of local conservation of energy and momentum, consider the following: if the metric is flat, it will admit ten Killing vectors, namely the ten isometries of the Poincaré group. If the metric is not flat, then it will not in general admit Killing vectors; however, since g_{ab} and $\Gamma^a{}_{bc}$ are nearly zero on small neighbourhoods, $K_{(a;b)}$ will be zero to first approximation. Thus, whether the metric is flat or not, we can do as follows: let $p^a = T^{ab} K_b$. Then

$$\nabla_a p^a = \nabla_a T^{ab} K_b + T^{ab} \nabla_a K_b$$

$$\Rightarrow \quad \nabla_a p^a = T^{ab} \nabla_a K_b \ .$$

But T^{ab} is symmetric and by Theorem 2.31 $K_{(a;b)} = 0$, thus we have

$$T^{ab} \nabla_a K_b = T^{ba} \nabla_b K_a$$

$$\Rightarrow \quad \nabla_a p^a = 0 \ .$$

Hence if $\mathfrak{u}$ is a compact orientable region with boundary $\partial\mathfrak{u}$, Gauss' Theorem implies that

$$\int_{\partial\mathfrak{u}} p^b \, d\sigma_b = \int_{\mathfrak{u}} \nabla_b p^b \, d\upsilon = 0$$

Since p^0 can be interpreted as the flow of energy and p^α (α = 1,2,3) as the flow of linear momentum through $\mathfrak{u}$, we may conclude from this result that the total flux of energy-momentum over a closed surface is zero. Therefore, even on small neighbourhoods, one has conservation of energy, momentum and angular momentum. Provided that the energy density of matter be non-negative ($T^{ab} v^a v^b \geq 0$, where v^a is the observer's 4-velocity), this result can be interpreted as Galileo's principle that all bodies fall at equal speeds. In Einsteinian terms, radiation must gravitate, a consequence of the equivalence principle. (In 1912, Einstein published a paper in which he shows that if the inertial mass of a body increases with the energy it contains, then its gravitational mass must also increase; otherwise a given body could fall at different accelerations depending on the energy it contains. This would make it possible to build a perpetual machine (Einstein,1924)).

To motivate Einstein's field equation, we now derive a characterization of the Reimann curvature. Let $\lambda_s(t)$ denote a smooth one-parameter family of s geodesics parametrized by t. The map $(s,t) \rightarrow \lambda_s(t)$ is smooth, one to one, and has a smooth inverse; let $\mathcal{M}(s,t)$ denote the two-dimensional submanifold spanned by the curves $\lambda_s(t)$. If we take s and t to be the coordinates of $\mathcal{M}(s,t)$ then $T^a = (\partial/\partial t)^a$ is tangent to $\lambda_s(t)$ and $X^a = (\partial/\partial s)^a$ represents the displacement from the actual geodesic to an infinitely close geodesic of the family; X^a is called the deviation vector. Thus

$$T^a \nabla_a T^b = 0 \qquad \text{and} \qquad T^b \nabla_b X^a = X^b \nabla_b T^a$$

The vector $v^a = T^b \nabla_b X^a$ represents the rate of change of the displacement to a nearby geodesic and may be interpreted as the relative velocity of an infinitely close geodesic. In the same way

$$A^a = T^b \nabla_b v^a = T^b \nabla_b (T^c \nabla_c X^a)$$

may be interpreted as the relative acceleration of an infinitely close geodesic. Hence we have

$$\begin{aligned} A^a &= T^b \nabla_b (T^c \nabla_c X^a) \\ &= T^b \nabla_b (X^c \nabla_c T^a) \\ &= T^b \nabla_b X^c (\nabla_c T^a) + T^b X^c \nabla_b \nabla_c T^a \end{aligned}$$

and by Theorem 2.26

$$\begin{aligned} &= X^b \nabla_b T^c (\nabla_c T^a) + T^b X^c \nabla_c \nabla_b T^a - T^b X^c R_{bcd}{}^a T^d \\ &= X^b \nabla_b (T^c \nabla_c T^a) - R_{bcd}{}^a X^c T^b T^d \end{aligned}$$

$$\Rightarrow \quad A^a = - R_{bcd}{}^a X^c T^b T^d$$

This result is called the geodesic deviation equation; it shows that the relative acceleration is zero if and only if space-time is flat. We also get, as a bonus, a characterization of the curvature.

In Newtonian theory, the tidal acceleration of two nearby particles is given by $-(\mathbf{x} \cdot \nabla)\nabla\phi$, where ϕ is the gravitational potential, $\mathbf{x}$ is the separation vector of the particles, and ∇ is the Laplacian operator. We can interpret the above characterization as follows: if $\lambda_s(t)$ are the geodesics of nearby particles, v^a their 4-velocity and x^a their deviation vector, then the tidal acceleration of two nearby particles is $-R_{bcd}{}^a x^c v^b v^d$. This suggests that in the Newtonian limit

$$R_{bcd}{}^a v^b v^d \longrightarrow \partial_c \partial^a \phi .$$

On the other hand, Poisson's equation being $\nabla^2 \phi = 4\pi G \rho$, we would like to think that in the limit

$$R_{bcd}{}^a v^b v^d \longrightarrow 4\pi G \rho .$$

Again, in this Newtonian limit, we should expect that $T_{ab} v^a v^b \rightarrow \rho$, which is the case. Therefore we could be tempted to write

$$R_{acd}{}^a v^c v^d = 4\pi G \, T_{cd} v^c v^d ,$$

but the last arrow is not an equal sign, and the previous one is only one way. Nevertheless this development suggests the field equation

$$R_{ab} = 4\pi G \, T_{ab} .$$

Indeed, Einstein first postulated this equation but realized that on the right-hand side $\nabla^a T_{ab} = 0$, while on the left we have by Corollary 2.27 4)

$$g^b{}_f \, g^{df} \, g^a{}_e \, (\nabla_a R_{bcd}{}^e - \nabla_c R_{abd}{}^e + \nabla_b R_{cad}{}^e) = 0$$

$$\Rightarrow \quad g^b{}_f \, g^{df} \, (\nabla_a R_{bcd}{}^a - \nabla_c R_{bd} + \nabla_b R_{cd}) = 0$$

$$\Rightarrow \quad g^b{}_f \, (\nabla_a R_{bc}{}^{fa} - \nabla_c R_b{}^f + \nabla_b R_c{}^f) = 0$$

$$\Rightarrow \quad \nabla_a R_c{}^a - \nabla_c R + \nabla_b R_c{}^b = 0$$

$$\Rightarrow \quad 2\nabla^a R_{ab} - \nabla^a R \, g_{ab} = 0$$

$$\text{or} \quad \nabla^a (R_{ab} - {}^1/_2 \, R \, g_{ab}) = 0$$

and hence $\nabla^a R_{ab} = 0$ if and only if $\nabla^a R = 0$ i.e. $R = 0$ and therefore $\nabla^a R_{ab} = 0$ if and only if $T = T^a{}_a$ is constant throughout the universe, which is impossible.

However, while this calculation forces us to reject the proposed field equation, it suggests that

$$R_{ab} - {}^1/_2 \, R \, g_{ab} = 8\pi G \, T_{ab} \, .^\dagger$$

This equation, known as Einstein field equation, not only restores the initial conflict between the Bianchi identity and the local conservation of energy and momentum, but also restores the correspondence which motivated a certain equality between the Ricci tensor and the energy-momentum tensor.

---

† $8\pi G/c^4$ is known as Einstein's constant ($= 2.073 \times 10^{-48}$ sec^2 cm^{-1} g^{-1})

By taking the trace of the field equation we find

$$R = -8\pi G\, T$$

and thus

$$R_{ab} = 8\pi G\,(T_{ab} - {}^1/_2\, T\, g_{ab}) \;.$$

To show that the field equation reduces to Poisson's equation in the Newtonian limit, we consider the weak field approximation $g_{ab} = \eta_{ab} + h_{ab}$, where the h_{ab} are so small that products between themselves are neglected (Rindler, 1977). This linear approximation will also point towards the Schwarzschild exact solution to the field equation; although this solution can be found without using the field equation, this latter method requires, after many calculations, that the Newtonian potential be 'recognized' from the particular form of an equation. This will not be necessary here. One should not be misled by the relative ease with which we shall solve the field equation for weak fields because in general it is not possible to specify T_{ab} and then find g_{ab} ; in most cases T_{ab} is expressed in terms of g_{ab} , and one must solve for the metric and the matter distribution simultaneously. Our task will be even more simplified by the fact that only the principal parts of g_{ab} , namely η_{ab} , will be used to raise and lower indices of quantities containing h; this principal part is known even though the metric has not been found yet.

In this context, the affine connection is given by

$$\Gamma^a{}_{bc} = {}^1/_2\, \eta^{ad} (\partial_b h_{cd} + \partial_c h_{bd} - \partial_d h_{bc}) + \text{order of } h^2$$
$$= {}^1/_2\, (\partial_b h^a{}_c + \partial_c h^a{}_b - \partial_a h_{bc}) \; .$$

Thus, the Ricci tensor is

$$R_{ab} = \eta^{cd} R_{abcd} + \text{order of } h^2$$
$$= {}^1/_2\, (\eta^{ab}\, \partial_a \partial_b h_{ab} + \partial_a \partial_b h - \partial_b \partial_c h^c{}_a - \partial_a \partial_c h^c{}_b)$$

where $h = \eta^{ab} h_{ab}$. Let $\Diamond = \eta^{ab}\, \partial_a \partial_b$ and let $X_{ab} = h_{ab} - {}^1/_2\, \eta_{ab} h$; it is not hard to see that

$$2\, R_{ab} = \Diamond\, h_{ab} - \partial_b \partial_c X^c{}_a - \partial_a \partial_c X^c{}_b \; .$$

By the coordinate transformation

$$\tilde{x}^a = x^a + \xi^a(x) \; ,$$

we intend to get rid of the terms containing X. By the transformation law for tensors we have

$$g_{ab} = \tilde{g}_{cd} \left[\frac{\partial \tilde{x}^c}{\partial x^a}\right] \left[\frac{\partial \tilde{x}^d}{\partial x^b}\right]$$

$$= \tilde{g}_{cd}\,(\delta^c{}_a + \partial_a \xi^c)(\delta^d{}_b + \partial_b \xi^d)$$

$$\Rightarrow \quad \tilde{h}_{ab} = h_{ab} - \partial_a \xi_b - \partial_b \xi_a \; .$$

$$\Rightarrow \quad \partial_b \tilde{X}^{ab} = 0$$

$$\Rightarrow \quad \tilde{R}_{ab} = \frac{1}{2}\,\tilde{\Diamond}\,\tilde{h}_{ab}$$

and omitting the tilde we get $R_{ab} = {}^1\!/_2 \Diamond h_{ab}$; thus,

$$R = {}^1\!/_2 \Diamond h$$

$$\Rightarrow \quad R_{ab} - {}^1\!/_2\, \eta_{ab} R = {}^1\!/_2 \Diamond (h_{ab} - {}^1\!/_2\, \eta_{ab} h) = {}^1\!/_2 \Diamond X_{ab}$$

and Einstein's equation takes the form

$$\Diamond X_{ab} = 16\pi G\, T_{ab}$$

$$\text{or} \quad \Diamond h_{ab} = 16\pi G\,(T_{ab} - {}^1\!/_2\, \eta_{ab} T) \; .$$

If the gravitational sources have negligible speed and negligible stress then T_{ab} is the diagonal matrix $(\rho, 0, 0, 0)$

$$\Rightarrow \quad T_{ab} - {}^1\!/_2\, \eta_{ab} T = \rho / 2 \quad \text{for } a = b, \; (=0 \text{ , for } a \neq b).$$

$$\Rightarrow \quad \Diamond h_{aa} = 16\pi G \rho \quad , \quad a = 0,1,2,3.$$

But since the sources move slowly, the field will also change slowly, and thus $\Diamond h_{ab} \approx -\nabla^2 h_{ab}$. The field equation becomes

$$\nabla^2 \left({}^1/_2 \, h_{00} \right) = 4\pi G \rho ,$$

which is Poisson's equation with ${}^1/_2 \, h_{00} = \phi$.

Furthermore $g_{ab} = \eta_{ab} + h_{ab}$ implies that .

$$g_{00} = 1 + 2\phi$$

$$\text{and} \qquad g_{\alpha\alpha} = 1 - 2\phi , \qquad \alpha = 1,2,3$$

$$\text{otherwise} \quad g_{ab} = 0.$$

This means that for $\phi = -GM/r$,

$$ds^2 = \left(1 - \frac{2GM}{r}\right) dt^2 - \left(1 + \frac{2GM}{r}\right) \left(dx^2 + dy^2 + dz^2\right)$$

Note that since this metric is for weak ϕ, $(1 + 2\phi) \approx (1 - 2\phi)^{-1}$, and thus we see that this latter metric is in accordance with the Schwarzschild solution (in spherical coordinates):

$$ds^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 dt^2 - \left(1 - \frac{2GM}{rc^2}\right)^{-1} dr^2 - r^2 \left(d\theta^2 + \sin^2\theta \, d\varphi^2\right)$$

There exist a few other solutions to Einstein's field equation such as De Sitter space-time, Robertson-Walker spaces, Reissner-Nordström solutions, the Kerr metric, Gödel's universe and others; these are all exact solutions. Since Schwarzschild's solution was used to calculate the perhilion advance of Mercury and the bending of a light ray around the sun, it is the best known; the others remain theoretical.

A last fact of interest should be pointed out: suppose that, due to the absence of matter, $T_{ab} = 0$. Then

$$
\begin{aligned}
& R_{ab} - {}^{1}/_{2}\, g_{ab} R = 0 \\
\Rightarrow\quad & R - {}^{1}/_{2}\, g^{ab} g_{ab} R = 0 \\
\Rightarrow\quad & R = 0 \\
\Rightarrow\quad & R_{ab} = 0 \quad .
\end{aligned}
$$

But this does not imply that $R_{abcd} = 0$; in fact when $T_{ab} = 0$,

$$R_{abcd} = C_{abcd}$$

the Weyl tensor. There are still $n^2(n^2-1)/12 - n(n+1)/2 = 10$ components which describe the gravitational field in free space (even if $\phi=0$). These ten equations can be characterized by using the fact that the Weyl tensor is conformally invariant.

To recapitulate we could say: space-time is a 4-dimensional C^∞ manifold with a non-flat Lorentz metric which is Minkowskian on small neighbourhoods; the curvature of this metric is related to the distribution of matter in space-time through Einstein's field equation, but is not identically zero in absence of matter; the matter fields, represented in this equation by the energy-momentum tensor, satisfy local causality and local conservation of energy and momentum.

4. Singularities

In 1799 Laplace published a paper in which he proves that the attractive force of a luminous body, of the same density of the earth and whose diameter is 250 times that of our sun, would be so large that none of its light would reach us. Consequently, the largest bodies of the universe could remain invisible to us. Today, such bodies are called black holes. Since these heavenly objects are not visible to the eye, their presence was not conjectured until this century with the event of X-ray telescopes and spectral analysis. One of the first black holes to be 'observed', named Cygnus X-1, was detected as a result of the following fact: when gases approach a certain critical radius (thought to be $r_g = 2GM/c^2$, the Schwarzschild radius), the magnetic field of the black hole being stronger, some particles are accelerated towards the centre; the falling matter emits radiation that can be detected by X-ray telescopes. In the case of Cygnus X-1, two facts were determinant as to the believe of the presence of a black hole: first, the source of X-rays Cygnus X-1 is part of a binary system of period 5 or 6 days which insures the increase of mass due to falling matter (a process called accretion), process that determines most certainly the X-ray \radiation. Secondly, unlike radiation emitted from pulsars such as Cen X-3 and Her X-1, the radiation emitted from Cygnus X-1 has a rapid fluctuation with no determined period. Its mass, which was known to be approximatively of 10 to 20 solar masses, has recently been found to be of at least 9.5 solar masses (B.Paczynski, 1974). This number is in agreement with the predictions made by the theory of stellar evolution.

Black holes can also evolve from massive stars like quasars and galactic centres.† In fact, in the seventies, evidence was found to support the hypothesis that there is a dead quasar in the centre of our galaxy (i.e. a quasar that has evolved to a black hole). The activity of such galactic nuclei is also associated to 'accrétion'. In 1978 two independent researches, (W.L.W. Sargent et al., 1978) and (P.J. Young et al., 1978), have come to the conclusion that there is a supermassive black hole (5×10^{9} solar masses) in the centre of the galaxy M87. The main idea of these two papers is that an unexpected spike of luminosity has been detected within the core radius of M87, the surroundings being of much lower intensity. As for your galaxy, J.Weber has built what should be called the first gravitational telescope (J. Weber, 1969 & 1970). It was used to detect a strong

† the evolution of stars will not be discussed here but can be founa in most text books considering the topic of black holes. For example see (Ginzburg, 1976).

gravitational force of the order of 10^{52} ergs$\cdot s^{-1}$ (and more) coming from the centre of our galaxy (the Reimann curvature tensor was used to find the directivity pattern). As the energy corresponding to our sun at rest is $Mc^2 \approx 10^{54}$ ergs, the centre of the galaxy would lose about 10^4 solar masses per year under the only effect of gravitational radiation. This is in contradiction with the fact that the core of the Milky Way has a mass of about 10^{11} solar masses, which means that it would take 10^7 years (approximatively) for the nucleus to lose the totality of its mass; this is impossible since the age of the Milky Way is known to be greater than 10^9 years. If Weber's result is true, the most possible explanation would be that of a black hole swallowing mass to live (accrétion).

Black holes are certainly the ~~most~~ studied singularities, but they are not the only ones. In fact the zero-th second of the expansion of Universe is called the initial singularity. This expansion is seen as a gravitational collapse but with time reversed (we do not mean that time is reversible). The isotropic background microwave radiation of 2.7 Kelvin and the fact that most galaxies are red shifted† are two examples which are thought to support the hypothesis of the initial expansion.

---

† for example Virgo travels away from us at 1200 km/s and Hydra at 61000 km/s.

It would be reasonable to define a space-time singularity as a point of the manifold where the metric is undefined or is not suitably differentiable. For example at $\phi = {}^1/_2$ in $(1 - 2\phi)^{-1}$ at the Schwarzschild radius r_g; another example arises in the Kerr solution (R.P. Kerr, 1963). The Kerr metric has the form (in Boyer-Lindquist coordinates (t, r, θ, φ)):

$$ds^2 = \rho^2\left(\frac{dr^2}{\Delta} + d\theta^2\right) + (r^2 + a^2)\sin^2\theta\, d\varphi^2 - dt^2 + \frac{2mr}{\rho^2}(a^2\sin^2\theta\, d\varphi - dt)^2$$

where $\rho^2 = r^2 + a^2\cos^2\theta$ and $\Delta = r^2 - 2mr + a^2$. m and a are constants: m represents the mass and ma the angular momentum; when $a = 0$ we recover the Schwarzschild solution. At $r = 0$ the singularity is not a point but a ring as can be seen by transforming to Kerr-Schild coordinates (x, y, z, t'):

$$x + iy = (r + ia)\sin\theta\, e^{i\int (d\varphi + a\Delta^{-1}dr)},$$

$$z = r\cos\theta, \quad t' = \int \{dt + (r^2 + a^2)\Delta^{-1}dr\} - r.$$

Then

$$ds^2 = dx^2 + dy^2 + dz^2 + dt'^2$$

$$+ \frac{2mr^3}{r^4 + a^2z^2}\left[\frac{r(x\,dx + y\,dy) - a(x\,dy - y\,dx)}{r^2 + a^2} + \frac{z\,dz}{r} + dt'\right]^2.$$

The ring $x^2 + y^2 = a^2$, $z = 0$ is a curvature singularity since $R_{abcd}R^{abcd}$ vanishes there.

It is of interest to note that while the Schwarzschild metric yields static black holes, the Kerr (-Newman) solution yields rotating (charged) black holes.

The problem with our definition of singularity is that one could simply cut out such points and say that space-time is constituted of the remaining points of the manifold; this would seem appropriate since the equations of physics break down at these points and no measurements can possibly be taken. Thus, we ensure that no singular or regular points be omitted by requiring that $(\mathcal{M}, \mathbf{g})$ cannot be extended. The search for singularities now becomes a question about space-time incompleteness, in some sense.

Let $\mathcal{M}$ be an n-manifold and $\lambda : I \to \mathcal{M}$ a curve with domain I in $\mathbb{R}$. λ is said to be extensible if there exists a curve $\lambda' : I' \to \mathcal{M}$ such that $I' \neq I$, $I' \supset I$ and λ is the restriction of λ' to I. λ is incomplete if it is inextensible and I is a finite interval. Let Γ be the linear connection of $\mathcal{M}$. Γ is said to be incomplete (and $(\mathcal{M}, \Gamma)$ to be geodesically incomplete) if there exists an incomplete geodesic in $\mathcal{M}$. If $\mathbf{g}$ is a positive definite metric, we can define the distance function $d(p,q) : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ as the greatest lower bound of the length of curves joining p and q. With this function, $\mathcal{M}$

is a metric space in the topological sense. $\mathcal{M}$ is then said to be metrically incomplete if and only if it is geodesically incomplete. (An alternative formulation is: $(\mathcal{M}, g)$ is metrically incomplete if every Cauchy sequence in $\mathcal{M}$ converges to a point in $\mathcal{M}$ with respect to the distance function d).

Contrarily to the positive definite metric used above, a Lorentz metric does not define a topological space. Thus we are left with geodesical incompleteness: timelike, spacelike and null. Timelike geodesic incompleteness can be interpreted as the existence of freely moving observers (or particles) whose histories only start a finite time ago, or end in a finite time.† The same interpretation holds for null geodesic incompleteness with the difference that the particle has zero rest-mass (such as a photon). Because nothing travels on spacelike geodesics, spacelike geodesic incompleteness is not clear. It is now possible to state that a singularity-free space-time is a space-time in which timelike and null geodesics are complete, i.e. defined on the entire field of real numbers.

In the late sixties and early seventies Hawking and Penrose proved a series of theorems asserting that a relativistic space-time is always geodesically incomplete

---

† such as particles created by the initial expansion of the Universe or swallowed by a black hole.

provided that some general conditions be fulfilled. Before stating these theorems we must introduce some concepts about the causal structure of space-time.

Let $(\mathcal{M}, g)$ be a space-time. For each $p \in \mathcal{M}$, V_p is isomorphic to Minkowski space-time. The light cone at p can be seen as the light cone passing through the origin of V_p, and thus as a subset of V_p. A timelike or null vector element of the future half of the cone at p is called future directed. A C^1 curve λ is said to be a future directed timelike curve if the tangent at each p in the range of λ is a future directed timelike vector. If these tangents are either future directed timelike or null vectors then λ is called a future directed causal curve. The chronological future of $p \in \mathcal{M}$ is defined as the set of all points on $\mathcal{M}$ which can be reached by material particles starting at p. That is

$$I^+(p) = \{ q \in \mathcal{M} \mid \text{there exists } \lambda(t) \text{ future directed timelike with } \lambda(0) = p \text{ and } \lambda(1) = q \}.$$

And for any subset S of $\mathcal{M}$ we define

$$I^+(S) = \bigcup_{p \in S} I^+(p)$$

The causal future of p is defined by replacing timelike curves with causal curves, i.e. we add curves of non-material particles; it is denoted by $J^+(p)$. All these definitions also apply to the past half of the cone, yielding the sets $I^-(p)$ and $J^-(p)$.

If for all $p \in \mathcal{M}$ and every neighbourhood $\mathcal{O}$ of p there exists a neighbourhood $\mathcal{U}$ of p contained in $\mathcal{O}$ such that no causal curve intersects $\mathcal{U}$ more than once, then the space-time $(\mathcal{M},\mathbf{g})$ is said to be strongly causal. It is generally thought that there does not exist space-times with closed causal curves.† If a space-time is not strongly causal at some point p, then there will exist causal curves arbitrarily close to intersecting themselves near p, and thus, by small changes of $\mathbf{g}$, closed curves could be produced.

A subset S of $\mathcal{M}$ is said to be achronal if there does not exist $p,q \in S$ such that $q \in I^+(p)$; this ensures that $I^+(S) \cap S = \emptyset$, the empty set.

Let S be a closed achronal set. The future domain of dependence of S is defined by

$$D^+(S) = \{\, p \in \mathcal{M} \mid \text{every past inextensible causal curve passing through p intersects S} \,\}.$$

The past domain of dependence $D^-(S)$ is defined by replacing 'past' with 'future'. The domain of dependence is

---

† $\lambda(t)$ is closed if there exists t and t', $t \neq t'$, such that $\lambda(t) = \lambda(t')$.

then

$$D(S) = D^+(S) \cup D^-(S).$$

If light is the fastest messenger, then any signal sent to $p \in D^+(S)$ must have been 'registered' in S. Hence if initial conditions are given on S, then the events at $p \in D^+(S)$ should be predictable. Similarly the knowledge of conditions on S should enable us to determine all conditions on the set of events D(S).

A Cauchy surface is a closed achronal set S for which $D(S) = \mathcal{M}$. In a space-time possessing a Cauchy surface, the entire past and future can be predicted from the conditions given at the interval of time in which the surface is defined. It is believed that all relativistic space-times have a Cauchy surface.

A space-time $(\mathcal{M}, g)$ is generic if it meets the following condition: for all timelike and null geodesics λ in $\mathcal{M}$ there exists a point $p \in \mathcal{M}$ at which the tangent v to λ at p satisfies

$$v_{[a}R_{b]cd[e}v_{f]}v^c v^d \neq 0.$$

This ensures that every timelike or null geodesic includes one event at which the gravitational curvature is 'effective'.

$(\mathcal{M}, g)$ is said to satisfy the weak or the strong convergence condition if $R_{ab}v^a v^b \geq 0$ for every null or timelike vector v respectively. These conditions can be seen as the assertion that matter attracts matter and radiation since, by Einstein's equation,

$$R_{ab}v^a v^b \geq 0 \quad \Rightarrow \quad T_{ab}v^a v^b \geq {}^1/_2\, T\, v^a v_a \,.$$

If for example

$$T_{ab} = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & p_1 & 0 & 0 \\ 0 & 0 & p_2 & 0 \\ 0 & 0 & 0 & p_3 \end{bmatrix}$$

where ρ is the energy density and p_α is the pressure ($\alpha =$ 1,2,3), then the strong convergence condition is just

$$\rho + p_\alpha \geq 0 \quad \text{and} \quad \rho + \Sigma_\alpha p_\alpha \geq 0 \,.$$

A trapped surface in $(\mathcal{M}, g)$ is a compact spacelike two-dimensional C^∞ surface which has the following property: incoming and outgoing future directed null geodesics that meet S orthogonally converge locally in S. This may be regarded as the formal statement that a black hole is black, i.e. a spacial region which absorbs nearby radiation and pulls back radiation originating from the inside. In this way all spheres inside a black hole are trapped surfaces. It has been shown that if the initial conditions of a stellar gravitational collapse are those of a spherical collapse, then trapped surfaces will form.

We can now state the main singularity theorems:† in these theorems all $(\mathcal{M},g)$ are relativistic space-times satisfying Eintein field equation together with the other properties stated in the last sentence of section 3.

I. Penrose, 1965: $(\mathcal{M},g)$ contains incomplete null geodesics if:

i) it satisfies the weak convergence condition,

ii) there is a non-compact Cauchy surface in $\mathcal{M}$,

iii) there is a closed trapped surface in $\mathcal{M}$.

II. Hawking, 1967: $(\mathcal{M},g)$ contains incomplete timelike geodesics if:

i) it satisfies the strong convergence condition,

ii) there exists a compact spacelike 3-surface H without edge,

iii) the unit normal vectors to H are everywhere converging (or everywhere diverging) on H.

---

† the proofs can be found in (Hawking & Ellis, 1973) or (R.M. Wald, 1984).

III. Hawking, 1967: $(\mathcal{M}, g)$ contains incomplete non-spacelike geodesics if:

i) it satisfies the strong convergence condition,

ii) the strong causality condition holds,

iii) there is some $p \in \mathcal{M}$ such that all future directed (or past directed) timelike geodesics through p are focused (i.e. expanded and contracted) by the curvature and start reconverging in a compact region in the future (or in the past).

IV. Hawking & Penrose, 1970: $(\mathcal{M}, g)$ contains incomplete non- spacelike geodesics if:

i) it satisfies the strong convergence condition,

ii) it is generic,

iii) it does not have closed timelike curves,

iv) there exists at least one of the following:

a) a compact achronal set without edge,

b) a closed trapped surface,

c) a worldpoint p such that all future directed (or past directed) null geodesics through p are focused by curvature and start reconverging in the future (or in the past).

In these theorems, the energy conditions will hold provided, as we said, that the energy density of matter be positive for all observers, a fairly reasonable condition. Since condition iii) of Theorem I is also realistic, the only way that this theorem could fail to provide singularities would be if there does not exist a Cauchy surface; this is the weakness of Penrose's Theorem. Thus it seems that Theorem I states that a collapsing star will end up either in a singularity or in creating a Cauchy horizon (i.e. the future boundary of $D^{+}(S)$ for some closed set S). However, this conclusion does not satisfy our 'quest' for singularities; we need a theorem that does not assume the existence of a Cauchy surface. Theorems III and IV have the most general conditions, as they apply to a number of physical situations. Nevertheless it can be that closed timelike curves occur (the past is reached again via the future) instead of a singularity. Theorem II asserts that this violation of the causality conditions does not prevent singularities from occurring provided that the universe be spatially closed (ii) and contracting or expanding (iii).

The success to prove the existence of space-time singularities by using the notion of geodesic incompleteness is not surprising when we try to imagine a closed geodesically complete universe, idea that was rejected from the early days of General Relativity. As R.Torretti phrases nicely and truly:

"After all, barring metaphysical insight or divine revelation I do not see how we could ever learn that spacetime is geodesically complete - and all the more reputable sources of supernatural information decidedly point to the opposite conclusion. Indeed the venerable vision of man as a microcosm, and hence of the universe as Man Writ Large, can only be upheld within the framework of General Relativity if spacetime contains incomplete timelike geodesics along which, so to speak, time runs out." †

† (R. Torretti, 1983)

AFTERTHOUGHT

The importance of the axioms, principles and postulates of a theory is now transparent to us since these have been shown to direct the mathematical formulation of the structural approach to space-time and hence to the understanding of Universe. Thus, natural philosophy (experimental and scientific) plays a determining role in the paths taken by 'exact' sciences. The relevancy of this role will increase as we search for a micro-mechanical explanation to gravity since such an understanding will inevitably constitute a step towards binding the known forces of nature into a unified field theory which will encompass the micro-universe of matter and the macro-universes.

To conclude, here are a few questions which I believe to be of some importance to achieve this understanding: do objects really fall at the same speed or is it that the gravitational field of the earth is too weak to make a significant difference ? How fast does the effect of gravity travel ? Is it really the exact same speed as light ? Is there an aethereal medium for light ? If there is, how can we relate it with gravity ? Does antigravity exist ? The density of a proton is of the order of one billion metric tons (10^{15} g.) per cubic centimetre; what is the density of a quark ? Is matter more and more dense as we approach the micro-world ? Why (not how) does matter attract matter ? When singularities occur through the formation of black holes, is matter so dense and so fast revolving on itself that eventually a supersmall hyperdense (> 10^{60} g/cm^3) 'object' is obtained ? Is matter made of the organized agglomeration of 'point' singularities forming an inpenetrable field of energy (to our senses) ?

BIBLIOGRAPHY

Borel, E. (1960), "Space & Time", (Dover Publications Inc., New York).

Dobbs, B.J.T. (1975), "The foundation of Newton's alchemy", (Cambridge University Press, Cambridge).

Einstein, A. (1950), "Out of my later years", (Philosophical Library, New York).

Einstein, A. (1974), "The meaning of relativity", fifth edition, (Princeton University Press, New Jersey).

Einstein, A., H.A. Lorentz, H. Minkowski and H. Weyl (1924), "The principle of relativity", (Dover Publications Inc., New York).

Fairchild, W.W. and C.I. Tulcea (1971), "Topology", (W.B. Saunders Comp., Philadelphia).

Ginzburg, V. (1976), "Sur la physique et l'astrophysique", (Editions Mir, Moscou).

Golovanov, L. (1981), "L'admirable harmonie de la nature", (Editions Mir, Moscou).

Hawking, S.W. and G.F.R. Ellis (1973), "The large scale structure of space-time", (Cambridge University Press, Cambridge).

Kerr, R.P. (1963), "Gravitational field of spinning mass as an example of algebraically special metrics", Phys. Rev.

Letters 11, 237-8.

Paczynski, B. (1974), "Mass of Cygnus X-1", Astron. & Astrophys. 34, 161-2.

Piaget, J. (1966), "La naissance de l'intelligence chez l'enfant", (Delachaux et Niestlé, Paris).

Poincaré, H. (1958), "The value of science", (Dover Publication Inc., New York).

Rindler, W. (1977), "Essential relativity", second edition, (Springer-Verlag, New York).

Sargent, W.L.W. et al. (1978), "Dynamical evidence for a central mass concentration in the galaxy M87", Astrophys. J. 221, 731-44.

Spivak, M. (1965), "Calculus on manifolds", (W.A. Benjamin Inc., New York).

Spivak, M. (1970), "A comprehensive introduction to differential geometry", Vol. I, (Publish or Perish Inc., Boston).

Torretti, R. (1983), "Relativity and geometry", (Pergamon Press, Oxford).

Touchinski, G. (1966), "Le cosmos et les rythmes de la nature sur la terre", (Editions Mir, Moscou).

Wald, R.M. (1984), "General relativity", (University of Chicago Press, Chicago).

Weber, J. (1969), "Evidence for discovery of gravitational radiation", Phys. Rev. Letters 22, 1320-4.

Weber, J. (1970), "Anisotropy and polarization in the gravitational raditation experiments", Phys. Rev. Letters 25, 180-4.

Young, P.J. et al (1978), "Evidence for a supermassive object in the nucleus of the galaxy M87 from SIT and CCD area photometry", Astrophys. J. 221, 721-30.