308

CAUSAL SPACES AND THE APPLICATION OF CRITICAL POINT THEORY TO GENERAL RELATIVITY

GEORGE CORDOULIS

SUMMER 1981

A thesis submitted as a partial requirement for the degree of Doctor of Philosophy in Mathematical Physics of the University of London

CAUSAL SPACES AND THE APPLICATION OF CRITICAL POINT THEORY TO GENERAL RELATIVITY

ACKNOWLEDGEMENTS

This is one of those cases where the hackneyed phrase "if it were not for the efforts of..., this work would not have been possible" applies without qualification to my family who supported me financially throughout the years of my studies and my tutor Dr. Patrick Dolan who not only thought that this work was possible but also worthwhile.

This seems to be the appropriate place to express my thanks to the indefatigable Librarian of the Mathematics Department of Imperial College Ms. J. Pindelska for her kindness and her expert help.

Thanks are due to Ms's J. Wilkinson, D. Abeysekera and S. Evans who performed the unthankful job of typing the manuscript.

Other acknowledgements to friends are too deep and intensive to tabulate.

ABBREVIATIONS

DFN., THR., PR., stand for definition, theorem and proposition in an obvious fashion

The quantifier "there exist" is denoted by $\exists$ and the quantifier "for each" is denoted by $\forall$

"iff" stands for "if and only if", where simple implication is denoted by $\Rightarrow$

$\square$ marks the end of a proof, whenever used.

$\emptyset$ is the void set and $\in$ means "belongs"

$\ni$ stands for "such that"

The strike / (e.g. $\nexists$, $\notin$, $\not\to$) denotes the negation of a proposition or a property.

The dash - over a set denotes its closure and in front of a set its complement .

Προλεγόμενα

This work has actually been written during two entirely different and indeed separate, periods of study. PART I was completed before 1975 and as a result we feel we should comment on some recent developments, however peripheral, as well as on the foundational aspects of the theory of causal spaces from the vantage point of knowing how far the theory has developed. PART II was originally motivated from the idea that instead of studying singularities of a rather exceptional nature, such as causal anomalies, incomplete geodesics and in general, sets of boundary points of our manifold spacetime or its various tangent bundles, one should try to understand the nature of a rather more prosaic type of anomalies such as the singularities, in the context of analysis, of well defined and general mappings on the manifold itself, that is, their topology and their genericity or their stable morphologies to use a fanciful term. Such singularities are obviously the caustics, the singularities of hypersurfaces and the asymptotic behaviour of various expressions defined on a spacetime.

The introductory sections of chapter 1 and chapter 2 contain our main motivation for their respective material and a description of what is being done; need only add the following remarks which are more of a postscript nature.

We now know a little more about the topologies on Minkowski space, whose group of homeomorphisms is either the augmented Poincaré group (full Lorentz plus translations plus dilations) or the augmented orthochronous Poincaré group, i.e. the group of causal automorphisms of the Minkowski spacetime (DFN.4 and PR.6, 1.2). First, the order topologies W_+ and W_- (PR.13, 1.2) have been proved to be superconnected (i.e. every open set is connected), pathwise connected, not arcwise connected (i.e. no continuous map from the unit interval into the Minkowski space with the order topology can be one to one), but they are simply connected; they are not comparable to the Euclidean topology and although they are coarse they are not minimal, in the sense that there exist strictly weaker topologies with the same homeomorphism group (S. NANDA, 4 or K. PANDA, 1). Second, there seem to be good reasons to include the null lines in any definition of topological nature, as well as to ignore them (see the remarks preceding (DFN.5, 1.2)), as trajectories of photons, i.e. continuous images of the unit interval; this is because, if f is a continuous chronology preserving map of the unit interval I into Minkowski space with the (Zeeman) fine topology, then f(I) is a connected union of time-like intervals. (E. C. ZEEMAN, 2). This motivates the definition of the finest topology with respect to which the induced topology on every time-like line and light-like line is one-dimensional Euclidean and the induced topology on every space-like hyperplane is three-dimensional Euclidean; such a topology is strictly finer than the Euclidean topology (and hence Hausdorff) and its group of homeomorphisms is the augmented Poincaré group. (S. NANDA, 6). S. Nanda has also considered some weaker versions of E. C. Zeeman's theorem (PR.6, 1.2) i.e. considering coarser topologies which still induce the Euclidean topology on time-like lines (t-topology) and space-like hyperplanes (s-topology) respectively (S. NANDA, 2,5); t- and s-topologies are neither normal, nor locally compact, still finer than the Euclidean topology (hence Hausdorff) and by definition first countable; their antispaces are therefore compact and non-Hausdorff but with the same group of homeomorphisms (S. NANDA, 3 or K. PANDA, 1); their supremum and their antitopological infimum do not however have as their group of homeomorphisms a causal group. (S. NANDA, 7). The answer to the question as to whether there exist maximal and minimal elements in the set of all causal topologies (i.e. topologies on the Minkowski space whose group of homeomorphisms is a causal group) in general, is still inconclusive.

Various topologies on Minkowski space make the full Lorentz group into a topological group (J. L. KELLEY, 3S) with a topology different from the usual Lie group topology (i.e. of a six-dimensional real manifold). The separating topology on the Minkowski space (i.e. intersection of the order (A) topology (C), 1.2) with the Zeeman fine topology (DFN.5, 1.2)) induces on the Lorentz group a strictly finer topology than the Euclidean one, makes it into a semitopological group and induces the same topology with the Euclidean one on every compact subgroup; no new representations have been found (P.G. VROEGINDEWEIT).

E. C. Zeeman's result (PR.6, 1.2) that the group of causal automorphisms of a (causal) Minkowski space is isomorphic to the augmented orthochronous Poincaré group has been taken to mean that the Lorentz group and hence the fundamental relativistic invariants in physics can be deduced from purely order-theoretic assumptions, i.e. causal, without any recourse to the metric and/or affine structure of the Minkowski space (i.e. the underlying spacetime). This idea originated with A. A. Robb who though that a set of events and a before-after ordering of them would be sufficient to describe the properties of spacetime. Robb's axiomatics has been largely simplified with the use of lattice theoretic considerations; one can define parallelism, orthogonality and causality (on linear geometric objects) on an n-dimensional affine lattice (Z. DOMOTOR, IV, Definition 4) and hope to extend to causal lattices the classical representational result that an abstract lattice is isomorphic to the lattice of all subgeometries of a suitable abstract geometry with finitary operations if and only if it is a relatively atomic, upper continuous lattice (F. Maeda, "Lattice theoretic characterization of abstract geometries", J. Sc. Hiroshima Univ., Series A, Vol. 15 (1951) 87-96); one can actually prove that given an abstract causal lattice there exist a finite-dimensional vector space and a quadratic function on it which makes it into a Minkowski space whose (causal) lattice is isomorphic to the given one (Z. DOMOTOR, IV, Theorem 12); we still do not know how the causal lattice group, defined in an obvious way, is related to the group of causal automorphisms of the corresponding (causal) Minkowski space.

The most difficult and the really new (A. D. ALEXANDROV and W. V. OVCHINNIKOVA) part in E. C. Zeeman's proof (E. C. ZEEMAN, 1) is the proof of the linearity of the map which transforms cones into cones; one can even dispense with continuity (A. D. ALEXANDROV and O. S. ROTHAUS)

My motivation for the work of PART II can be summarized in the following: Geometric optics is a method for approximating solutions of Maxwell equations (eikonal approximation). It can be generally set up in a Riemannian manifold, whose positive-definite curvature is determined by the refractivity; light travels along the geodesics, i.e. the projection on the base manifold of the integral curves of the Hamiltonian vector field, defined on the cotangent bundle with the help of the metric tensor; these lie on the level surface of the Hamiltonian function. The Hamilton-Jacobi equation, determining the characteristics of the wave equation suggests that the method can be generalized for more general differential operators with a characteristic equation of the Hamilton-Jacobi type. Solutions are described by subsets of the cotangent bundle, in particular by what are called Lagrangian submanifolds (of a symplectic manifold in general). The projection on the base manifold of the points at which such a submanifold is not transverse to the fibers of the cotangent bundle (singular set) is called the caustic set of the corresponding submanifold. This definition agrees with the caustic's definition in geometric optics as envelope surfaces of light rays (see: J. G. Dubois, J. P. Dufour, O. Stanek "La théorie des catastrophes III Caustiques de l'optique géometrique", Ann. Inst. Henri Poincaré, Vol. 24, No. 3 (1976) 243-60).

A generalization is attempted to the case of normal hyperbolic manifolds (spacetimes of General Relativity) and the inhomogeneous Hamilton-Jacobi equation. In this case it is known that fundamental solutions for the wave equation near caustics do look different from the positive-definite case (see: F.G. Friedlander, "The Wave equation in curved spacetime", Cambridge Univ. Press, 1975). However is not obvious if caustics are the same with catastrophe sets of a gradient model or a family of functions, if their morphologies are stable and if so, what their normal forms look like.

Finally a note on the bibliography; that of PART I is thorough and as a result extensive; the material of PART II is part of a number of different fields of study and to survey the literature seemed a pointless exercise; so in BIBLIOGRAPHY II only the publications that are being used have been cited; however we have included in the main text and in particular in section 4.1, a number of related studies in order to place this work in a certain perspective.

The emphasis throughout has been on the mathematical, rather than the physical, aspects of the theories and the concepts considered; we hope this, although easier to conform to, has the advantage of being rigorous and clarifying rather than alienating.

CHAPTER 1

1.1 Introduction

Far from being ambitious to give a detailed account on the category of causation, whatever is meant by causation for the moment, I will not even attempt to arrive at a generally acceptable definition. The literature on causality is enormous and the number of definitions of the term almost equals the number of authors. An elementary discussion of causality from the point of view of a physicist see M. BUNGE. The everyday usage of the words 'cause' (C) and 'effect' (E), which everybody is familiar with, will suffice to the purpose of the following. Let us start with the formulation below:
'If C occurs, then and only then E is always produced by it'
or translating into the categorical mode:
'Every event of a certain class C necessarily produces an event at a certain class E'. (M. BUNGE, page 47).

From a philosopher's point of view it might be of some interest to inquire about the adequate logical correlates of the above (and various others) verbal statements of the causal principle. This attempt of the logical formalization of causal statements (if proved successful) is by no means a reduction of the causal problem to logical terms if only because what is logically possible need not be physically e.g. causally possible, and conversely the laws of nature, whether causal or not, are by no means logically necessary (causal relations are referred to a trait of reality and consequently cannot be settled a priori by purely logical means; see D. HUME, Section IV, Part II). On the other hand such an attempt is perfectly legitimate and let us get into it. Thus the logician is interested in studying the sentence: 'If C happens and C and E are causally connected then E must happen' by abstracting from the nature of the entities designated by 'C' and 'E' as well as from the specific character of the connection between C and E. One could argue about the formulation above that determines the cause as a sufficient reason only (not necessary or both); there is strong evidence that this is most likely to be the case (R.G. NEWTON). The sentence then can be regarded in any of the following alternative ways:

a) As an inclusion relation $C \subset E$ among classes

b) As a relation of implication $P_C \Rightarrow P_E$ among propositions

c) As a dyadic relation xRy among members x and y of the classes C and E.

It is not hard to see that b) leads us straight to paradoxes as: 'anything is self-caused' since proposition P always implies itself, and 'a cause if absent entails any effect' because anything follows from a false proposition. Moreover the relation of implication need not be irreversible (asymmetrical) where the cause-effect link is essentially one way. A few attempts* to adopt the cause-effect propositional relation to somehow modified one of implication has been proved unsuccessful.

Almost nothing has been said about a) and still c) remains as the most appropriate candidate. Indeed it summarises a few remarkable properties: it is a dyadic relation holding among elements interpretable as events and can be postulated to satisfy antireflexivity, which in turn can be interpreted as 'nihil est causa sui', transitivity and antisymmetry (i.e. $XRy \Rightarrow -yRx$). But the above properties are common to all ordered sets such as the succession of dawns and sunsets and although they specify the topology (J.L. KELLEY, Page 58) of the set do not say anything about the nature of its terms.

So we can conclude with an aphorism: the logical aspect of the causal problem is semantical rather than syntactical i.e. what is required is not an extension of formally logical relations but the determination of a type of semantic connection among terms that are relevant to each other.

* A.W. Burks: The logic of causal propositions, Mind (U.S.) 60, 363 (1951)
G.P. Henderson: Causal implication, Mind (U.S.) 63, 504 (1954)
H. Simon: On the definition of the causal relation, Journal of Philosophy 49, 517 (1952).
B.F. Chellas: Journal of Philosophical Logic 4 (1975) 133-153.

The most common way a physicist perceives the function of a causal connection is in terms of theories which allow, at least in principle, the calculation of the future state of the system under consideration from data specified at a time t_o. No specific reference to causes or effects is needed, but it is understood that all the phenomena (or variables) which can influence the system have been taken into account in the initial specification; which means that the system is closed. But if the system was indeed closed at all $t < t_o$, its past behaviour can also be calculated for all earlier times; so to call all our initial configurations 'cause' of all other configurations determined by it mathematically, would simply imply that effects can exist before as well as after the cause and this is a very unpleasant nomenclature. So functional interconnecteness is equivalent to deterministic causalism.

The fortunate thing is that such closed systems are ideal structures, approximately realised as closed subsystems of our universe, whenever we can attribute within a theory the separability condition to it i.e. if our universe is divided into subsystems (which are sufficiently far removed from each other) each subsystem can be described in terms of variables referring to it alone (e.g. Newtonian point mechanics can accommodate separability without difficulty).

If the system is open i.e. if interference by an outside agent is allowed, prediction of the future state of the system from its present one is not possible. If the interference is arbitrary no scientific statement can be made at all; if it is specified as definite function of time instead, the state of the system at times $t > t_o$ may still be calculated, but it is not a function of the state at t_o alone. One usually finds himself in a much more complicated situation than having a finite aggregate of physical objects (which can be considered as closed in certain circumstances) i.e. one has to deal with a portion of a mass continuum or a field.

The behaviour of a limited portion of a medium can no longer be calculated just from the initial data, as such a subsystem is in actual contact with the rest of the system and boundary conditions must be stated explicitly for all times. Fields necessarily extend over all space, so they can never describe a limited closed system, and the prescription of boundary conditions can not be dispensed with in this case. But boundary conditions such as required for the aforementioned cases can only be observed or set up arbitrarily and the calculations have to be tested experimentally.

Still, it seems absolutely necessary to deal with a theoretical description of open systems at some level; simply because we must communicate somehow i.e. use signals and a signal is an open system. Signals are necessitated by the Special Relativity Theory (definition of simultaneity and global spacetime coordinates); moreover are indispensible elements of the theory from an operational point of view and after all we have to rely on them if any degree of objectivity of our knowledge is to be attained. But a signal has to be an open system because if the initial state of the system would determine the future (past) state completely no information could be transmitted. For example in the case of the Maxwell's theory of electromagnetism the fact that we can set up arbitrarily the initial values, the changes of the charges and the currents and create discontinuities of the field which allow formation of a message, is the very proof that the theory can apply to an open system (the signal). To quote P. Havas (P. HAVAS, 1, Page 86)*'the program of a T.V. station for the day cannot be predicted by even the most detailed description of the station's condition at 6 a.m. - even though we viewers do get this impression at times'.

So one is interested in the following question: given for such a complicated system the initial data over a finite region and provided that non-local field theories are excluded (near action only is assumed

* A clear distinction between open and closed systems is essential and as far as I know has only been put manifestly together with some of the arising problems in Newtonian, Special and General Relativity Theory in P. HAVAS, 1 and 2.

as taking place) can we calculate something about the past and future behaviour of our finite region considered as a subsystem (closed apart from boundary effects)?

To summarize, we have started from the concept of causality in the ontal sphere and focussing on its epistemological and logical aspects we have reduced it to that of classification, particularly of ordering, of the elements of the class of the events.* A very popular representation of the class of events is one by a set endowed with a manifold structure One then can ask questions of the following kind:

Upon which events does the event p depend, (Domain of dependence)

Can the event p influence the event q by means of a signal? (Domain of influence).

The study of a relatively simple model of space-time (or event space) that of the Minkowski space-time of the Special Theory of Relativity, will be our next step.

1.2 Special Relativity Theory and Causal Considerations

The Minkowski spacetime M is supposed to be a four-dimensional, C^{∞}-manifold which admits a single coordinate chart i.e. the $\mathbb{R}^4$ with a quadratic form defined by:

$$Q(x) = x_0^2 - x_1^2 - x_2^2 - x_3^2 \qquad \forall x \in M$$

It is obvious that the group of homeomorphisms of M endowed with the Euclidean topology for $\mathbb{R}^4$ is much richer than the full Poincaré group which preserves the Q(x). It is reasonable then to ask for 'natural' (with respect to Q(x)) topologies on M, compatible or not with its natural as a manifold topology (i.e. the Euclidean topology of $\mathbb{R}^4$), but always compatible**with its linear vector space structure.

Using the quadratic form Q(x) we define the following dyadic relations:

* By event one usually understands the abstract concept of pointlike spacial and/or temporal phenomena

** G. Choquet, Topology, Academic Press, 1966, Chapter III, definition 1.1 and proposition 1.2.

DFN. 1 The chronology ($\ll$) by: $x \ll y$ iff $Q(x-y) > 0$ and $x_o \leq y_o$

DFN. 2 The causality ($<$) by: $x < y$ iff $Q(x-y) \geq 0$ and $x_o \leq y_o$

DFN. 3 The horismos ($\rightarrow$) by: $x \rightarrow y$ iff $Q(x-y) = 0$ and $x_o \leq y_o$

PR. 1 The chronology and the causality relations are transitive.

Hence they constitute partial orderings (J.L. KELLEY, Page 13).

PR. 2 $x \rightarrow y$ iff $x < y$ and $x \not\ll y$

PR. 3 The horismos is not a transitive relation

PR. 4 The chronology is an antireflexive relation where the causality is a reflexive one (i.e. $x \not\ll x$ but $x < x$)

The proofs are obvious from the null cone geometry.

One is tempted to ask which transformations on M preserve the above relations.

DFN. 4 Let $f : M \rightarrow M$ be a one-to-one map. We call f a causal automorphism iff both f and f^{-1} preserve the chronology relation (i.e. $x \ll y \Leftrightarrow fx \ll fy \quad x, y \in M$)

If f is a causal automorphism, preserves the causality (and hence the horismos too by PR. 2, 1.2) by:

PR. 5 Let $f : M \rightarrow M$ be a one-to-one map. Then f, f^{-1} preserve the partial ordering $\ll$ iff they preserve the relation $<$.

For a proof see E.C. ZEEMAN (1, Lemma 1). E.C. ZEEMAN in 1 has also proved:

PR. 6 Causal automorphisms form a group (causality group) and this is the group G generated by the orthochronous Lorentz group plus the translations plus the dilatations (multiplication by a scalar).

It is easy to see that G is contained in the causality group but the inverse is not at all obvious and the result depends essentially upon space being more than one-dimensional as can be seen by the following counter example:

Let K denote two-dimensional Minkowski space with $Q(x) = x_0^2 - x_1^2$ $x = (x_0, x_1) \in K$. Choose new coordinates $y_o = x_0 - x_1$, $y_1 = x_0 + x_1$. Then $Q(y) = y_0 y_1$. Let $f_0, f_1 : \mathbb{R} \to \mathbb{R}$ be two arbitrary, nonlinear, orientation preserving homeomorphisms of the real line into itself. Define $f : K \to K$ by : $f(y_0, y_1) = (f_0\, y_0, f_1 y_1)$. Then f is a causal automorphism but $f \notin G$ because f is non-linear.

PR. 6 above shows that causality requirements on M are in absolute accord with the Lorentz structure.

To return where we started i.e. to talk about topology and their homeomorphisms, we observe that all that has been said above is referred to M endowed with its Euclidean topology where the quantities $N_\varepsilon^E =$ $\{y \in M : \rho(x,y) < \varepsilon\}$ $\rho(x,y) = (x_o - y_o)^2 + (x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2$ (Euclidean ε-neighbourhoods) form a base for the neighbourhood system of the point x. (Appendix I). But observe that Euclidean topology is locally homogeneous (in a loose sense) whereas M is not (every point has its associated light cone separating space vectors from time vectors) and to repeat it again the group of homeomorphism of Euclidean space is vast and of no physical significance.

But how does one understand the function of neighbourhoods, open sets and all that? What more a topological structure says about the the elements of a set is to ask what our neighbourhoods tell us when we say we are 'close' to a point. A relation of 'closer than' can be defined (G. WILLIAMS, 1) in terms of our Lorentzian (quadratic form) in an invariant way for all timelike and spacelike events (separately) but not for events on distinct null geodesics.* It turns out that the definition is not only invariant under the group of transformations leaving invariant Q(x) but it is not also devoid of any physical meaning e.g. if 'a' and 'b' are two events in the future

* For the general problem in case of pseudo-Riemmanian geometry see: E. Schrodinger: Space time structure, Cambridge Univ. Press, 1950.

(past) of x, we say that 'a' is closer than 'b' to x if the proper distance along the line joining x to 'a' is strictly smaller than that along the line joining x to 'b'; i.e. the time recorded by an inertial observer between the events x and 'a' will be smaller than that recorded by such an observer travelling between x and 'b'. Hence an intrinsic topology for Minkowski space should 'ignore' the null lines.

As a first attempt we should consider Zeeman's fine topology (E.C. ZEEMAN, 2).

DFN. 5 The fine topology for M is the finest topology for M with the property to induce the one-dimensional Euclidean topology on every time axis where a time axis is any subset of M of the form $\mathbb{R}_t = \mathbb{R} \times \{0\}$ and a space axis $\mathbb{R}_3 = \{0\} \times \mathbb{R}^3$.

Typical open sets in M with the fine topology (M^F) are the ε-fine neighbourhoods $N_\varepsilon^F(x) = N^E(x) \cap (C^T(x) \cup C^S(x))$ with $C^T(x) = \{y \in M: y = x$ or $Q(y-x) > 0\}$ and $C^S(x) = \{y \in M : y = x$ or $Q(y-x) < 0\}$. They are obtained from $N_\varepsilon^E(x)$ by removing the light cone and replacing the point x.

PR. 7 $N_\varepsilon^F(x)$ are open in M^F but not in M^E.*

PROOF To show that $N_\varepsilon^F(x)$ is open in M^F it is enough to show that $N_\varepsilon^F(x)$ is open in any space or time axis A i.e. $N_\varepsilon^F(x) \cap A$ is open in the Euclidean topology. Indeed:

$$N^F(x) \cap A = \begin{cases} N_\varepsilon^E(x) \cap A & \text{if } x \in A \\ (N^E(x) - C^T(x)) \cap A & \text{if } x \notin A \end{cases}$$

Both the right hand sides are open in A because $N_\varepsilon^E(x)$, $N_\varepsilon^E(x) - C^T(x)$ are open in M^E.

Typical closed sets in M^F but not closed in M^E are the Zeno sequences defined by:

DFN. 6 A Zeno sequence converging to z is a disjoint sequence of points in M that converge to z in M^E but does not converge in M^F.

PR. 8 A Zeno sequence $Z = \{z_n\}$ is closed in M^F but not closed in M^E.

* By M^E we denote the Minkowski spacetime with its Euclidean topology.

PROOF $z_n \overset{\lim}{\to} z$ in M^E implies $z \in \overline{Z}^E$ (closure in Euclidean topology). Since $z \notin Z$ it follows that Z is not closed in M^E.

Also, $z_n \overset{\lim}{\not\to} z$ in M^F means that there exists a neighbourhood 0 of z in fine topology such that $0 \cap Z = \phi$. Hence z is an interior point of $-Z^F$ (complement in the fine topology). We claim that $-Z^F$ is open in M^F. It is enough to show that every $x \neq z$ of $-Z^F$ is an interior point of $-Z^F$. Let $x \neq z$ be any point of $-Z^F$; then there exists a Euclidean neighbourhood of x not meeting Z (which also serves as neighbourhood in the fine topology since the fine topology is finer than the Euclidean topology). For if not, every Euclidean neighbourhood of x will contain infinitely many points of Z and consequently x will be the limit point of Z in M^E, thus giving a contradiction since M^E is Hausdorff and $x \neq z$.

PR. 9 The ε-fine neighbourhoods of x do not form a base of neighbourhoods of x in the fine topology.

PROOF If a Zeno sequence is removed from a ε-fine neighbourhood of x then what is left is an open neighbourhood of x (in M^E) containing no ε-fine neighbourhood.

Summarising the results of E.C. ZEEMAN, 2, the fine topology is Hausdorff, induces a discrete topology on a light ray,* is not normal, does not have a countable base of neighbourhoods at any point (i.e. is not first countable and hence is not second countable as well), is not locally compact, is locally connected and connected, and has the very attractive feature:

PR. 10 The group of homeomorphisms of M^F is the group of automorphisms of M given by the full Lorentz group plus the translations plus the dilatations. (E.C. ZEEMAN, 2, Theorem 3).

E.C. Zeeman in (E.C. ZEEMAN, 2) conjectures that alternatives topologies

* Intuitively this says, that the track of a photon is not a continuous path: we have no evidence of a photon other than the discrete events of its emission and absorption.

for M could be obtained each of which has the same homeomorphism group as that of the fine topology. The general procedure was to reduce the information available about the induced topologies on spacelike hyperplanes, timelike and lightlike line. S. Nanda has proved (S. NANDA, 1 theorem 2) :

PR. 11 The group of homeomorphisms of the finest topology which induces the Euclidean topology on every spacelike hyperplane is the full Lorentz group plus translations plus dilatations.

Another approach to arrive at topologies with a physical significance or counterpart is to depart from the causal relations.

DFN. 7 The chronological future (past) of a set A is given by

$$I^+(A) = \{y \in M : x \ll y \quad \forall x \in A\} \quad (I^-(A) = \{y \in M : y \ll x \quad \forall x \in A\})$$

In particular $I^{\pm}(x)$ is the chronological future (past) of a single event.

Consider the following families of sets:

a) $W_+ = \{I^+(x): \forall x \in M\}$, $W_- = \{I^-(x): \forall x \in M\}$

b) $W = \{I^+(x),\ I^-(x): \forall x \in M\}$

c) $(A) = \{I^+(x) \cap I^-(y): \forall x, y \in M\}$

and ask whether or not can each be bases or subbases for a topology for M. (For a summary of the theory of bases (subbases) for a topology, see Appendix I). Using the results of Appendix I, it is easy to prove:

PR. 12 W can be the subbase for some topology but not a base.

PR. 13 (A), W_+, W_- are bases for some topologies on M.

The topologies generated by W_+, W_- have been studied by G.S. WHISTON They have the same group of homeomorphisms and this is the orthochronous Lorentz group plus dilatations. They are rather complicated topologically as they are only T_0, not T_2 (hence not Hansdorff), not T_3, not paracompact, not locally compact, but they are locally connected and connected.

E.C. Zeeman's work was followed by a number of attempts to generalize its theorems (PR. 10, I.2) (J.L. ALONSO - F.J. YUDURAIN and C. GHEORGHE - E. MIHUL) and extend it (G. WILLIAMS, 2). Also in a number of publications one can find a detailed study of the causality group itself (G. TEPPATI, G. BARUCCHI).

1.3 Causal Spaces - The Abstract Theory

We represent the work done by E.H. Kronheimer and R. Penrose (E.H. KRONHEIMER, R. PENROSE, 1) in 1966. We rearrange the material and provide some more detailed proofs and proofs of side conjectures in the original.

DFN. 8 Any set X endowed with three dyadic relations ($<$, $<<$, $\rightarrow$) (named causality, chronology and horismos respectively) satisfying the properties below is called a causal space:

1. $x < x \quad \forall x \in X$
2. $x < y$ and $y < z$ imply $x < z$
3. $x < y$ and $y < x$ imply $x = y$
4. $x \not<< x \quad \forall x \in X$
5. $x << y$ implies $x < y$
6. $x < y$ and $y << z$ imply $x << z$
7. $x << y$ and $y < z$ imply $x << z$
8. $x \rightarrow y$ iff $x < y$ and $x \not<< y$

PR. 14 Given X a causal space and $x < y < z$*, $x \rightarrow z$, then $x \rightarrow y \rightarrow z$.

PROOF Suppose $x << y$ ($y << z$). By axioms 6 and 7 of the DFN. 8, 1.3, $x << y < z$ ($x < y << z$) imply $x << z$ which contradicts $x \rightarrow z$ ($x < z$ and $x \not<< z$).

An alternative, equivalent axiom system should be:

I. $x < x \quad \forall x \in X$

II. $x < y < z \Longrightarrow x < z$

* $x < y < z$ stands for $x < y$ and $y < z$ in an obvious abbreviation.

III. $x < y < x \Rightarrow x = y$

IV. $x \rightarrow x \quad \forall x \in X$

V. $x \rightarrow y \Rightarrow x < y$

VI. $x < y < z$ and $x \rightarrow z \Rightarrow x \rightarrow y \rightarrow z$

VII. $x \ll y \Leftrightarrow x < y$ and $x \not\rightarrow y$

In view of PR. 14, 1.3, the equivalence is easily proved. Note that in each axiom system, one out of the three relations is defined trivially in terms of the other two (the horismos for the former and the chronology for the later one).

One can even construct a causal space, but not in a unique fashion, just from one relation adding an appropriate constrain. When certain conditions are met the resultant causal relations are compatible or even coincide. For this purpose, it is useful to regard causal relations on X as subsets of the Cartesian product X x X. (An order relation can be defined between order relations (!) $R_i / i \in I$ on X by the natural inclusion relation for subsets of X x X i.e. $R_i < R_j$ iff $R_i \subset R_j$. It is called the natural order in any class of dyadic relations $R_i / i \in I$ forming a set). Following E.H. KRONHEIMER (AND R. PENROSE) we introduce further the symbol {hor/cau <} for the collection of horismos compatible with the causality (<) to mean that for some relation '<<' the quadruple $(X, <, \rightarrow, \ll)$ is a causal space. (Similarly one introduces {chr/cau <}, {hor/chr <<}, e.t.c.).

At this stage we construct a causal space, given X and an horismotic* relation '→' on X :

DFN. 9 $x \overset{u}{<} y$ iff $\exists\ x_i / i = 1,2,\dots,n \ni x = x_1 \rightarrow x_2 \rightarrow \dots \rightarrow x_n = y$

DFN. 10 $x \overset{u}{\ll} y$ iff $x \overset{u}{<} y$ and $x \not\rightarrow y$

PR. 15 $(X, \overset{u}{<}, \rightarrow, \overset{u}{\ll})$ is a causal space.

DFN. 11 $(X, <, \rightarrow, \ll)$ is called a u-space iff '<' (and '<<') coincide with '$\overset{u}{<}$' and '$\overset{u}{\ll}$')

* A reflexive relation R is called 'horismotic' iff $\forall\ x_i / i = 1,2,\dots,n \ni x_i R x_{i+1} / i = 1,2,\dots,n$, and h and k integers $\ni 1 \leq h \leq k \leq n$, $x_1 R x_h \Rightarrow x_h R x_k$ and $x_n R x_1 \Rightarrow x_h = x_k$ e.g. any reflexive partial ordering with the property: xRy and $yRx \Rightarrow x = y$, is a horismotic relation. The horismos is a horismotic relation too (notwithstanding that it is not a partial order).

PR. 16 If $< \subset \overset{u}{<}$ (i.e. if $x < y$ implies $x \overset{u}{<} y$) then $(X, <, \to, <<)$ is a u-space.

PROOF $\forall < \in \{cau/hor \to\} \Rightarrow \overset{u}{<} \subset <$, i.e. $\overset{u}{<} = \bigcap \{cau/hor \to\}$. Since if $x \overset{u}{<} y \Rightarrow \exists x_i / i = 1,2,\ldots,n \ni x = x_1 \to x_2 \to \ldots x_n = y$ and by axiom 8 of the DFN. 8, 1.3, $x = x_1 < x_2 < \ldots < x_n = y$, which in turn by axiom 2 of the DFN. 8, 1.3, imply $x < y$. Hence $\overset{u}{<} \subset <$ combined with our hypothesis $< \subset \overset{u}{<}$ imply $\overset{u}{<} = <$. Furthermore if $< = \overset{u}{<} \Rightarrow << = \overset{u}{<<}$ by DFN. 10, 1.3, and axiom VII ($\overset{u}{<<} = \bigcap \{chr/hor \to\}$).

DFN. 12 The causal, chronological and null future (past) of a set A are given by:

$J^+(A) = \{x \in X : a < x \ \forall a \in A\}$ $(J^-(A) = \{x \in X : x < a \ \forall a \in A\})$

$I^+(A) = \{x \in X : a << x \ \forall a \in A\}$ $(I^-(A) = \{x \in X : x << a \ \forall a \in A\})$

$E^+(A) = J^+(A) - I^+(A)$ $(E^-(A) = J^-(A) - I^-(A))$

Introduce $[a,b]$ for $\{x \in X: a < x < b\}$, $\langle a,b \rangle$ for $\{x \in X: a << x << b\}$ and $a//b$ iff neither $a < b$ or $b < a$. Finally we define a topology on X namely:

DFN. 13 The Alexandrov topology J^* on a set X equipped with an anti-reflexive partial ordering '<<' is the smallest topology on X in which $I^{\pm}(x)$ are open $\forall x \in X$

DFN. 14 A set linearly ordered* by the causality (chronology) is called a causal (chronological) chain.

DFN. 15 A causal space is called regular iff for any distinct $x_i, y_j \ni x_i \to y_{j/i,j = 1,2}$ then the horismos orders $x_{1,2}$ iff it orders $y_{1,2}$.

In view of PR. 14, 1.3, this is equivalent to: For any distinct $x_i, y_j \ni x_i \to y_j / i,j = 1,2$ $x_1 // x_2$ iff $y_1 // y_2$.

* For the relation R on X linearly orders it iff R is a partial ordering and xRy or yRx $\forall x \in X$, $y \in X$, $x \neq y$.

Given a set X and an antireflexive partial ordering '<<' on X (e.g. chronology) a number of situations can arise. The relation '<<' is called future (past) reflecting iff $I^-(x) \subset I^-(y)$ $(I^+(x) \supset I^+(y))$ whenever $I^+(x) \supset I^+(y)$ $(I^-(x) \subset I^-(y)$. The relation '<<' is future (weakly) distinguishing iff $x = y$ whenever $I^+(x) = I^+(y)$ $(I^+(x) = I^+(y)$ and $I^-(x) = I^-(y))$. Finally '<<' is full iff $\forall x \in X\ \exists\ p \in X \ni p << x$; and if $p_1 << x$, $p_2 << x \Rightarrow \exists q \in X \ni p_1 << q$, $p_2 << q$ and $q << x$.

At this stage we can state a few theorems. They relate the Alexandrov topology for X to the above mentioned properties of the chronology. (Recall the notation $(A) = \{< x,y > : x \neq y,\ < x,y > \neq \emptyset,\ x,y \in X\}$ and PR. 13, 1.3, which turns out not to be true in general).

THR. 1 '<<' is full iff (A) is a base for J^*.

PROOF Assume fullness. We shall show that (A) is a base for $\cup\{A : A \in (A)\}$

It is enough to show that $\forall A_{1,2} \in (A)$ and $x \in A_1 \cap A_2 \Rightarrow \exists A \in (A)$ $\ni x \in A \subset A_1 \cap A_2$. ($A_{1,2}$ stands for: A_1, A_2).

Let $A_{1,2} = I^+(x_{1,2}) \cap I^-(y_{1,2})$. $\forall\ x \in A_1 \cap A_2 \neq \emptyset \Rightarrow x \in A_{1,2} \Rightarrow$ $x_{1,2} << x << y_{1,2}$ and by fullness $\exists q,q' \ni x_{1,2} << q << x << q' << y_{1,2}$ i.e. $x \in < q,q' > \subset A_1 \cap A_2$. Take $A = < q,q' >$.

To complete the proof note that $\forall x \in X \Rightarrow \exists p(p') \ni p << x$ $(x << p')$ i.e. $\exists < p, p' > \in (A) \ni x \in < p, p' >$. So $X \subset \cup\{A: A \in (A)\}$ and any set $I^+(x)$ $(I^-(x))$, open in J^*, can be written as: $I^+(x) = \{y \in X : x << y\} = \cup\{< x,y > : y \in X\}$ $(I^-(x) = \{y \in X : y << x\} = \cup\{< y,x > : y \in X\})$ i.e. as a union of the base elements. It follows then that (A) is a base for J^*.

Suppose (A) is a base for J^*. $\forall x \in X \Rightarrow x \in A$ for some $A \in (A)$ i.e. $x \in < p, p' >$ i.e. $\exists\ p(p')\ p << x$ $(x << p')$.

Since (A) is a base and $X \subset \cup\{A: A \in (A)\}$ $\forall A_{1,2}$ and $\forall x \in A_1 \cap A_2$, $x \in X \Rightarrow \exists A \ni x \in A \subset A_1 \cap A_2$ i.e. $\forall A_{1,2}$ and

$\forall\ x \in I^+(x_1) \cap I^-(y_1) \cap I^+(x_2) \cap I^-(y_2)\ \exists < q', q > \subset A_1 \cap A_2$

and $x \in \langle q', q \rangle$ i.e. $\forall x_{1,2} \ni x_{1,2} \ll x (\forall y_{1,2} \ni x \ll y_{1,2})$
$\Rightarrow \exists q(q') \ni x_{1,2} \ll q \ll x \quad (x \ll q' \ll y_{1,2})$.

THR. 2 If $(X, \mathcal{J}^*)$ is Hausdorff and '$\ll$' full then '$\ll$' is future (past) distinguishing.

PROOF Suppose $p,q \in X$, $p \in \langle x_1, y_1 \rangle$, $q \in \langle x_2, y_2 \rangle$, $\langle x_1, y_1 \rangle \cap \langle x_2, y_2 \rangle = \emptyset$, $I^+(p) = I^+(q)$ and $p \neq q$.

We'll show that they lead to a contradiction.

Let $z_1 \in \langle x_1, y_1 \rangle \ni p \ll z_1$ i.e. $z_1 \in I^+(p) = I^+(q)$, hence $q \ll z_1$

Let $z_2 \in \langle x_2, y_2 \rangle \ni q \ll z_2$ i.e. $z_2 \in I^+(q) = I^+(p)$, hence $p \ll z_2$

From $q,p \ll z_1 \quad p, z_1 \in \langle x_1, y_1 \rangle$ and

$q,p \ll z_2 \quad q, z_2 \in \langle x_2, y_2 \rangle \Rightarrow \exists w \ni p \ll w < z_{1,2}$

$w \in \langle x_1, y_1 \rangle$. But $p \ll w \Rightarrow w \in I^+(p) \Rightarrow w \in I^+(q) \Rightarrow q \ll w$

From $q \ll w \ll z_2 \Rightarrow w \in \langle x_2, y_2 \rangle$ which is a contradiction to $w \in \langle x_1, y_1 \rangle$ with $\langle x_1, y_1 \rangle \cap \langle x_2, y_2 \rangle = \emptyset$

Moreover if $(X_1 \mathcal{J}^*)$ is T_1 (i.e. $\forall p,q \in X$, $p \neq q \Rightarrow \exists \langle x,y \rangle \in \mathcal{J}^* \ni p \in \langle x, y \rangle$ and $q \notin \langle x, y \rangle$) and '$\ll$' is full, then X has to be weakly distinguishing. Because if we assume that $I^{\pm}(p) = I^{\pm}(q) \Rightarrow y \in I^+(p) = I^+(q) \Rightarrow q \ll y$ and $x \in I^-(p) = I^-(q) \Rightarrow x \ll q$ i.e. $q \in \langle x, y \rangle$, a contradiction. As T_1-spaces are T_o the argument is true for T_1-spaces too.

THR. 3 Let '$\ll$' be an antireflexive partial ordering on X. The following are equivalent statements:

1. '$\ll$' is full
2. $x \in I^+(x)^{cl*} \cap I^-(x)^{cl*} \forall x \in X$ (cl* means closure with respect to *-topol.).
3. $I^+(x)^{cl*} = \{y \in X : I^+(y) \subset I^+(x)\}$ and $I^-(x)^{cl*} = \{y \in X : I^-(y) \subset I^-(x)\} \ \forall x \in X$

PROOF Since $I^{\pm}(x) \subset I^{\pm}(x)$ 3. trivially implies 2.

2 implies 1. $\forall x \in X$ since $x \in I^-(x)^{cl*}$, $I^-(x)$ open in *-topology and $x \notin I^-(x)$, follows that x is an accumulation

point of $I^-(x)$ i.e. $\exists$ N(x) (a neighbourhood of x) $\ni$ $N(x) \cap I^-(x) \neq \phi$; which in turn proves that exist $y << x$. ($y \in N(x) \cap I^-(x)$ $y \neq x$) If two $y_{1,2} \in N(x) \cap I^-(x)$ exist, it follows that $y_{1,2} << x$ and since $I^+(y_1) \cap I^+(y_2) \neq \phi$ (as $x \in I^+(y_1) \cap I^+(y_2)$) and open (definition of $*$-topology) exist $N'(x)$ (neighbourhood of x) $\ni N'(x) \subset I^+(y_1) \cap I^+(y_2)$. But again $N'(x) \cap I^-(x) \neq \phi$ (x is an accumulation point of $I^-(x)$) hence $\exists$ $p \in N'(x) \ni y_{1,2} << p << x$ and this completes the requirements of fullness.

1 implies 3. (a) $\forall y \in I^+(x)^{cl*}$ and $\forall z \in I^+(y)$ i.e. $y << z$ it follows that $y \in I^-(z)$ and $I^-(z)$ is a $*$-neighbourhood of y; hence intersects $I^+(x)$ ($y \in I^+(x)^{cl*}$) i.e. $\exists$ $w \in I^-(z)$ $x << w << z$ i.e. $z \in I^+(x)$ consequently $I^+(y) \subset I^+(x)$. (b) $\forall y \in X \ni I^+(y) \subset I^+(x)$ and any $*$-neighbourhood of y, there exists a $< u, v >$-neighbourhood of it $\ni (y, v) \neq \phi$. Hence $\exists$ $w \in I^+(y) \subset I^+(x)$ i.e. $w \in I^+(x)$; and $y \in I^+(x)^{cl*}$

These are enough to cover the equivalence between 1., 2. and 3.

THR. 4 Let $A \subset X$ and X a full causal space. Then $J^+(A) \subset I^+(A)^{cl*}$ and $I^+(A) = J^+(A)^{int*}$

PROOF $J^+(a) = \{y \in X : a < y\}$. $\forall x \in I^+(y)$ i.e. $y << x$ and $y \in J^+(a)$ i.e. $a < y \Rightarrow a << x$ which means that also $y \in \{y \in X: a << x \ \forall x \in I^+(y)\} = \{y \in X: x \in I^+(a) \ \forall x \in I^+(y)\} = \{y \in X: I^+(y) \subset I^+(a)\} = I^+(a)^{cl*}$ by THR. 3 (3.); and this is true $\forall a \in A$.

$\forall x \in J^+(A)^{int*}$ $\exists$ a $*$-neighbourhood of x in $J^+(A)$ i.e. $\exists$ p, q, r $\ni x \in < p, q > \subset J^+(A)$ and $p << r << x$ i.e. $x \in I^+(A)$; therefore $J^+(A)^{int*} \subset I^+(A)$. Now note that $I^+(A) \subset J^+(A)$ and $I^+(A)$ is open. But $J^+(A)^{int*}$ is the greatest open set contained in $J^+(A)$ hence $I^+(A) \subset J^+(A)^{int*}$. So $J^+(A)^{int*} = I^+(A)$

PR. 17 Given X a causal space a necessary and sufficient condition for H to satisfy $I^+(A) \subset H \subset J^+(A)$ for some A is that $I^+(H) \subset H$ ($H \subset X$).

PROOF Assume $I^+(A) \subset H \subset J^+(A)$ for some A. $\forall\, y \in H \Rightarrow y \in J^+(A) =$ $\{z \in X:\ a < z\ a \in A\}$ i.e. $a < y$; note that $\forall z \in I^+(H) =$ $\{z \in X:\ y << z,\ y \in H\}$ and $\forall\ y \in H \Rightarrow a < y << z$ i.e. $a << z$ or equivalently $z \in I^+(A) \subset H$. So $I^+(H) \subset H$.

Conversely given $I^+(H) \subset H$, take A = H.Then $J^+(H) =$ $\{y:\ h < y,\ h \in H\}$ and since $h < h \Rightarrow H \subset J^+(H)$ i.e. $I^+(H) \subset H \subset J^+(H)$.

PR. 18 Given $H \subset X$, X causal space and

1. $I^+(H) \subset H$
2. $I^-(-H) \subset -H$
3. $I^+(H) \cap I^-(-H) = \phi$

1. and 2. are equivalent and imply 3.

3. imply 1. if the space is 'partially full' i.e. given $x \in X$ and $y \in X \ni x << y \Rightarrow \exists\, z \in X \ni x << z << y$

PROOF $1. \Rightarrow 2.$

$J^-(-H) = \{x:\ x << y,\ y \in -H\}$ and note that $x \in H$ and $y \in -H$ imply $x << y$; since if $x << y$ was true $\Rightarrow y \in I^-(x)$ i.e. $y \in H$ by 1., contradicting $y \in -H$. Hence $x \notin H$ and $I^-(-H)\quad H = \emptyset$ i.e. $I^-(-H) \subset -H$.

$2. \Rightarrow 1.$

$I^+(H) = \{x:\ y << x,\ y \in H\}$ and again $x \in -H$, $y \in H$ imply $y \not<< x$; since if $y << x$ was true $\Rightarrow y \in I^-(x)$ i.e. $y \in I^-(-H)$ i.e. $y \in -H$ by 2., contradicting $y \in H$. So $I^+(H) \cap (-H) = \phi$ and $I^+(H) \subset H$.

$1.(2.) \Rightarrow 3.$ in an obvious way.

$3. \Rightarrow 1.(2.)$ whenever X is 'partially full'.

By 3. $\nexists\ x \in I^+(H) \cap I^-(-H)$ i.e. $\nexists\ x \ni h << x << h'$ $h \in H$, $h' \in -H$.

Moreover $I^+(H) \cap (-H) = \phi$ $(I^-(-H) \cap H = \phi)$ since if $x \in I^+(H) \cap (-H)$ $\exists\ h \in H \ni h << x$ and by partial fullness $\exists\, y \ni h << y << x$ i.e. $y \in I^+(H)$ and $y \in I^-(-H)$ contradicting our hypothesis (if $x \in I^{-}(-H) \cap H \ \exists \ h' \in -H$ and $y \ni x \ll y \ll h'$ i.e. $y \in I^{+}(H)$ since $x \in H$ and yet $y \in I^{-}(-H)$, the same contradiction too.)

Sets H, satisfying the condition $I^{+}(H) \subset H$ are called future sets; e.g. $I^{+}(N)$, $J^{+}(N)$ are future sets for every $N \subset X$.

THR. 5 Given X a full causal space and $A \subset X$, then all sets H satisfying $I^{+}(A) \subset H \subset J^{+}(A)$ have the same closure and the same interior with respect to *-topology. Moreover $E^{+}(A) \subset H^{bdy*}$ $(E^{+}(A) = J^{+}(A) - I^{+}(A))$

PROOF By THR. 4: $H \subset J^{+}(A) \subset I^{+}(A)^{cl*} \Rightarrow H^{cl*} \subset I^{+}(A)^{cl*}$

$I^{+}(A) \subset H \Rightarrow I^{+}(A)^{cl*} \subset H^{cl*}$ hence $H^{cl*} = I^{+}(A)^{cl*}$, the same $\forall$ H.

$I^{+}(A) = J^{+}(A)^{int*}$, $H \subset J^{+}(A)$ and $I^{+}(H) \subset H^{int*}$ ($I^{+}(H)$ being open in *-topology) $\Rightarrow I^{+}(H) = H^{int*}$.

$E^{+}(A) = (J^{+}(A) - I^{+}(A)) \subset (I^{+}(A)^{cl*} - I^{+}(A)) = (H^{cl*} - H^{int*}) = H^{bdy*}$

A 'boundary' can be defined for such a set $(I^{+}(H) \subset H)$ without resort to any topology by: $B^{+}(H) = \{x \in X : I^{+}(x) \subset H \text{ and } I^{-}(x) \subset -H\}$.

THR. 6 Given $I^{+}(H) \subset H \Rightarrow H^{bdy*} \subset B^{+}(H)$ and $H^{bdy*} = B^{+}(H)$ whenever X is full.

PROOF Suppose X is full and $x \in B^{+}(H)$. Any nbd. N of x is a set of the form $< p, q >$ where $p \ll x \ll q$. Choose p', q' such that $p \ll p' \ll x \ll q' \ll q$. Then $p' \in I^{-}(x) \subset X - H$ so $< p, q >$ intersects $X - H$. Also $q' \in I^{+}(x) \subset H$, so $< p, q >$ intersects H. Hence N intersects both H and X-H i.e. $z \in H$ bdy*.

Let $x \in H^{bdy*}$ (I); suppose $I^{-}(x) \not\subset X - H$, then $\exists \ h \in I^{-}(x) \cap H$ i.e. $x \in I^{+}(h) \subset H$. So $I^{+}(h)$ is a nbd. of x which fails to intersect $X - H$ contradicting (I).

Suppose $I^{-}(x) \not\subset H$; then $\exists \ z \in I^{+}(z) - H$. Since $z \notin H \supset I^{+}(H)$, $I^{-}(z) \cap H = \phi$. Hence $I^{-}(z)$ is a nbd. of x which fails to intersect H, contradicting (I). Thus $x \in B^{+}(H)$.

To state and prove the next theorem, the notion inherent to the construction of a causal space from a partial antireflexive ordering (chronology) is necessary.

Given X and a partial antireflexive ordering '<<' on X, define:

DFN. 16 $x \overset{B}{<} y$ iff $I^+(x) \supset I^+(y)$ and $I^-(x) \subset I^-(y)$

DFN. 17 $x \overset{B+}{<} y$ iff $I^+(x) \supset I^+(y)$ Note that: $\overset{B}{<} \subset \overset{B^+}{<}$

DFN. 18 $x \overset{B}{\rightarrow} y$ iff $x \overset{B}{<} y$ and $x \not\ll y$

PR. 19 $(X, \overset{B}{<}, \overset{B+}{\rightarrow}, \ll)$ is a causal space iff '<<' is weakly distinguishing (to ensure the validity of axiom 3, DFN. 8; 1.3).

Correspondingly:

$(X, \overset{B+}{<}, \overset{B+}{\rightarrow}, \ll)$ is a causal space iff '<<' is future reflecting and weakly distinguishing.

DFN. 19 $(X, <, \rightarrow, \ll)$ is called B-space (future-reflecting B-space) iff '<'(and '→') coincide with $\overset{B}{<}, \overset{B}{\rightarrow}$ $(\overset{B+}{<}, \overset{B+}{\rightarrow})$

PR. 20 If $\overset{B}{<} \subset <$ (i.e. if $x \overset{B}{<} y \Rightarrow x < y$) then $(x, <, \rightarrow, \ll)$ is a B-space (and hence '<<' is weakly distinguishing).

Correspondingly:

PR. 21 If $\overset{B+}{<} \subset <$ (i.e. if $x \overset{B+}{<} y \Rightarrow x < y$) then $(X, <, \rightarrow, \ll)$ is a future reflecting B-space (and hence '<<' is future-reflecting and future distinguishing).

PROOF of PR. 20 if $< \in \{caus/chr \ll\}$ then $< \subset \overset{B}{<}$ (i.e. if $x < y \Rightarrow x \overset{B}{<} y$) since $x < y \Rightarrow x \in I^-(y)$ and $y \in I^+(x)$; $\forall z \in I^-(x)$ and $z' \in I^+(y) \Rightarrow z \ll x \ll y$ and $x \ll y \ll z'$ i.e. $z \ll y$ and $x \ll z'$ i.e. $I^-(x) \subset I^-(y)$ and $I^+(y) \subset I^+(x)$. Furthermore given $x \overset{B}{<} y$ $\exists < \in \{caus/chr. \ll\} \ni x < y$ ($u < v$ iff $u = v$ or $u \ll v$ or $u = x$ and $v = y$), so $\overset{B}{<} = \bigcup \{cau/chr \ll\}$($\overset{B}{\rightarrow} = \bigcup\{hor/chr \ll\}$)

DFN. 20 Denote by J^+ the smallest topology with respect to which $J^-(x)\ \forall\ x \in X$ are closed.

THR. 7 Let X be a full causal space. Then the following statements are equivalent

1. X is a future-reflecting B-space (i.e. '<<' is future reflecting, future distinguishing and $\overset{B+}{<} = <$).

2. $J^+(x) = I^+(x)^{cl*} \quad \forall\, x \in X$

3. $J^+(A) = I^+(A)^{cl*} \quad \forall\, A \subset X$, A compact with respect to J^+ topology

4. $J^+(A)^{bdy*} = E^+(A)\ \forall\, A \subset X$, A compact with respect to J^+ topology

PROOF 2. $\Rightarrow$ 1.

By virtue of PR. 21, Chapter 1, sec. 3, it is enough to show that $\overset{B+}{<} \subset <$ i.e. if $x \overset{B+}{<} y \Rightarrow x < y$. But $x \overset{B+}{<} y \Leftrightarrow I^+(x) \supset I^+(y)$ by DFN. 17, 1.3. Since $I^+(x)^{cl*} = \{y \in X : I^+(x) \supset I^+(y)\}$ by THR. 3 and $I^+(x)^{cl*} = J^+(x) = \{y \in X : x < y\}$ by our hypothesis, $I^+(x) \supset I^+(y)$ implies $x < y$.

3. $\Rightarrow$ 2. obviously

3. $\Rightarrow$ 4.

Assume $J^+(A) = I^+(A)^{\;*}$ Apply THR. 5 with $J^+(A) = H$: $I^+(A) \subset I^+(A)^{cl*} = J^+(A) \subset J^+(A)$ and $J^+(A)^{bdy*} = (J^+(A)^{cl*} - J^+(A)^{int*}) = ((I^+(A)^{cl*})^{cl*} - I^+(A))$, using $I^+(A) = J^+(A)^{int*}$ by THR. 4, 1.3, and $I^+(A)^{cl*} = J^+(A)$ by our hypothesis. (Note that $(A^{cl})^{cl} = A^{cl}$). Hence $J^+(A)^{bdy*} = J^+(A) - I^+(A) = E^+(A)$.

4. $\Rightarrow$ 3.

Suppose $I^+(A)^{bdy*} = E^+(A) = J^+(A) - I^+(A)$ and apply THR. 5 with $I^+(A) = H$: $I^+(A) \subset I^+(A) = J^+(A)^{int*} \subset J^+(A)$ where $I^+(A) = J^+(A)^{int*}$ by THR. 4. Also $I^+(A)^{bdy*} = (I^+(A)^{cl*} - I^+(A)^{int*}) = I^+(A)^{cl*} - I^+(A)$ since $I^+(A)$ is open with respect to the *-topol. Compare with $I^+(A)^{bdy*} = J^+(A) - I^+(A)$ (our hypothesis) to get $J^+(A) = I^+(A)^{cl*}$.

Finally one can prove that 1. $\Rightarrow$ 3. (E.H. KRONHEIMER, R. PENROSE, 1, p. 494) (and it is at this part of the proof that compactness is employed and all these establish the equivalence.

The next theorem states a condition of comparability of J^+ (DFN. 20, 1.3) and J^* (DFN. 13) topologies.

THR. 8 Let X be a full causal space. $J^+ \subset J^*$ iff X is a past reflecting B-space.

PROOF Assume X to be full and $J^+ \subset J^*$. Then $J^-(x)$ is closed in the *-topology; this is so since $J^-(x)^{cl*} \subset J^-(x)^{cl+}$ and $J^-(x)$ closed in the J^+ topology by definition. But this is the statement 2. of the dual or the THR. 7, 1.3, and so X is past-reflecting B-space.

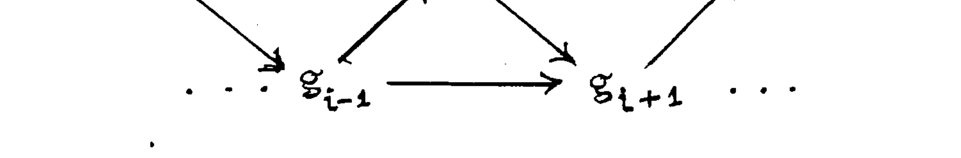
Conversely if X is past-reflecting B-space, by the dual of the THR. 7 (2), 1.3, $J^-(x) = I^-(x)^{cl^*}$ i.e. $J^-(x)$ is *-closed. Since J^+ is the smallest topology with respect to which $J^-(x)$ are closed $\Rightarrow J^+ \subset J^*$.

Consider a set X with a horismotic relation '$\rightarrow$' on it. Construct a partial reflexive ordering defined by $\overset{u}{<}$ (to recall DFN. 9, 1.3, $x \overset{u}{<} y$ iff $\exists\ x_i/i = 1,2,\ldots,n \ni x = x_1 \rightarrow x_2 \rightarrow \ldots \rightarrow x_n = y$) Suppose that a finite subset F of X is linearly ordered by $\overset{u}{<}$, and arrange the points of F in their ordering by $\overset{u}{<}$ (It follows that any subset of F is linearly ordered by $\overset{u}{<}$ too). One is interested to know which are the minimal chains with respect to '$\rightarrow$' in F.

Any two elements of F form a (trivial) chain whenever $x \rightarrow y$. If any three successive elements of F form a chain then F is called a girder. This is equivalent to Penrose's definition:

DFN. 21 F is a girder means F is a finite sequence ($g_i/i = 1,2,\ldots,N$) of elements of X linearly ordered by $\overset{u}{<}$ (any causality relation in general), $N \geq 3$, and $g_i \rightarrow g_j$ whenever $j-i = 2$.

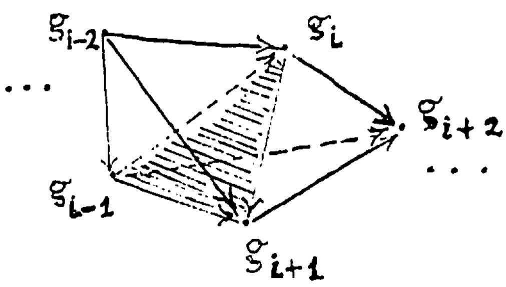
By PR. 14, 1.3, $g_i \rightarrow g_j$ whenever $j-i = 1$ and the following diagram is representative of the situation:

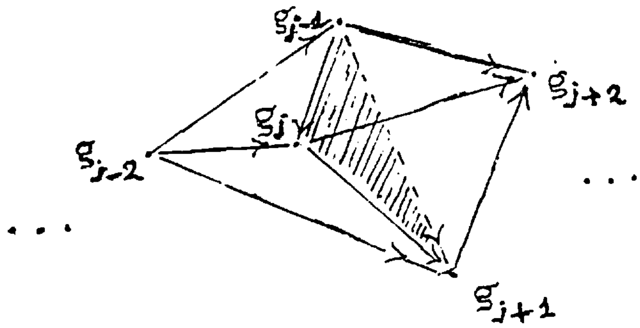


One can go a step further and consider the case where any four successive elements are linearly ordered by '$\rightarrow$' i.e.

$g_{i-2} \rightarrow g_{i-1} \rightarrow g_i \rightarrow g_{i+1}$ with $g_{i-2} \rightarrow g_i$, $g_{i-1} \rightarrow g_{i+1}$ and $g_{i-2} \rightarrow g_{i+1}$.

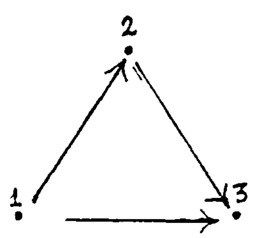
Pictorially:



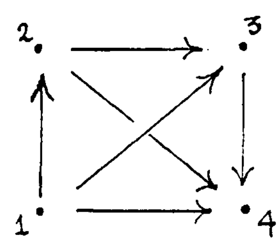


One can even go to larger subsets of F linearly ordered by '$\rightarrow$', until one ends up with all the N elements of F linearly ordered by the horismos.

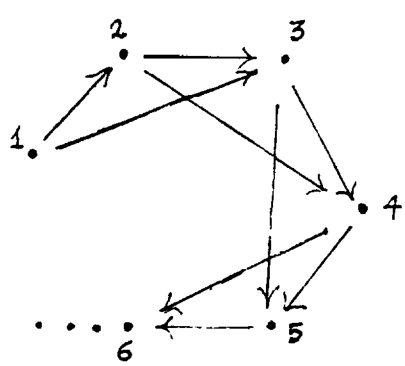
Another way to represent the situation above is through polygons and their diagonals:



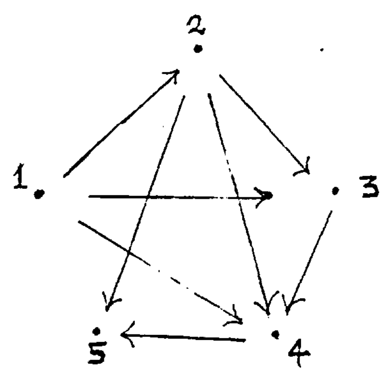
A three element unit



A four element unit



A girder



A set of five points linearly ordered by the horisoms

We will come back to subsets linearly ordered by the horismos when we talk about the links etc.

PR. 22 If $(g_1, \ldots, g_N)$ is a girder and h lies strictly between g_r and g_{r+1} (with respect to the causality ordering), then $(g_1, \ldots, g_r, h, g_{r+1}, \ldots, g_N)$ is a girder.

DFN. 22 H ($H \subset X$, X endowed with a horismotic relation) is called a hypergirder iff $H \neq \phi$ and $\forall\, x,y \in H\ \exists$ girder $G \ni x,y \in G$ and $G \subset H$.

PR. 23 If F is any finite subset of the hypergirder H, there exists a girder G such that $F \subset G \subset H$.

DFN. 23 A proper beam is a maximal hypergirder (i.e. a hypergirder which is not a proper subset of any hypergirder).

PR. 24 Every hypergirder is contained in some proper beam (by Zorn's Lemma).

DFN. 24 Two points are called proximate iff they are successive points of some girder.

THR. 9 Each pair of proximate points belong to precisely one proper beam iff the causal space is regular.

PROOF E.H. KRONHEIMER, R. PENROSE, 1, pages 490-491.

Note that the 'proximate' cannot be weakened to 'ordered by the horismos'.

One can generalize the notion of the proper beam to the following:

DFN. 25 A subset B of a causal space is called a beam iff either

1. B is a proper beam (i.e. a maximal hypergirder) or
2. $B = \{a,b : a,b \in X\ ,\ a \neq b,\ a \rightarrow b\}$ and B is not contained is any girder.

Then a theorem in analogy to THR. 9, 1.3, can be stated, dispensing with the regularity constrain.

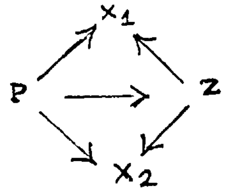
THR. 10 In any causal space each non-trivial set linearly ordered by the horismos is a subset of some beam.

PROOF Let $P \neq \phi$ linearly ordered by the horismos. Then $K = \dot{\cup}\{[p,q] : p,q \in P\}$ is partially ordered by '<' and choose a maximal chain containing P (P is a chain with respect to '$<$' as well by axiom 8 of DFN. 8, 1.3). If H has more than two points, it is a hypergirder and by PR. 24, 1.3, is contained is some proper beam (and hence a beam). If H has just two points (i.e. P has just two points) and it is maximal, it means that these two points cannot be proximate and condition (2) of DFN. 25, 1.3, are met to give us a beam.

Consider a subset F of X , where X is a causal space, and suppose F is linearly ordered by the horismos. Then F is linearly ordered by the causality too (axiom 8, DFN. 8, 1.3) i.e. F is a causal chain. It is somehow interesting to inquire the converse; i.e. when a causal chain is linearly ordered by the horismos? At first the most probable subset of X linearly ordered by the causality is of theform $L = [p,q]$. Almost nothing can be said without imposing further conditions. For finite L the answer was given with the construction of girders. Now suppose all you only know is $p \rightarrow q$. One can prove:

PR. 25 If X is a full and regular causal space then $L = [p,q]$, $p \rightarrow q$, is linearly ordered by the horismos (hence it is a causal chain in a trivial sense).

PROOF Fullness imply that $\forall x_{1,2} \in L$ i.e. $p < x_{1,2} < q \; \exists z \in L \ni p < z < x_{1,2}$. PR. 14, 1.3, implies $p \rightarrow z$, $p \rightarrow x_{1,2}$, $z \rightarrow x_{1,2}$, i.e.



and by regularity $x_1 \rightarrow x_2$ (or $x_2 \rightarrow x_1$)

To answer our question:

PR. 26 If $L = [p,q]$, $p \rightarrow q$ and it is a causal chain then L is linearly ordered by the horismos.

PROOF $\forall x_{1,2} \in L \Rightarrow x_1 < x_2$ $(x_2 < x_1)$. Since $p < x_{1,2} < q$ and $p \rightarrow q \Rightarrow p \rightarrow x_{1,2} \rightarrow q$ (by PR. 14, 1.3) and $p < x_1 < x_2$, $p \rightarrow x_2$ $(p < x_2 < x_1,\ p \rightarrow x_1) \Rightarrow x_1 \rightarrow z_2$ $(x_2 \rightarrow x_1)$.

Another possible answer is to postulate that in a set endowed with a reflexive, transitive and antisymmetric relation (i.e. a causality relation), the causal chains define horismodal connections, i.e. given X and a relation '$<$' on it satisfying axioms 1, 2 and 3 of DFN. 8, 1.3, define:

DFN. 27 $x \overset{C}{\to} y$ iff $x < y$ and '$<$' linearly orders, any proper subset $[u,v]$ of $[x,y]$.

DFN. 28 $x \overset{C}{\ll} y$ iff $x < y$ and $x \overset{C}{\not\to} y$

DFN. 29 A causal space $(X, <, \to, \ll)$ is called a C-space iff $\to$ and $\ll$ coincide with $\overset{C}{\to}$ and $\overset{C}{\ll}$.

Neither simple expressions for $\overset{C}{\to}$ and $\overset{C}{\ll}$, nor criteria for X being a C-space, can be found as in PR. 16, 1.3, and PR. 21, 1.3, for the time being. We will be able to state some conclusions but let us develop an appropriate machinery.

DFN. 30 The subset $L = [p,q] \subset X$, X a causal space is called a link iff L is a non-trivial chain and $p \to q$.

By PR. 26 a link is linearly ordered by the horismos too; by THR. 10, 1.3, it is therefore a subset of some beam.

DFN. 31 Given $Y \subset X$, X a causal space; a linkage of Y from p to q is a finite sequence $x_i / i = 1,2,\ldots, n \ni x_1 = p,\ x_n = q$, $[x_i, x_{i+1}]$ is a link and $[x_i, x_{i+1}] \subset Y \quad i = 1,2,\ldots, n-1$.

DFN. 32 Y is called linked iff $\forall\, p,q \in Y \ni p < q\ \exists$ a linkage of Y from p to q.

A further refinement of THR. 10 can be achieved for the regular spaces since in this case a beam is a linked chain. This can be seen from the following considerations: In case of a proper beam B i.e. of a maximal hypergirder any two points belong to some girder by DFN. 22, 1,3, hence $\forall\, a,b \in B \Rightarrow a,b$: are proximate. The space being regular by THR. 9, 1.3, a,b belong to precisely one proper beam. Moreover $[a,b]$ is a link because if $x \in [a,b]$ i.e. if x strictly lies between a and b by PR. 22, x belongs to the same girder with a and b. Given two such x's i.e. $x_{1,2} \in [a,b]$ since they belong to the same beam they are ordered by the causality i.e. $x_1 < x_2$ or $x_2 < x_1$, i.e. $[a,b]$ is linearly ordered by the causality and by PR. 26, 1.3, DFN. 30, 1.3, $[a,b]$ is a link hence B is a linked chain. In case of $B = \{a,b\}$ $a \neq b$, $a \to b$ and B not belonging to any girder $\Longrightarrow [a,b] - \{a,b\} = \phi$ and $[a,b]$ is a trivial chain. Therefore:

THR. 11 In a regular space any two points ordered by the horismos belong to some linked chain.

Note that THR. 11, 1.3, may at first sight make superfluous the assumption of fullness in PR. 25, 1.3. This is not so, since $L = [p,q]$ need not be a subset of the beam which contains p and q and hence not a causal chain to apply PR. 26, 1.3.

Finally come back to the construction of a C-space.

PR. 27 $\overset{C}{\to} = \bigcup\{\text{reg.hor/cau} <\}$, $\overset{C}{<<} = \bigcap\{\text{reg.chr/caus} <\}$

PROOF E.H. KRONHEIMER, R. PENROSE, 1, page 489.

Note that a C-space need not be regular; but a sufficient condition for a regular causal space to be a C-space is that $x \overset{C}{\to} y$ implies $x \to y$ (as $x \to y \Longrightarrow x \overset{C}{\to} y$ by PR. 27).

As a corollary of THR. 11, 1.3, we get a criterion for regular u-spaces:

PR. 28 If a regular space is a u-space then its underlying set is linked. The converse is true without the space being regular.

PROOF Let $x, y \in X$, X regular and u-space. Suppose $x < y$; then $x \overset{u}{<} y$ and by DFN. 9, 1.3, $\exists\, x_i / i = 1,2,\dots,n \ni x = x_1 \to x_2 \to \dots \to x_n = y$. Since X is regular $[x_i, x_{i+1}] / i = 1,2,\dots,n-1$ are links by THR. 11, 1.3, and therefore there is a linkage of X from x to y whenever $x < y$ i.e. X is linked by DFN. 32, 1.3.

Conversely suppose X is linked. X is a u-space iff $\overset{u}{<}$ $(\overset{u}{<<})$ coincide with $<$ $(<<)$. Given that $\overset{u}{<} = \bigcap\{\text{cau/hor} \to\}$ i.e.

$x \overset{u}{<} y \Rightarrow x < y$ it is enough to show that: $x < y \Rightarrow x \overset{u}{<} y$

Let $x < y$; X being linked implies the existence of a linkage $\{x_i\}$ from x to y and hence $x = x_1 \to x_2 \to \ldots \to x_n = y$ i.e. $x \overset{u}{<} y$.

E.H. KRONHEIMER and R. PENROSE devote a whole chapter to the null future ($E^+(A) = J^+(A) - I^+(A)$) and its intersection properties with causal chains. In view of the most common application of the abstract causal spaces theory, that on the causal structure of manifolds, where these results are more easily accessible through direct calculations, we stop at this point.

CHAPTER 2

2.1 Introduction

For about fifty years general relativity theory was rather a field for applied mathematicians than a description of nature. The situation has greatly changed with the identification of extremely distant extragalactic radio sources (quasars), with the discovery of a background radiation (in the microwave range), and with an improvement of radar techniques for precision measurements within the solar system. An interpretation of such experiments demands more than an understanding of first order post-Newtonian corrections to planetary kinematics. If the history of our universe is to be postdicted from experiments and predictions are to be made, we have to understand the large scale consequences of Einstein's theory.*

A characteristic feature of general relativity considered as a field theory is that the solutions of the field equations provide the metric tensor of the spacetime itself. Consequently, on the one hand one cannot speak of a spacetime (in the presence of a gravitational field) without solving the field equations while on the other hand, almost every problem one faces in the solution of the field equations has a counterpart in the structure of spacetime itself.

It seems natural to start with an abstract mathematical model for our spacetime and study the various underlying and/or primitive substructures, hoping to discover those features of spacetime geometry which are independent of the field equations.

In the following, we concentrate on the possibilities of defining causal relations (in the sense of Chapter 1, Section 3, on our model and study their properties.

* see: W. Kundt, Recent Progress in Cosmology, Springer Tracts in Modern Physics, Vol. 47, 1968 and
C.W. Misner, K.S. Thorne, J.A. Wheeler, Gravitation, W.H. Freeman and Co., 1973.

2.2 A Model for Spacetime

DFN. 1 A spacetime is a real, four dimensional, connected, Hausdorff and C^{∞}-differentiable manifold with a globally defined C^{∞}- tensor field g of type (0,2) which is nondegenerate and Lorentzian.

Nondegenerate means: $\forall$ p $\in$ M and $\forall x \in T_p(M)$ $g(x,y) = 0 \Rightarrow y = 0$.

Lorentzian (or hyperbolic normal) means that $\forall x \in M$ $\exists$ a basis in $T_p(M)$, the tangent space to M at p, relative to which g_p has the matrix form: diag (1, -1, -1, -1).

In an attempt to justify the above choice we remark the following.

The connectedness condition is imposed because one would like to think that spacetime consists of one piece so that communication is possible.

The C^{∞}-differentiable manifold structure is a rather natural choice on the assumption that the universe "looks the same" everywhere. This is because the group of automorphisms of a connected topological (C^0) manifold acts transitively i.e. $\forall$ x, x' $\in$ M there is a homeomorphism which carries any neighbourhood of x onto a neighbourhood of x'. The degree of differentiability (finite, infinite, real analytic) is a subtle point indeed; form a physical point of view (namely propagation of data for the Einstein's vacuum field equations) infinitely differentiable initial data on some spacetime hypersurface imply infinitely differentiable solutions, while solutions whose initial data are k-times differentiable ($k < \infty$) do lose derivatives, but not significantly i.e. their integrated squared derivatives up to order k stay bounded for any k ($k \leq \infty$) when bounded for the initial data*; finally from a purely mathematical point of view note that real analytic manifolds are not so easy to handle because real analytic functions are problematic (e.g. there are no real analytic partitions

* Proceedings of the Thirteenth Biennial Seminar of the Canadian Mathematical Congress on: Differential Topology, differential geometry and applications, edited by J.B. Vanstone, Canad. Math. Congress, Montreal, 1972.

of unity); anyway each C^k-structure is C^k-equivalent (i.e. C^k-compatible)* to a C^{ω}-structure** (or the weaker statement that each C^k-structure is C^k-equivalent to a C^{∞}-structure).

The T_2-separability (Hausdorff property) is imposed mainly for mathematical convenience as can be seen from the following propositions. A topological manifold is already a T_0 and T_1 topological space.

PR. 1 A connected and Hausdorff topological manifold is a T_3 (regular) topological space.

PROOF $\forall$ $x \in M$ choose a chart (U, f) at x (i.e. U open nbd. of x, $f: U \longrightarrow \mathbb{R}^n$ is a homeo). Then $U \cap V$, where V is an open nbd. of x, is an open nbd. of x too and $f/_{U \cap V} : U \cap V \to \mathbb{R}^n$ is a homeo. onto an open subset of $\mathbb{R}^n$. Choose $\varepsilon > 0$ such that the closed ε-ball $\overline{B}_{\varepsilon}(f(x))$, $\overline{B}_{\varepsilon}(f(x)) \subset f(U \cap V)$ (because of the regularity of $\mathbb{R}^n$); $C = f^{-1}(\overline{B}_{\varepsilon}(f(x))$ is closed in $M, C \subset V$ and $\dot{C}$ is a nbd. of x as $B_{\varepsilon}(f(x))$ contains an open nbd. of f (x).

PR. 2 A connected, Hausdorff differentiable manifold with a connection is second countable.

PROOF See R. GEROCH, 1, Appendix.

Consequently our spacetime is separable; the proof goes as follows: choose a point out of each member of the countable base thus obtaining a countable set A. The complement of the closure of A is an open set which, being disjoint from A contains no nonvoid member of the base and is hence void.

The main property of second countable spaces is that they are Lindelöf i.e. every open covering has a countable subcovering. By a theorem of K. Morita (J. DUGUNDJI, VIII, 6.5, page 174) in Lindelof spaces the

* Two atlases A and B are C^k - compatible iff $A \cup B$ is a C^k-atlas.

** H. Whitney - Differentiable manifolds. Ann. of Math. 37, pp.645-680, 1956.
J. Munkres - Elementary differentiable topology. Ann. of Math. Studies 54, § 4, Princenton.

concept of regularity and paracompactness are equivalent. In general, paracompactness implies not only regularity but normality as well (J. DUGUNDJI, VIII, 2.2, page 163). Hence our spacetime is paracompact and normal in its manifold topology.

Notwithstanding, mathematical convenience is not the sole reason for imposing the Hausdorff property. From an operational point of view, as it is impossible to reduce errors in a measuring process to zero, one must always say that a measuring process maps a physical condition onto a neighbourhood of the space of parameters used to characterize the measurements' values; so one must be able to state that the (respective) minimal nbd.s of two measured values do not overlap (i.e. they are disjoint, in order to be able to assert that the two measured values are distinct.. Given that minimal nbd.s must be considered open, the Hausdorff property must be assumed.* Finally P. Hajicek** has shown that reasonably defined non-Hausdorff spacetimes exhibit causal anomalies.

PR. 3 Any connected, Hausdorff and paracompact differentiable manifold M admits a global Riemannian metric tensor field.

PROOF The manifold being paracompact ensures the existence of a partition of unity on M i.e. of a pair((V),(F))where (V) is a locally finite covering of M and $(F) = \{f_V : V \in (V)\}$ is a collection of real valued, C^∞, functions on M such that each $f_V \geq 0$ $\forall\, V \in (V)$, the support of f_V (= closure of the set $\{x \in M : f_V(x) \neq 0\}$) is contained in V and $\sum_{V \in (V)} f_V = 1$ (which makes sense since for each $x \in M$ $f_V(x) = 0$ for all but finitely many $V \in (V)$. Then consider each $V \in (V)$ as a coordinate neighbourhood and define a Riemannian C^∞ structure R_V by $R_V\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right) = \delta_{ij}$ where x_i are the coordinate functions on V . Then define R on M by $R(x,y) = \sum_{V \in (V)} f_V R_V(x,y)$***.

* For a detailed study see: Michael Cole - Int. J.Theor.Phys. 1, No. 1, pp.115-151, 1968.

** Comm. Math.Phys. 21, 75, 1971.

*** Note that the converse is also true i.e. every Riemannian manifold is paracompact and although paracompactness ensures the existence of a partition of unity the inverse is not true.

DFN. 2 Any tangent vector $X \in T_x(M)$ is said to be timelike (TL), spacelike (SL) or null (N) according as $g_x(X,X)$ $(=g_{x,ab}X^aX^b)$ is positive, negative or zero.

The null-cone at x ($x \in M$), i.e. the set of null vectors in $T_x(M)$, disconnects the TL vectors into separated components.

DFN. 3 A spacetime M is said to be time orientable iff it is possible to make a consistent (continuous) choice of one component of the set of TL vectors $\forall x \in M$. To label the TL vectors so chosen future pointing (future directed) (FD) and the remaining ones post-pointing (past directed) (PD) is to make the spacetime time-oriented.

The Lorentzian (i.e. locally Minkowskian) character is mathematically equivalent to the existence of a nowhere vanishing vector field which is continuous (R. GEROCH, 2, page 79). Once such a vector field has been constructed we can assert that either the manifold is noncompact or that it cannot be simply connected. The proof goes as follows: if the spacetime is simply connected, then the generalized STOKES' theorem assures us that there exists on M a single valued scalar function of which X is the gradient field. But if M is compact this function must assume both its maximum and minimum values on M and at these extreme points the gradient must vanish. This contradicts the hypothesis that X is nowhere zero.

There is another reason to exclude a compact spacetime and this is the fact that in a compact model there exist closed TL curves (R.W. BASS, L. WITTEN). (This is a causal anomaly and a proof based on purely causal considerations will be given at a later stage.)

PR. 4 Any non compact spacetime (DFN. 1, 2.2) is time orientable (DFN. 3, 2.2).

PROOF By PR. 3, 2.2, our spacetime carries a global Riemannian metric tensor field R. Being Lorentzian too, it admits a non-vanishing continuous vector field V if M is non-compact and has dimension greater than or equal to two. (L. MARKUS and N. STEENROD, § 39.6, § 39.7, § 40.10). The tensor field $\mathcal{L}$ defined by: $\mathcal{L}(X,Y) = R(X,Y) - R(X,V)\cdot R(Y,V) / 2R(V,V)$ is Lorentzian and V is nowhere vanishing and TL with respect to it. Consequently $(M, \mathcal{L})$ admits a nowhere vanishing, continuous and TL vector field and by DFN. 3, 2,2, is time-oriented.

DFN. 4 A path in M is a C^0 map $\alpha: I \to M$ ($I \subset \mathbb{R}$ is an interval) and its image $\alpha(I)$ specifies a curve.* Any inverse image $\omega : \alpha(I) \to \mathbb{R}$ is called a parameterization of the curve.

Piecewise smoothness (C^∞-differentiability) and continuity are defined as usual. Future (past)-directed (FD(PD)), timelike (TL), causal (C) and null (N) piecewise smooth paths are characterized by their tangent vectors; at a join if $t_1^{\,a}$, $t_2^{\,b}$ are the tangent vectors one further assumes for a C-path that $g_{ab}\, t_1^{\,a}\, t_2^{\,b} \geq 0$. Similar definitions for curves can be assigned through their defining paths.

DFN. 5 An end-point p of a path α (or its associated curve) is defined by: $\forall \{x_i / i = 1,2,\ldots\}$ (a sequence) in I such that $x_i \overset{\lim}{\to} a(b)$, where $a = \inf I$ ($b = \sup I$) $\Rightarrow \alpha(x_i) \overset{\lim}{\to} p$.

For FDTL (FDC) paths past and future end-points are defined by a and b respectively. (Dually for PDTL (PDC) paths). TL (C) curves need not contain their end-points (and hence be closed) unless I is closed. Whenever a future (past) end-point does not exist the path (curve) is called future (past) endless.

DFN. 6 A TL(C) trip is a FD piecewise TL(C) geodesic. A trip from $x \in M$ to $y \in M$ is a trip with past end-point x and future end-point y.

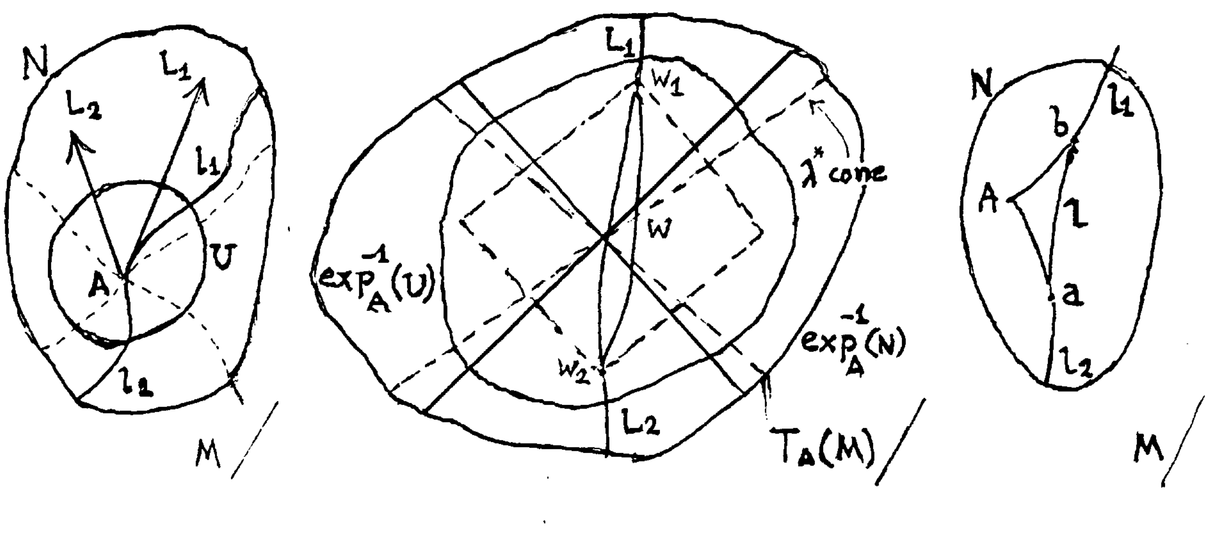
DFN. 7 $\forall x \in M$, the light cone at x, denoted by L_x, is the set: $\{y \in M : y \text{ lies on a null geodesic through } x\}$.

* More rigorously a curve should be the equivalence class of paths equivalent under parameter change (i.e. homeo. or diffeo. of the path domain).

THR. 1 Suppose that a FDTL or FDC trip exists from x to y ($x, y \in M$, M a spacetime). Then a smooth FDTL path exists from x to y, unless the FDC trip is a single null geodesic. Provided a FDTL or FDC smooth path exists from x to y, then a FDTL trip exists from x to y unless the FDC smooth path is a null geodesic.

PROOF Let a trip exist from x to y with a joint at A. We will show that the joint can be smoothed (the piecewise geodesic be replaced by a smooth curve). There is a simple region N containing A and no other joints, since the set of joints accumulate nowhere i.e. form a discrete set. Let l_1 and l_2 be TL(C) geodesics meeting at A and let L_1, L_2 be the tangents to l_1 and l_2 respectively at A.

I. Suppose that both the tangents to l_1 and l_2 are not null. Choose the one, L_1 say, to be the time axis in a new coordinate system (t, x, y, z). The locus $t^2 = \lambda\cdot(x^2 + y^2 + z^2)$, $\lambda > 1$ is a cone whose generators are TL lines. Let L_2 lay on $t^2 = \lambda\cdot(x^2 + y^2 + z^2)$. Choose λ^* so that $\lambda_2 > \lambda^* > 1$; then the cone $t^2 = \lambda^*(x^2 + y^2 + z^2)$ contains both L_1 and L_2 in its interior. Construct a copy of this cone at every point in $\exp_A^{-1}(N)$; since the image of the light cone at A under $\exp_A^{-1}$ contains the λ^*-cone and since the light cones in N

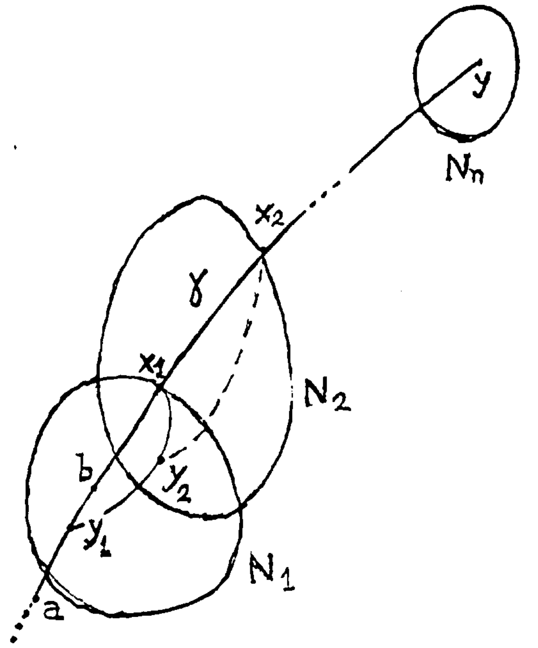


are a smoothly varying family of hypersurfaces, there exists an open nbd. U of A, $U \subset N$ such that the image of the light cones under $\exp_A^{-1}$ $\forall p \in U$ contains the λ^*-cone at $\exp_A^{-1}(p)$. Choose W_1, W_2 on L_1, L_2 respectively in $\exp_A^{-1}(U)$. Since the lines L_1, L_2 lay within the λ^*-cones at W_1 and W_2 respectively the joint can be replaced by smooth curve W whose tangent vector at each point lies in the λ^* cone at that point and which joins smoothly L_1 and L_2 at W_1 and W_2. By construction $\exp_A(W)$ is TL smooth.

II. In case that l_1 and l_2 possess null tangents at A (of different direction to ensure the existence of the joint) the above construction of λ^* cones cannot be carried out and to prove the result (mainly the case for a FDC-trip) we resort to the following: Consider a and b on l_1 and l_2 respectively, in N. Consider also the function $\Phi(x) = \Phi(a,x)$, the world function of a and x, as x varies from a to A on l_2 and from A to b on l_1. I will show that $\Phi(a,x) > 0$ when x is past A (on l_1), which means that the geodesics ax, $x \in l$, (and consequently the geodesic ab) are TL. First of all as x proceeds in future causal direction the geodesics ax are FD. More than this $\Phi(a,x) \geq 0$ and the rate of change of $\Phi(a,x)$ as measured along a FDC vector field T_i is given by: $T^j \nabla_j \Phi = g^{ij} T_i \nabla_j \Phi$ which is non-negative. But since x is past the point A and since $\Phi(a,x)$ is continuous $\Phi(a,x)$ has to be strictly positive; this is because $g^{ij} T_i \nabla_j \Phi = 0$ means that T_i and $\nabla_i \Phi$ must be null and proportional, which at least is violated at A where $\nabla_j \Phi$ can be identified as the tangent to l_2 at A and T_j as the tangent to l_1 at A (in the direction in which x proceeds), a contradiction as at A there is a joint. Surely the above is equally well true with one of the tangents at A being TL. So a

FDC trip exists from x to y with at least one TL part (from a to b or from W_2 to W_1).

III. The FDC-trip (γ) from b to y, being compact, can be covered by a finite number of simple regions N_1, $N_2 \ldots, N_n$. Let x_1 be the future end-point of $\gamma \cap \overline{N}_1$ from b. Choose



$y_1 \in N_1$ on γ with $y_1 \neq b$. Then apply II above; y_1x_1 is FDTL. Now either $x_1 = y$ or $x_1 \in N_2$ in which case let x_2 be the future end-point of $\gamma \cap \overline{N}_2$ from x_1; choose $y \in N_2$ on y_1x_1 with $y_2 \neq x_1$. Then either $x_2 = y$ or we can repeat the construction. The process must eventually terminate, since there is a finite number of connected components $\gamma \cap \overline{N}_i$. So a FDTL trip exists from b to y. The same procedure applied backwards from a to x gives us eventually a FDTL trip from x to y. Then part I provides us with a FDTL smooth curve from x to y.

So in case a joint exists applying I. or II. a TL part always can be inserted in our FDC trip and applying III a FDTL trip joins x to y. Then I. gives the FDTL smooth curve. Given that no joints exist our FDC trip is already a geodesically FDC smooth curve which is either TL or N and this proves the first part of the theorem.

To prove the second part of the theorem suppose that that a FDTL smooth path (γ) exists from x to y. The corresponding curve being compact can be covered by a finite number of simple regions N_1, N_2, $\ldots, N_n$. Let x_1 be the future end-point of $\gamma \cap \overline{N}_1$ from x. The geodesic from x to x_1 is FDTL as $\exp_x^{-1} y$ (with

$y \in \gamma$ between x and x_1) never leaves the future component of the TL vectors at $T_x(M)$, and repeating the argument we end up with a FDTL trip from x to y.

Given a FDC smooth path and repeating the above construction we get a FDC trip. Applying the first part of the theorem, unless the given path is a single null geodesic, we get a FDTL smooth curve and again applying the above we get a FDTL trip.

2.3 The Causal Structure

Causal relations can be defined on our spacetime using either trips or smooth curves in an equivalent fashion by virtue of THR. 1, 2.2.

DFN. 8 Let $x, y \in M$, M a spacetime.

$x < y$ iff a FDC trip exists from x to y.

$x << y$ iff a FDTL trip exists from x to y.

$x \rightarrow y$ iff $x < y$ and $x \not<< y$.

PR. 5 if $x \rightarrow y$ then a FD single null geodesic exists from x to y.

PROOF Let $x \rightarrow y$ i.e. $\exists$ a FDC trip and $\nexists$ a FDTL trip from x to y; by THR. 1, 2.2, either a FDTL smooth path exists from x to y or a single null geodesic. By THR. 1, 2.2, the FDTL smooth curve gives a FDTL trip which has been excluded by our hypothesis; hence a FD single null geodesic joins x to y.

Note that the converse of PR. 5 is not true in general.

DFN. 9 Let $x, y \in M$, M a spacetime.

$x \underset{C}{<} y$ iff a FDC smooth path exists from x to y.

$x \underset{C}{<<} y$ iff a FDTL smooth path exists from x to y.

$x \underset{C}{\rightarrow} y$ iff a FD null geodesic joins x to y and $\nexists$ a FDTL smooth path from x to y.

PR. 6 The relations $\underset{C}{<}$, $\underset{C}{<<}$, $\underset{C}{\rightarrow}$ coincide with the relations $<$, $<<$, $\rightarrow$.

PROOF (Straightforward using THR. 1, 2.2).

Although curves and trips are equivalent in this sense, they motivate two technically different approaches in any subsequent development. There are two major publications respectively by S.W. Hawking and G.R.R. Ellis (S.S. HAWKING, 1, G.R.R. ELLIS) and R. Penrose (R. PENROSE, 2), which combine the results of a number of previous publications started around the mid-sixties. Both include a treatment of systems of curves (trips), of the Jacobi fields and conjugate points, and of the existence of geodesics as maximal curves. An account of all these, for geodesics only, can be found in N.M.J. Woodhouse's thesis (N.M.J. WOODHOUSE, 1).

To be consistent, we adopt DFN. 8, 2.2, in terms of trips, throughout the rest, unless otherwise stated.

A spacetime M endowed with the relations defined above (DFN. 8, 2.2 and DFN. 9, 2.2) is a causal space in the sense of E.H. Kronheimer and R. Penrose (DFN. 8, 1.3) provided that no closed (self-intersecting FDC paths (trips) exist).

The fact that no closed FD (or PD) TL (resp. C) trips are allowed can be stated as the chronology (resp. causality) conditions; these are the first two conditions in a causality condition hierarchy.

DFN. 10 We say that our spacetime is a causal (chronological) (or it satisfies causality (chronology)) iff it does not contain closed FDC (FDTL) and PDC (PDTL) trips.

PR. 7 Causality implies chronology.

Note that violation of causality does not necessarily imply violation of chronology i.e. there may be closed null geodesics (e.g. identify two null hypersurfaces $t-x = t_i / i = 1,2$ in Minkowski spacetime).

PR. 8 A spacetime M considered as a causal space is a regular causal space.

PROOF According to the definition of regularity (DFN. 15, 1.3) it is enough to prove that given $x_{1,2}$, $y_{1,2}$ such that $x_{1,2} \rightarrow y_{1,2}$

and $y_1 \to y_2$ then $x_1 \to x_2$. Let $x_1 << x_2$ (or $x_2 << x_1$) then $x_1 << y_1$ (or $y_1 << x_1$) a contradiction to $x_1 \to y_1$. By DFN. 9, 2.3, and PR. 6 $x_1 \to y_1 \to y_2$ ($x_2 \to y_1 \to y_2$) mean that FD null geodesics join x_1 to y_1 and y_1 to y_2 (x_2 to y_1 and y_1 to y_2). These geodesic paths form a single geodesic because otherwise by THR. 1, 2.2, and DFN. 8, 2.2, $x_1 << y_2$ ($x_2 << y_2$) a contradiction to $x_1 \to y_2$. The portion of these geodesics between y_1 and y_2 coincide and these two geodesics ($x_1y_1y_2$ and $x_2y_1y_2$) must be two portions of the same geodesic i.e. $x_1 \to x_2$ or $x_2 \to x_1$.

PR. 9 $I^+(A)$ ($I^-(A)$) is open in the manifold topology $\forall A \subset M$.

PROOF see R. PENROSE, 2, Proposition 2.8.

PR. 10 The chronology relation (<<) defined in a spacetime (DFN. 8, 2.3) is full (THR. 1, 1.3).

PROOF By PR. 9, 2. 3, $I^{\pm}(x)$ $\forall x \in M$ are open, hence $\exists\, y_{1,2} \ni y_1 \in I^+(x)$ i.e. $x << y_1$ and $y_2 \in I^-(x)$ i.e. $y_2 << x$. Also if $y_{1,2} \in I^-(x) \Rightarrow x \in I^+(y_{1,2})$ but $I^+(y_1) \cap I^+(y_2)$ is open and $x \in I^+(y_1) \cap I^+(y_2)$. Choose $y \in I^-(x) \cap I^+(y_1) \cap I^+(y_2)$ to get $y_{1,2} << y << x$.

Recalling the distinguishing properties of the chronology relation (1.3) we define:

DFN. 11 A spacetime M is future (past) distinguishing at $x \in M$ iff $I^+(x) \neq I^+(y)$ ($I^-(x) \neq I^-(y)$) $\forall y \in M \ni x \neq y$.

DFN. 12 A spacetime is future (past) distinguishing iff M is future (past) distinguishing at every point.

DFN. 13 A spacetime M is weakly distinguishing iff $\forall x, y \in M$, $x \neq y$ $\Rightarrow$ either $I^+(x) \neq I^+(y)$ or $I^-(x) \neq I^-(y)$ (i.e. $I^{\pm}(x) = I^{\pm}(y) \Rightarrow x = y$).

PR. 11 A future and past distinguishing spacetime is future or past distinguishing; and a future or past distinguishing spacetime is weakly distinguishing.

PR. 12 A weakly distinguishing spacetime (DFN. 13, 2.3) is causal (DFN. 10, 2.3).

PROOF Because of proposition 3.8 in R. PENROSE, 2, (i.e. $x < y$ $I^+(y) \subset I^+(x)$) no spacetime containing closed causal trips can be either future or past distinguishing.

Although the causality and chronology restrictions (DFN. 10, 2.3) do not allow for closed trips one would like to rule out the possibility for trips leaving the vicinity of a point and then returning close to it, even though an actual closed trip need not be the result. This is partially achieved with the future (past) distinguishing properties if one realises the fact that DFN. 11, 2.3, is equivalent to:

DFN. 14 A spacetime M is future (past) distinguishing at $x \in M$ iff every nbd of x contains a nbd. of x which no FD (PD) causal trip from x intersects more than once.

There is in fact an infinite hierarchy of such higher degree causality conditions as has been pointed out by B. CARTER. First we extend our definition of causal relations.

DFN.15 Let A, B and C be subsets of M where M is a spacetime. Then it will be said that B lies in the causal future of A with respect to C (and write $A \underset{C}{<} B$ iff from every point of B there is a PDC trip contained entirely in C which intersects some point of A (qualified causality relation (B. CARTER, page 353)).

Dual definitions and symbols as well as chronology relations (qualified) are given in B. CARTER but we will not use them here.

DFN.16 Let A, B be subsets of M. The causal future of A with respect to B is given by: $J_B^+(A) = \{x \in M: A \underset{B}{<} x\}$

Whenever no reference set is mentioned it is understood that it is the whole spacetime M.

Qualified causal relations for points, whenever the reference set is open and hence a spacetime manifold in its own right (or, if C is not connected, a disjoint union of spacetime manifolds), possess the same properties as unqualified ones. In particular, the sets $I_A^+(x) \cap I_A^-(y) = \{z \in M: x \underset{A}{<<} z \underset{A}{<<} y\}$ are open if A is open. We recall the notation: $I_A^+(x) \cap I_A^-(y) = < x,y >_A$ (DFN. 12, 1.3); $< x,y >$ stands for $< x,y >_M$.

Since it is not always true that $J^+(A)$ is a closed set one can, performing closure operations of causal futures (and pasts), end up with non trivial extensions.

DFN.17 Let $A \subset M$. The first degree future of A is defined as

$$\overset{1}{J}{}^+(A) = \{x \in M: J^+(A) \cap \overline{J^-(x)} \cup \overline{J^+(A)} \cap J^-(x) \neq \emptyset\}$$

DFN.18 We say B belongs to the first degree future of A iff

$$B \subset \overset{1}{J}{}^+(A) \quad (A \overset{1}{<} B)$$

Nth degree causality relations can be defined by induction.

DFN.19 Let A and B be subsets of M and let A contain at least two points. A is called virtuous with respect to B iff $\forall\, x, y \in A$ with $x \neq y$ at most one of the relations: $x \underset{B}{<} y$ and $y \underset{B}{<} x$ is satisfied.

A is called globally virtuous iff B = M.

A is called virtuous to the n^{th} degree $(n \geq 0)$ iff no pair $x, y \in A$ satisfies $x \overset{r}{<} y$ and $y \overset{s}{<} x$ with $r+s \leq n$.

DFN.20 A spacetime M is called n^{th} degree causal iff M is globally virtuous to the n^{th} degree.

Zero degree virtue is equivalent to the causality condition.

PR. 13 A future and past distinguishing spacetime M is 1st degree causal and vice versa.

PR. 14 Second degree virtue is equivalent to the fact that $\forall\, x \in M$ and every nbd. N_x of x there exist a nbd. of x contained in N_x which no causal trip intersects more than once.

DFN.21 A spacetime which is 2nd degree causal is called strongly causal.

One can obviously restrict the definition of strong causality just to a point.

So far every reference to topological properties was made with respect to the manifold topology; but causal spaces can be topologized intrinsically (i.e. using only the causal relations) as follows. We define J^* to be the smallest topology on M in which $I^+(x)$ are open $\forall\ x \in M$. J* is called the Alexandrov topology. Incidentally $I^{\pm}(x)$ are open in the manifold topology J^M (PR. 9, 2.3) and hence $J^* \subset J^M$.

One is interested in a base for J^* and under which circumstances J^* agrees with J^M. By THR. 1, 1.3 and PR. 10, 2.3, the set (A) = $\{I^+(x) \cap I^-(y),\ x,\ y \in M\}$ is a base for J^*. It turns out that the two topologies agree iff the spacetime is strongly causal; but first we have to introduce some new technicalities.

DFN.22 An open subset A of M is called causally convex iff A intersects no TL trip in a disconnected set or equivalently $\forall x,\ y \in A$ and $x \ll z \ll y \Rightarrow z \in A$ i.e. $\forall\ x,\ y \in A \Rightarrow \langle x,y \rangle \subset A$.

e.g. the sets $\langle x,y \rangle_A$ are causally convex; also a spacetime would be called strongly causal at $x \in M$ iff x has arbitrarily small causally convex nbd.s. This is because:

PR. 15 Given $x,\ y \in N$ (a simple region), no C trip lying in M can intersect $\langle x,\ y \rangle_N$ in a disconnected set.

PROOF R. PENROSE, 2, Proposition 4.8.

Consequently $\langle x,y \rangle_N$ are causally convex in N.

PR. 16 Given A an open subset of a simple region N and $x \in A$ then there exist $x_{1,2} \in A \ni x \in \langle x_1,\ x_2 \rangle_N \subset A$

PROOF R. PENROSE, 2, Proposition 4.9.

DFN.23 A local causality nbd. is a causally convex open set whose closure is contained in a simple region.

PR. 17 A spacetime is strongly causal at x iff x is contained in some local causality nbd.

PROOF R. PENROSE, 2, Proposition 4.12.

PR. 18 The following restrictions on a spacetime M are equivalent;

(a) M is strongly causal

(b) J^* agrees with J^M.

PROOF (a) $\Rightarrow$(b): Since $J^* \subset J^M$ it is enough to show that every nbd. N_x of $x \in M$ in the manifold topology contains an Alexandrov nbd; choose a simple region $N \ni x \in N \subset N_x$; by strong causality a causally convex, open, set 0 containing x exists in N and is a local causality nbd as required by DFN. 23, 2.3. By PR. 16, 2.3, $\exists\ x_{1,2} \in 0 \ni x \in < x_1, x_2 >_N \subset 0$. But if $< x_1, x_2 >_N \neq < x_1, x_2 >$ (an Alexandrov nbd.) then there exists a trip from x_1 to x_2 which leaves and re-enters N. Thus it would have to leave and re-enter 0 also, violating its causal convexity. Thus $x \in < x_1, x_2 >$, $< x_1, x_2 > \subset 0 \subset N \subset N_x$.

(b) $\Rightarrow$(a): $\forall\ x \in M$ and every nbd N_x of x in J^M, there is an Alexandrov nbd contained in N_x; it is a causally convex set and (a) follows from DFN. 21, PR. 14, and PR. 15, 2.3.

Note that since J^M is assumed Hansdorff the Alexandrov topology is Hansdorff for strongly causal spaces. The converse is true but a proof requires a study of the strong causality failure regions.

<u>Futures, Pasts, Cauchy developments, horizons</u>

Since a spacetime M is a full causal space (PR. 10, 2.3) theorems 3, 4, 5 and 6, and PR. 17, 18 of Chapter 1 are true. These results cover part of the properties of futures and pasts.

PR. 19 Let Q(P) be subsets of a spacetime M. Then the following relations (and their duals) are equivalent:

1. $I^+(Q) \subset Q$	1'. $I^-(P) \subset P$
2. $I^-(-Q) \subset -Q$	2'. $I^+(-P) \subset -P$
3. $I^+(Q) \cap I^-(-Q) = \emptyset$	3'. $I^-(P) \cap I^+(-P) = \emptyset$
4. int $Q = I^+(Q)$ and int $-Q = I^-(-Q)$	4'. int $P = I^-(P)$ and int $-P = I^+(-P)$

5. $\partial Q = (-I^+(Q) \cap (-I^-(-Q)$ 5'. $\partial P = (-I^-(P)) \cap (-I^+(-P))$

6. $I^+(Q) \subset Q \subset J^+(Q)$ 6'. $I^-(P) \subset P \subset J^-(P)$

PROOF By PR. 18, 1.3, $1 \Leftrightarrow 2 \Leftrightarrow 3 \Leftrightarrow 1$. By PR. 17, 1.3, $1 \Leftrightarrow 6$.

Introducing the J^* topology (since $J^* \subset J^M \Rightarrow \forall A \subset M$ $\text{int}^* A \subset \text{int } A$, $\bar{A} \subset \bar{A}^*$, $\partial A \subset \partial^* A$) by THR. 5 and THR. 6, 1.3, $\text{int}^* Q = I^+(Q)$ and $\partial^* Q = B^+(Q) = \{x \in M: I^+(X) \subset Q \text{ and } I^-(x) \subset -Q\}$ $\forall Q \subset M \ni I^+(Q) \subset Q$. But $B^+(Q) = \{x \in M: I^+(x) \subset Q\} \cap \{x \in M: I^-(x) \subset -Q\}$ $= \{x \in M: y \in I^+(x) \Rightarrow y \in Q\} \cap \{x \in M: y \in I^-(x) \Rightarrow y \in -Q\} = (-\{x \in M: y \in I^+(x) \Rightarrow$ $y \notin Q\}) \cap (-\{x \in M: y \in I^-(x) \Rightarrow y \notin -Q\}) = (-\{x \in M: x \in I^-(y) \text{ and } y \notin Q\}) \cap$ $\cap (-\{x \in M: x \in I^+(y) \text{ and } y \in Q\}) = (-I^-(-Q) \cap (-I^+(Q))$; consequently 1 (or 2 or 3 or 6) imply 4 and 5 in the Alexandrov topology.

4. $\Rightarrow$1. (rather trivially) since $I^+(Q) = \text{int}^* Q \subset \text{int } Q \subset Q$

5. $\Rightarrow$4. $\partial Q = (-\text{int } Q) \cap (-\text{int } (-Q)) = -(I^+(Q) \cup I^-(-Q)$ i.e. $\forall x \in M$: $x \in \partial Q \Leftrightarrow x \notin I^+(Q)$ and $x \notin I^-(-Q)$ (I) or equivalently [$\forall x \in M$: $x \notin \partial Q$ i.e. $x \in \text{int } Q$ or $x \in \text{int } -Q \Leftrightarrow x \in I^+(Q)$ or $x \in I^-(-Q)$] (P). (P) implies $I^+(Q) = \text{int } Q$ (and $I^-(-Q) = \text{int } (-Q)$ iff $\text{int } Q \cap I^-(-Q) = \emptyset$ (and $\text{int } -Q \cap I^+(Q) = \emptyset$). Indeed, let $x \in \text{int } Q \cap I^-(-Q)$; $\exists$ a trip γ from $q \in -Q$ to $x \in \text{int } Q$ ($x << q$) which has to cross the ∂Q; let $z \in \partial Q \cap \gamma$; then $z \in I^-(-Q)$ contradicting (P). ($4 \Leftrightarrow 5$ trivially).

Obviously the equivalence of 4. and 5. is true independently of the topology used; all one needs is that 1 (or 2 or 3 or 6) imply 4 (or 5) in the manifold topology.

DFN. 24 A subset A of a spacetime M is called a future (past) set iff $A = I^+(B)$ ($I^-(B)$) for some $B \subset M$.

PR. 20 $A \subset M$ is a future (past) set iff $A = I^+(A)$ ($A = I^-(A)$).

PR. 21 Let $A \subset M$ and $A = I^+(A)$. Then:

1. $\bar{A} = \{x \in M: I^+(x) \subset A\}$
2. $\bar{A} = -I^-(-A)$
3. $\partial A = -A \cap (-I^-(-A) = \{x \in M: I^+(x) \subset A \text{ and } x \notin A\}$

PROOF: 1. (R. PENROSE, 2. Proposition 3.3)

2. By 1 of PR. 21 and PR. 19, 2.3.

PR. 22 $A \subset M$, $I^+(A) \subset A$ and A open in M. Then A is a future set.

PROOF: By PR. 19, 2.3.

PR. 23 The union of any number of future (past) sets is a future (past) set; the intersection of finite number of future (past) sets is a future (past) set.

PROOF: (R. PENROSE, 2, Proposition 3.7).

DFN.25 A subset A of a spacetime M is called achronal (or semi-spacelike) iff $\forall x, y \in A \Rightarrow x \not\ll y$.

Note that a subset A can be locally spacelike without being achronal. No achronal set need exist if M violates chronology.

In the introduction to Chapter I the concept of the domain of dependence was mentioned; its physical significance lies mainly in the fact that it is precisely that part of the future (past) which is determined by initial data on some hypersurface, preferably an achronal one, provided that local physical laws (the possibility of superluminal field equations has been ignored) are deterministic (the spacetime being locally Lorentzian, the bicharacteristics of the partial differential equations involved should be null geodesics).

DFN.26 Let A be an achronal subset of a spacetime M. The future, past and total domain of dependence of A (or future, past and total Cauchy developments of A) are defined by:

$D^+(A) = \{x \in M$: every PDTL and past endless trip containing x intersects A$\}$

$D^-(A) = \{x \in M$: every FDTL and future endless trip containing x intersects A$\}$

$D(A) = D^+(A) \cup D^-(A)$

The use of curves instead of trips makes no difference as long as A is closed or achronal. Also S.W. Hawking in his early papers and in his book with G.F.R. Ellis (S.W. HAWKING, 1) uses C rather than TL curves and there would be some physical justification for this; such a domain of dependence: $\tilde{D}^+(A) = \{x \in M$: every PDC and past endless curve containing x intersects $A\}$ is related to $D^+(A)$ by:

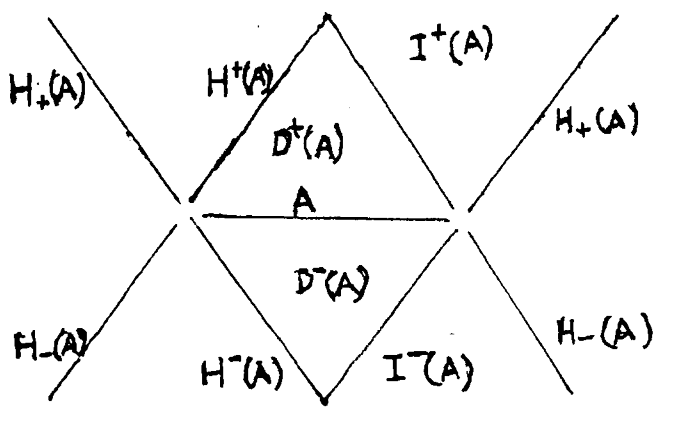
PR. 24 $\tilde{D}^+(A) = \overline{D^+(A)}$

PROOF: (S.W. HAWKING, 1 and G.F.R. ELLIS, Proposition 6.5.1)

For a fairly extensive account of the properties of the domain of dependence see: S.W. HAWKING 1, and G.F.R. ELLIS, R. PENROSE, 2, R. GEROCH, 3.

The boundaries of any $I^{\pm}$ or $D^{\pm}$ are called horizons. Four types of horizons ($H\pm$, $H^{\pm}$) are possible.

e.g. in two-dimensional Minkowski spacetime:



In particular if γ is a TL curve considered as a point set in our spacetime M we define the $\partial I^-(\gamma)$ (= $H^-(\gamma)$) as an event horizon and the $\partial I^+(\gamma)$ (= $H^+(\gamma)$) as a particle horizon; physically, the event horizon of γ separates those events which are observable by an observer whose world line is γ from those events unobservable by him, and similarly the particle horizon of γ separates those events from which a particle with world line γ can be observed from those events from which the particle cannot be so observed.

$H^{\pm}(A) = D^{\pm}(A) - I^{\mp}(D^{\pm}(A))$ and are called future (past) Cauchy Horizons.

The following two concepts are useful in the study of horizons.

DFN.27 A subset B of a spacetime is called an achronal boundary (or semispacelike boundary) iff B is the boundary of a future set.

Note that the definition is time symmetric since $B = \partial I^{+}(A) = \partial I^{-}(-I^{+}(A))$ (R. PENROSE, 2, Proposition 3.14); an achronal boundary is an achronal and closed set which is a C^{0} 3-dimensional submanifold of M (i.e. a continuous imbedded hypersurface) (R. PENROSE, 2, Lemma 3.17 or S.W. HAWKING, 1, and G.F.R. ELLIS, Proposition 6.5.1).

DFN.28 Let A be an achronal and closed subset of a spacetime M edge $A = \{x \in M: \forall$ nbd. N_x of $x \; \exists \; x_{1,2} \ni x_1 \in I_{N_x}^{-}(x)$ and $x_2 \in I_{N_x}^{+}(x)$ and a TL trip from x_1 to x_2 in N_x not intersecting $A\}$. If edge $A = \emptyset$ we call A edgeless (e.g. any achronal boundary in M is edgeless).

Note that $\bar{A} - A \subset$ edge $A \subset \bar{A}$ and consequently if A is edgeless A must be closed; edge A is the set of limits points of A not in A together with the set of points in whose vicinity A fails to be a topological submanifold.

PR. 25 Let $B = \partial I^{+}(A)$, $A \subset M$ $\forall x \in B - \bar{A}$ $\exists$ a null geodesis in B with future end-point x and which is either past endless or has a past end-point on $\bar{A}$. In case that x is the end-point of two geodesics either the one is contained in the other or every extension of either geodesic into the future must leave B and enter $I^{+}(A)$.

PROOF See R. PENROSE, 2. Propositions 3.20 and 3.22.

PR. 26 Let A be achronal. Every $x \in H^{+}(A)$ - edge A is the future end-point of a null geodesic on $H^{+}(A)$ which is either past endless or has past end-point on edge A.

PROOF (R. PENROSE, 2, Theorem 5.12; S.W.HAWKING, 1, and G.F.R.ELLIS,Pro.6.5.3)

So horizons are (pieces of) null hypersurfaces except where they meet $\bar{A}$; analogous statements hold for past horizons. Horizons can contain closed null lines.

2.4 Stable causality

Recalling the causality conditions hierarchy (DFN. 20, PR. 13 and PR. 14, 2.3), it seems that apart from actually closed TL or C trips (curves) (chronology or causality conditions respectively) we have excluded 'almost closed' TL curves (strong causality) and furthermore the case of TL curves which pass arbitrarily close to other TL curves, which then come arbitrarily close to the first curves (higher than second order causality condition). However one may ask for an ultimate causality condition which is stronger than all this hierarchy and which corresponds to spacetimes not being on the verge of violating causality.

Besides, for operational reasons, any physically significant property or condition ought to have some form of stability, that is to say, it should also be a property of 'nearby' spacetimes. In order to give precise meaning to 'nearby' one has to define a topology on the set of all spacetimes. In particular one has to topologize the set $T^{o}_{S2}(M)$ of symmetric, second rank, covariant tensors on M; the set of Lorentz metrics is a subset of $T^{o}_{S2}(M)$ and so will inherit each topology from $T^{o}_{S2}(M)$ (a Lorentzian metric tensor field is a cross section $\hat{g}: M \rightarrow T^{o}_{S2}(M)$ assigning an element g of $T^{o}_{S2}(M)$ at each point $P \in M$ such that $\pi . \hat{g}$ = identity, where π is the projection: $T^{o}_{S2}(M) \rightarrow M$).

There is a number of different topologies that can be placed on $T^{o}_{S2}(M)$; the topologies differ in how many derivatives of a metric have to be near to those of another metric for the two metrics to be considered 'near' to each other and in what region they are required to do so. A sophisticated approach makes use of the bundle of jets over the manifold M;* for an elementary treatment see S.W. HAWKING, 2, and R. GEROCH, 4.

* R. Palais: Foundations of Global Non-Linear Analysis - Benjamin - 1968.

By PR. 3, 2.2, a spacetime M admits a global Riemannian metric tensor field i.e. a (positive definite) metric h_{ab} with respect to which covariant derivatives (∇_a) of tensor fields on M can be defined; with these data we define a distance function on $T^{o}_{S2}(M)$ by:

$$\rho(g_{ab}, g'_{ab}) = \inf_{C} \sum_{n=0}^{p} 2^{-n} \frac{|g-g'|_n}{1 + |g-g'|_n}$$

where C any closed subset of M, p any nonegative integer and

$$|g-g'|_n = \{[\nabla_{a_1}\cdots\nabla_{a_n}(g_{rs}-g_{rs}')]\,[\nabla_{b_1}\cdots\nabla_{b_n}(g_{ar}-g_{ar}')]h^{a_1b_1}\cdots h^{sr}\}^{\frac{1}{2}}$$

Each arbitrary choice of h_{ab} and C defines via the distance function a topology on $T^{o}_{S2}(M)$ (a family of neighbourhoods of $g_{ab}\in T^{o}_{S2}(M)$ consists of all $\{g'_{ab}\in T^{o}_{S2}(M) \ni \rho(g_{ab}, g'_{ab}) < \varepsilon\}\forall\varepsilon > 0$). However we are interested in topologies which are independent of such arbitrary choices; all we have then to do is to specify in some invariant way an appropriate collection of pairs (h_{ab}, C) and consider the aggregate of all finite intersections and arbitrary union of open sets (defined via the distance functions from each pair (h_{ab}, C) of our collection) as a topology on $T^{o}_{S2}(M)$.

Rigorously this can be done as follows: Let $\mathcal{B}$ be the Cartesian product of the set of all positive definite metric tensor fields on M with the set of all nonempty closed subsets of M. Each subset of $\mathcal{B}$ defines a topology on $T^{o}_{S2}(M)$. As the group of diffcomorphisms on M acts as a transformation group on $T^{o}_{S2}(M)$ (and also on $\mathcal{B}$) an invariant topology is one for which these transformations are homeomorphisms (or an invariant topology is defined by a subset of $\mathcal{B}$ invariant under the action of the transformation group).

The C^p-compact open topology for $T^{o}_{S2}(M)$ is defined by the collection of all pairs (h_{ab}, C) with C compact; so metrics are required to be near only on compact regions of the spacetime M. The C^p-open topology for $T^{o}_{S2}(M)$ is defined by the collection of all pairs (h_{ab}, C) with C always taken to be the whole spacetime M; so nearby metrics must be nearby everywhere and must have the same limiting behaviour at infinity. Finally, if $F(g_{ab}, C)$ is the set of all metrics satisfying $\rho(g_{ab}, g'_{ab}) < \epsilon$ and coincide with g_{ab} outside the compact set C, define a nbd. of g_{ab} to be the union of all $F(g_{ab}, C)$ with $C \subset M$ and C compact. The resulting topology is called the fine topology on $T^{o}_{S2}(M)$.

The fine topology is finer than the open topology which in turn is finer than the compact-open topology. As an example we give explicitly the C^{o} open sets; these are defined to be the sets L(U), where U is an open set in $T^{o}_{S2}(M)$ and L(U) consists of all Lorentz metrics g such that $\hat{g}(M) \subset U$.

DFN.29 A spacetime M satisfies the stable causality conditions (equivalently M is called stable) iff the Lorentz metric tensor has an open nbd. in the C^{o} open topology such that M satisfies chronology for any metric belonging to the nbd.

What this condition means intuitively is that one can expand the null cones slightly at every point without introducing closed TL curves.

In general:

DFN.30 A property P of a metric tensor g is stable in a given topology on $T^{o}_{S2}(M)$ iff there is an open nbd of g, every metric tensor of which has the property P.

The following concept would also be in some use in the study of spacetimes (i.e. of the set $T^{o}_{S2}(M)$).

DFN.31 A theorem holds generically or a property is generic in a subset of $T^{o}_{S2}(M)$ iff it holds almost everywhere in that subset (i.e. it holds on an open dense outset of the subset).

Note that, given two topologies $J_1 \subset J_2$, for a property P to be stable in J_1 is a stronger requirement than for P to be stable in J_2. This, together with the fact that in the compact-open topologies any nbd. of any metric tensor g contains metric tensors in which there are closed TL curves (because outside the compact set C the metric tensors

can differ by an arbitrary amount), justify the choice of the open topology in the DFN. 29, 2.4, of the stable causality condition. To use the fine topology results in a definitely weaker condition.
An important consequence of stable causality condition is stated in the following proposition.

PR. 27 A spacetime M satisfies the stable causality condition iff there is a real valued function on M whose gradient is everywhere timelike.

PROOF G.F.R. ELLIS, and S.W. HAWKING, 1, Proposition 6.4.9 and S.W. HAWKING, 3.

The function mentioned in PR. 27, 2.4, can be thought of as a 'cosmic time' in the sense that it increases along every FDTL or FDN curve. Surfaces of constant cosmic time are slices in M i.e. closed, spacelike, 3-dimensional submanifolds without boundary (properly imbedded in M); such slices may be thought of as surfaces of simultaneity in spacetime although they are not unique. If they are all compact they are all diffeomorphic to each other, but this is not necessarily true if some of them are non compact. PR. 27, 2.4, can be rephrased as: a spacetime admits a cosmic time function iff it is stably causal in the C^0-open topology.

Next to stable causality in the causality condition hierarchy is the concept of a causally continuous spacetime (S.W. HAWKING, 4, and R.K. SACHS, 1). But let us first introduce some new concepts:

DFN. 32 Given A an open subset of a spacetime M, the chronological past of A is defined by: $\downarrow A = I^-(x \in M\colon x \ll a \ \forall a \in A)$.

Clearly $\{x \in M\colon x \ll a \ \forall a \in A\}$ is a past set and by PR. 19, 2.3, $\downarrow A = \mathrm{int}\{x \in M\colon x \ll a \ \forall a \in A\}$. Properties of common futures (pasts) are reviewed in R.K. SACHS, 2. We recall:

PR. 28 Given A and B subsets of a spacetime M,

1. $A \subset B \Rightarrow \uparrow B \subset \uparrow A$
2. $A \subset \downarrow\uparrow A$
3. $A \subset \uparrow B$ iff $B \subset \downarrow A$
4. $\forall x \in M \quad I^{-}(x) \subset \downarrow I^{+}(x)$

PROOF of 1, 2, 3 (R.K. SACHS, 2. Lemma 1.4) of 4 (S.W. HAWKING, 4 and R.K. SACHS, 1, Proposition 1.1).

DFN. 33 A function F from a spacetime M into the power set of M which maps the whole of M into open subsets of M is called inner (outer) continuous iff $\forall x \in M$ and $\forall$ compact set C(K) $C \subset F(x)$ $(K \subset (M - \bar{F}(x)), \exists$ a open nbd. N_x of x $\ni$ $C \subset F(y) \; \forall y \in N_x$ $(K \subset (M - \bar{F}(y)) \; \forall y \in N_x$).

As an example note that:

PR. 29 $\forall x \in M$, $I^{\pm}(x)$ are inner continuous.

PROOF see S.W. HAWKING, 4, and R.K. SACHS, 1, Proposition 1.7.

One can define an ordering on the set of the Lorentz metric tensors: $g < g'$ iff $\forall$ non-zero vector X, $g(X,X) \leq 0$ implies $g'(X,X) < 0$; then the Seifert past is defined by: ${}_{S}J^{-}(x) = \bigcap_{g'>g} J^{-}(x;\, g')$

Another useful concept is the future (past) volume of a point x in a spacetime M; it is always possible to find an additive measure μ on M which assigns positive volume V(U) to each open set U and assigns finite volume to M (R. GEROCH, 3, footnote 24). The future (past) volume of a point $x \in M$ is given by $V^{\pm}(x) = \int_{I^{\pm}(x)} d\mu$. That V^{-} be increasing along TL curves is necessary and sufficient for the absence of closed TL curves; however $V^{\pm}(x)$ will not in general be continuous.

PR. 30 Given a spacetime M, the following are equivalent:

1. M is a reflecting causal space (page 14, 1.3)
2. $\forall x \in M$ and $\forall y \in M$ $x \in \overline{J^{+}(y)}$ iff $y \in \overline{J^{-}(x)}$
3. $\forall x \in M \; \downarrow I^{+}(x) = I^{-}(x)$ and $\uparrow I^{-}(x) = I^{+}(x)$

If in addition the spacetime is future and past distinguishing the equivalence is extended to

4. $\forall x \in M$ $I^{\pm}(x)$ are outer continuous.

5. $\forall x \in M$ $V^{\pm}(x)$ are continuous (and hence, by definition, are C^{0} 'global time' functions)

6. $\forall$ $x \in M$ ${}_{S}J^{\pm}(x) = \overline{J}^{\pm}(x)$

PROOF: S.W. HAWKING, 4, and R.K. SACHS, 1, Proposition 3 and Theorem 2.1.

DFN.34 A future and past distinguishing spacetime M is called causally continuous iff it obeys any one of the equivalent condition of PR. 30, 2.4.

PR. 31 A causally continuous spacetime is stably causal.

PROOF: S.W. HAWKING, 4, and R.K. SACHS, 1, Proposition 2.3 and H.J. SEIFERT.

DFN.35 An open subset A of M is called causally simple iff $\forall$ compact set $C \subset A$, $J^{+}(C) \cap A$ and $J^{-}(C) \cap A$ are closed in A.

Consequently a spacetime is causally simple iff $\forall x \in M, J^{\pm}(x)$ are closed. Furthermore:

PR. 32 A causally simple spacetime (DFN. 35, 2.4) is causally continuous (DFN. 34, 2.4).

PROOF: By PR. 30, 2.4.

The most severe restriction one can put on a spacetime (i.e. a manifold without boundary) is that of global hyperbolicity.

DFN.36 A subset A of a spacetime M is said to be globally hyperbolic iff 1st, it is strongly causal, 2nd $\forall$ x, $y \in A$ $J^{+}(x) \cap J^{-}(y) \subset A$, and 3rd $J^{+}(x) \cap J^{-}(y)$ is compact.

PR. 33 An open, globally hyperbolic set is causally simple.

PROOF S.W. HAWKING, 1 and R.K. SACHS, 1, Proposition 6.6.1.

PR. 34 Given A a closed and achronal subset of a spacetime then int D(A), if nonempty, is globally hyperbolic.

PROOF: S.W. HAWKING, 1, and G.F.R. ELLIS, Proposition 6.6.3.

Global hyperbolicity is related to the existence of Cauchy surfaces.

DFN. 37 A partial Cauchy surface is a spacelike hypersurface which no causal curve intersects more than once.

DFN. 38 A partial Cauchy surface S is said to be a (global) Cauchy surface iff D(S) = M.

PR. 35 If $S \subset M$ is an achronal set which intersects every endless null geodesic in M in a nonempty compact set, then S is a Cauchy surface for M.

PROOF: R. PENROSE, 2, Proposition 5.14.

By PR. 34, DFN. 38, 2.4, and R. GEROCH, 3:

PR. 36 A spacetime is globally hyperbolic iff a Cauchy hypersurface exists for M.

PR. 37 If a Cauchy surface S exists for M, then M is homeomorphic to $\mathbb{R} \times S$. Further if $f : \mathbb{R} \times S \to M$ is the homeo, we can arrange it so that f(t,S) is a Cauchy surface $\forall t \in \mathbb{R}$ and $f(\mathbb{R}, s)$ is a TL curve $\forall s \in S$.

A number of minor technicalities such as limit curves, imprisonment and trapped surfaces have not been tackled in the above presentation. Our main aim was to expose the causality conditions hierarchy. Another major subject, that of extension of spacetimes, (spacetimes with boundary) is somehow related to further causality restrictions (weakly asymptotically simple and empty spacetimes, future asymptotically predictable from a partial Cauchy surface spacetimes et.al.) has not been mentioned at all; the main difficulty is how the causality relation can be extended to the boundary points. Physically the boundary of a spacetime is related to the nature of singular points (singularities) and asymptotic properties (infinity).

2.5 An algebraic view

A causal spacetime (DFN. 10, 2.3), being partially ordered by chronology and causality, qualifies for some applications of well known results of lattice theory.

Let M be a causal spacetime. We denote by J its manifold topology; $J \subseteq B(M)$, B(M) is the power set of M. By $(P) \subseteq J$ and $(F) \subseteq J$, the collections of past and future sets (DFN. 24, 2.3)

By $\hat{L}$ $(\check{L})$ the sets $\{L \in B(M): L = \downarrow A,\ A \in J\}$*$(\{L \in B(M): L = \uparrow A,\ A \in J\})$; $\hat{L} \subseteq (P)$ $(\check{L} \subseteq (F))$ by DFN. 32, 2.4, and PR. 9, 2.3.

A pair $(P, F) \in (P) \times (F)$ is called a hull pair iff $P = \downarrow F$ and $F = \uparrow P$.

PR. 38 In a causal spacetime the following are true:

1. $\hat{L} = \downarrow(F)$ (dually $\check{L} = \uparrow(P)$)
2. $\downarrow.\uparrow$ $(\uparrow.\downarrow)$ is the identity map on $\hat{L}$ $(\check{L})$

PROOF: Since $(F) \subseteq J \Rightarrow \downarrow|(F) \subseteq \downarrow|J$ i.e. $\downarrow(F) \subseteq \downarrow J = \hat{L}$.

Conversely let $L = \downarrow U$, $U \in J$; by the dual of PR. 28, 2.4, $U \subseteq \uparrow.\downarrow U = \uparrow L$ and $L = \downarrow U \supseteq \downarrow.\uparrow L$; but $L \subseteq \downarrow\uparrow L$; therefore $L = \downarrow.\uparrow L$ (this proves (2)). $\uparrow L \in (F)$ hence $\downarrow\uparrow L \in \downarrow(F)$

Thus $\hat{L} = \downarrow(F)$.

PR. 39 The mappings $\downarrow, \uparrow : J \rightarrow J$ are antitone (J is partially ordered by the set inclusion relation)

PROOF: PR. 28 and its dual, 2.4.

THEOREM: $\hat{L}$ is a complete lattice partially ordered by set inclusion, under the meet operation defined by: $\sqcap(N) = \text{interior} \bigcap_{N \in (N)} N$, $(N) \subseteq \hat{L}$ and the join by:

$$\sqcup(N) = \downarrow.\uparrow \bigcup_{N \in (N)} N.$$

PROOF: J is a complete lattice partially ordered by set inclusion, with 'interior $\bigcap$' as meet and '$\bigcup$' as join. By PR. 39. 2.5 and PR. 28, 2.4, $(\downarrow, \uparrow)$ is a Galois connection from J to itself; hence $\downarrow.\uparrow$ $(\uparrow.\downarrow)$ is a closure operation on J. By PR. 38, 2.4, $\hat{L}$ is the set of closed elements.

* See Appendix II.

The partial order on $\hat{L}$ $(\check{L})$ induced from its lattice structure i.e. $L_1 < L_2$, $L_{1,2} \in \hat{L}$ iff $L_1 \sqcap L_2 = L_1$ ($\Rightarrow$ int $(L_1 \cap L_2) = L_1 \Rightarrow$ $L_1 \subseteq L_2$) or $L_1 \sqcup L_2 = L_2$ ($\Rightarrow \downarrow . \uparrow$ $(L_1 \cup L_2) = L_2 \Rightarrow L_1 \subseteq L_2$ again), being a set inclusion relation and hence reflexive, qualifies for a causality rather than a chronology relation. Consequently the question remains how one could define a consistent chronology on $\hat{L}$ $(\check{L})$ and how these definitions should extend to $\hat{L} \cup \check{L}$ and further to more general sets (e.g. future and past sets: $(P) \cup (F)$). The motive behind such an attempt is closely related to the various procedures involved in the completion of a spacetime (i.e. attachment of a boundary representing 'singular' points and points at 'infinity') and cannot be fully appreciated unless one examines thoroughly this problem.

A general scheme has been proposed by R.K. Sachs and R. Budic (R.K. SACHS, 2, and R. BUDIC) given on $(P) \cup (F)$ by the following table.

		causality	chronology	Woodhouse's chronology
$(P) \times (F)$		$P_1 \subseteq P_2$	$\uparrow P_1 \cap P_2 \neq \emptyset$	$\uparrow\downarrow P_1 \cap \downarrow P_2 \neq \emptyset$
$(F) \times (F)$		$F_1 \supseteq F_2$	$F_1 \cap \downarrow F_2 \neq \emptyset$	$\uparrow F_1 \cap \downarrow\uparrow F_2 \neq \emptyset$
$(P) \times (F)$	$\exists \; \hat{L}, \check{L}$	$F \subseteq \check{L}$, $\hat{L} \supseteq P$	$\uparrow P \cap \downarrow F \neq \emptyset$	$\uparrow\downarrow P \cap \downarrow\uparrow F \neq \emptyset$
$(F) \times (P)$	$\ni$	$F \supseteq \check{L}$, $\hat{L} \subseteq P$	$F \cap P \neq \emptyset$	$\uparrow F \cap \downarrow P \neq \emptyset$

In a specific treatment of chronological sequences (N.M.J. WOODHOUSE, 2) a definition of the chronology relation is provided which is a special case of the one cited above as can be seen from the last column of the above table and the fact that $\downarrow P \subseteq P$ and $\uparrow F \supseteq F$. (The results of N.M.J. Woodhouse have been translated into our notation).

PART 2 - SINGULARITY THEORY AND GENERAL RELATIVISTIC SPACE-TIMES

CHAPTER 3

3.1 The Characteristics of a characteristic equation

The characteristic of the wave equation, that is the first-order partial differential equation

$$\frac{1}{c^2} \cdot \left[\frac{\partial f}{\partial t}\right]^2 - \left[\left[\frac{\partial f}{\partial x}\right]^2 + \left[\frac{\partial f}{\partial y}\right]^2 + \left[\frac{\partial f}{\partial z}\right]^2\right] = 0 \qquad (3.1)$$

where c is a constant, $(t,x,y,z) \in \mathbb{E}^4$ and f a continuously differentiable real valued function on $\mathbb{E}^4$, is linked with Realitivity Theory in more than one significant ways.

Let us introduce a new notation : (x_o, x_1, x_2, x_3) for the coördinates in $\mathbb{E}^4$ with $x_o = ct$, $x_1 = x$, $x_2 = y$, $x_3 = 2$, and a quadratic form denoted by ds^2 and given by

$$ds^2 = \sum_{\alpha=0}^{3} e_\alpha (dx_\alpha)^2 \qquad (e_o = 1\ e_1 = e_2 = e_3 = -1) \qquad (3.2)$$

Relation (3.1) then looks like

$$(\nabla f)^2 = \sum_{\alpha=0}^{3} e_\alpha \left(\frac{\partial f}{\partial x_\alpha}\right)^2 = 0 \qquad (3.3)$$

Upon a coördinate transformation

$$x'_\alpha = X_\alpha(x_\beta) \qquad (\alpha,\beta = 0,\ 1,\ 2,\ 3)$$

with a non-vanishing Jacobian, i.e. there exist functions $X'_\alpha(x'_\beta)$ such that

$$x_\alpha = X'_\alpha(x'_\beta) \qquad (\alpha,\beta = 0,\ 1,\ 2,\ 3)$$

(3.2) and (3.3), transforms into

$$ds^2 = \sum_{\alpha,\beta=0}^{3} g_{\alpha\beta}\ dx'_\alpha\ dx'_\beta \qquad (3.3)$$

and

$$(\nabla f')^2 = \sum_{\alpha,\beta=0}^{3} g^{\alpha\beta} \frac{\partial f'}{\partial x'_\alpha} \frac{\partial f'}{\partial x'_\beta} = 0 \qquad (3.5)$$

with $f'(X_\alpha) = f(x_\alpha)$, $g_{\alpha\beta} = \sum_{\gamma=0}^{3} e_\gamma \frac{\partial X'_\gamma}{\partial x'_\alpha} \frac{\partial X'_\gamma}{\partial x'_\beta}$, $g^{\alpha\beta}\ e_\gamma \frac{\partial X_\alpha}{\partial x_\gamma} \frac{\partial X_\beta}{\partial x_\gamma}$

satisfying

$$\sum_{\beta=0}^{3} g_{\alpha\beta}\, g^{\gamma\beta} = \delta^{\gamma}_{\alpha} = \begin{cases} 1 & \alpha = \gamma \\ 0 & \alpha \neq \gamma \end{cases} \tag{3.6}$$

The Principle of Special Relativity asserts that phenomena occuring in a closed system are independent of any non-accelerated motion of the system as a whole. In the domain of pre-relativistic mechanics the principle has long been known; it is the Galilean relativity principle. Einstein's achievement was to extend it to all phenomena (though in the first place to electromagnetism) and to derive from it certain consequences regarding the interrelation of space and time. The key concept in trying to make the content of this postulate precise is that of an inertial frame; in pre-relativistic physics the notion of an inertial frame was related to the laws of mechanics and an inertial frame was defined as one with respect to which a body moves uniformly and in a straight line, provided no forces act on it (Newton's first law of motion); to relate this notion of an inertial frame to electromagnetism one has to consider Maxwell's equations for which has always been assumed, even before Relativity, that at least one reference frame exists that is inertial with respect to mechanics and in which at the same time they (Maxwell's equations) are valid. One can show that an electromagnetic wave front, a characteristic surface for the Maxwell Equations, is given by an equaiton

$$f(x_\alpha) = 0 \quad (\alpha = 0,1,2,3) \tag{3.7}$$

where f is a solution of equation (3.3) (V. FOCK, §3). Relation (3.7) is clearly the equation of a certain hypersurface in the four-dimensional space-time continuoum; define a displacement (dx_1, dx_2, dx_3) given by $dx_i = {f_{x_i}}/{|{}_3\nabla f|}$. dm with $f_{x_i} = \frac{\partial f}{\partial x_i}$ $(i = 1, 2, 3)$, ${}_3\nabla f = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3})$

and η the conormal vector on the hypersurface defined by (3.7). Relation (3.7) implies $|{}_3\nabla f|d\eta + f_{x_o}\, dx_o = 0$ and (3.7) may represent a surface (in three dimensional space) in motion, if the desplacement velocity defined by $v^2 = c^2 \left(\frac{d\eta}{dx_o}\right)^2$ is equal (or less) the velocity of light (c^2) i.e.

$$(\nabla f)^2 \leq 0 \qquad (3.8)$$

A frame that is inertial both in the mechanical and in the electromagnetic senses could therefore be characterized by the following two properties :

1st A body moves uniformaly and in a straight line, provided no forces act on it.

2nd. The equation of propagation of an electromagnetic wave front has the form (3.3)

It is a consequence of the second condition that the velocity of light is independent of the velocity of its source; but the condition is capable of a more general interpretation, such as :

> "There exists a maximum speed for the propagation of any kind of action; this is numerically equal to the speed of light in free space".

This is consistent with the Principle of Relativity, for if there was no single limiting velocity but instead different agents (light and gravitation, propagated in vacuo with different speeds) then the Principle of Relativity would necessarily be violated as regards at least one of the agents.

The existence of a general upper limit for all kinds of action, endows the speed of light with a universal significance and equation (3.3) acquires a general character; it becomes more general than Maxwell's equations from which it is derived. As a consequenc of the principle of the existence of a universal limiting velocity one can assert that the differential equations describing any field that is capable of transmitting signals must be of such a kind that the equation of their characteristics is the same as the equation for the characteristic of light waves (i.e. equation (3.3)).

Equation (3.5) is a partial differential equation of the following general form:

$$H(x_i, f_{x_i}) = 0 \qquad (i = 1,\ldots,n) \qquad (3.9)$$

where $H(x_i, y_i)$ is a function on $\mathbb{E}^{2n}$, at least twice continuously differentiable, and such that in the neighborhood of the point at which $H^2_{y_i} \neq 0$ it is also true that $H(x_i, y_i) = 0$; $f(x_i)$ is an at least twice continuously differentiable function on $\mathbb{E}^n$ and a solution of equation (3.9). By partial differentiation of the relation (3.9) with respect to x_i we obtain the identities

$$H_{x_i} + H_{y_j} f_{x_j x_i} = 0 \qquad (i = 1,\ldots,n) \qquad (3.10)$$

where $y_i = f_{x_i}$ (and $f_{x_i} = \partial f / \partial x_i$).

We consider in the space of the x_i an arbitrary differentiable curve $x_i = x_i(W)$.

By substitution of these values of x_i into $y_i = f_{x_i}(x_i)$ and differentiation with respect to W (denoted by a dot) we obtain the relations:

$$\dot{y}_i = f_{x_i x_j} \dot{x}_j \qquad i = 1,\ldots,n) \qquad (3.11)$$

Adding equations (3.11) to equations (3.10) we obtain the identities :

$$\dot{y}_i + H_{x_i} = f_{x_i x_j}(\dot{x}_j - H_{y_j}) \qquad (i = 1,\ldots,n) \qquad (3.12)$$

Let us now specialize our curve $x_i(W)$ by requiring

$$\dot{x}_i = H_{y_i}(x_j, f_{x_j}) \qquad (3.13)$$

and from (3.12)

$$\dot{y}_i = -H_{x_i}(x_j, f_{x_j}) \qquad (3.14)$$

If we consider that in (3.13) and (3.14) we must take

$f_{x_j} = f_{x_j}(x_i(W)) = y_j(W)$ $(y_i = f_{x_i}(x_j(W)) = y_i(W))$ then it can be seen that the 2n functions $x_i(W)$ and $y_i(W)$ are solution of a system of ordinary differential equations, first used in mechanics, the so called canonical differential equations :

$$\dot{x}_i = H_{y_i}(x_j, y_j) \qquad (i = 1,\ldots,n) \qquad (3.15)$$

$$\dot{y}_i = -H_{x_i}(x_j, y_j) \qquad (i = 1,\ldots n) \qquad (3.16)$$

In order to determine the family of curves which is defined by equations (3.15), we must have an at least twice continuously differentiable solution $f(x_i)$ of the partial differential equation (3.9) where existence is in no way ensured. On the other hand by the existence theorems of the theory of ordinary differentiable equations the system (3.15), (3.16) can determine each curve of the family if at point x_i^o of this curve we merely know the values y_i^o $(=f_{x_i}(x_j^o))$. To state this fact in geometric language consider in an $(n+1)$-dimensional space (x_i, z) a surface element which passes through the point $(x_i^o, 0)$ and whose normal has the components $(y_i^o, 1)$; if there exists at least one twice continuously differentiable solution $f(x_i)$ of the partial differential equation (3.9) which considered as a hypersurface $z = f(x_i)$ contains this surface element, then firstly the entire curve $x_i = x_i(N)$ which we calculated as a solution of (3.15) with the given initial conditions must lie on the surface $z = f(x_i)$ and secondly , the direction components of the normals to the surface are determined along this curve by the functions $y_i(W)$ which likewise are uniquely determined by the prescribed initial values and (3.16). One can show that one can pass infinitely many distinct surfaces $z = f(x_i)$ through a given surface element such that f is a solution of the partial differential equation (3.9); all these surfaces must therefore touch along the curve just calculated which is called a characteristic of the differential equation (3.9). From the single characteristics which one obtains as solutions of the ordinary differential equations (3.15) and (3.16) one can construct families of curves which constitute the totality of the characteristics lying on a solution of the partial differential equation; the value of the method, developed by A.L. Cauchy as early as 1819, is that it reduces partial differential equation of the form (3.9) to the integration of ordinary differential equations; (C. CARATHEODORY §40, §41, §42) i.e. one can construct from the characteristics all, at least twice continuously differentiable solutions of the equation (3.9). Specifically, for partial differential equation of the form (3.9), it is possible to establish necessary and sufficient conditions in order that the family of functions

$$\begin{aligned} x_i &= \xi_i(W,u_\alpha) \qquad & i &= 1,\dots,n \\ y_i &= \eta_i(W,u_\alpha) \qquad & \alpha &= 1,\dots,n-1 \end{aligned} \tag{3.17}$$

represent all characteristics of a solution of this differential equation which pass through a certain region of the space of the x_i; more precisely.

PR. 1 If the functions (3.17) denote continuously differentiable solutions of the canonical differential equations (3.15), (3.16), then these represent the characteristics of a solution $f(x_i)$ of the partial differential equation $H(x_i, f_{x_i})$ = constant for suitable value of these last constants if and only if the functional derivative

$$\frac{\partial(\xi_1,\dots,\xi_n)}{\partial(W,u_1,\dots,u_{n-1})} \neq 0$$ and all relations $[u_\alpha, W] = 0$, $[u_\alpha, u_\beta] = 0$ are fulfilled (where the symbol [ ] denotes the *Lagrange brackets* defined by $[u_\alpha,u_\beta] = \frac{\partial\xi_i}{\partial u_\alpha}\frac{\partial\eta_i}{\partial u_\beta} - \frac{\partial\xi_i}{\partial u_\beta}\frac{\partial\eta_i}{\partial u_\alpha}$ and $[u_\alpha, W] = \frac{\partial\xi_i}{\partial u_\alpha}\frac{\partial\eta_i}{\partial W} \quad \frac{\partial\xi_i}{\partial W}\frac{\partial\eta_i}{\partial u_\alpha}$)

PROOF (C. CARATHEODORY, §46, Theorem 2)

In the four dimensional case (n=4) and for the equation (3.5) $\left(H(x_\alpha,y_\alpha) = \frac{1}{2}\sum_{\alpha,\beta=0}^{3} g^{\alpha\beta}(x_\alpha)y_\alpha y_\beta\right)$ the canonical differential equations (3.13) and (3.14) look like :

$$\frac{dx_\alpha}{dW} = \sum_{\beta=0}^{3} g^{\alpha\beta} y_\beta \qquad (\alpha,\beta = 0,1,2,3) \qquad (3.18)$$

$$\frac{dy_\alpha}{dW} = -\tfrac{1}{2} \sum_{\beta,\gamma=0}^{3} \frac{\partial g^{\beta\gamma}}{\partial x_\alpha} y_\beta y_\gamma \qquad (\alpha,\beta,\gamma = 0,1,2,3) \qquad (3.19)$$

In the special case of equation (3.3), equations(3.18) transform into:

$$\frac{dx_o}{dW} = y_o \qquad \frac{dx_i}{dW} = -y_i \qquad (i = 1,2,3) \qquad (3.20)$$

Upon integration with respect to W and with initial conditions (x_α^o, y_α^o) ($\alpha = 0,1,2,3$) we obtain (substituting W with the time t.c. after the integration of the first of equations (3.20)) the equation of a straight line

$$x_i = x_i^o - c\,\frac{y_i^o}{y_o^o}\,t \qquad i = 1,2,3$$

As a consequence the second condition in the definition of an inertial system is equivalent to the fact that light travels in straight lines. At this point one can pose a purely mathematical problem, i.e. find a transformation between two inertial frames without any further physical assumptions, other than the definition of an inertial frame given in page 59. More precisely, find the coördinate transformation

$$x'_\alpha = X_\alpha(x_\beta) \qquad (\alpha,\beta = 0,1,2,3) \qquad (3.21)$$

for which the following two conditions are fulfilled:

1st To a uniform rectilinear motion in the coördinates (x_α) there corresponds a motion of the same nature in the coördinates (x'_α).

2nd To a uniform rectilinear motion with light-velocity in the coördinates (x_α) there corresponds a motion of the same nature in the coördinates (x'_α).

or equivalently : To the wave front equation (3.3) in the coördinates (x_α) there correspond just such an equation in the coördinates (x'_α) .

The results, obtained by V. Fock (V. FOCK, Appendix A) and also in H. Weyl's Mathematische Analyse des Raumproblems (1923) can be summarized as follows :

1st The most general form of the transformation (3.21) which satisfies the 1st condition is a transformation involving linear fractions, all with the same denominator

2nd The most general form of the transformation (3.21) which satisfies the 2nd condition is a product of a linear transformation followed by a Möbius transformation, known as special conformal map (J. WESS, §1) and a change in scale.

A special conformal map is generated in general by a triple product of an inversion by reciprocal radii(i.e. $t_\alpha \to x_\alpha \cdot (\sum e_\alpha x_\alpha^2)^{-1}$) , a translation by (c_α) and the same inversion, where defined, and it has the explicit form:

$$x^*_\alpha(x_\beta) = \frac{x_\alpha - c_\alpha \sum_{\beta=0}^{3} e_\beta x^2_\beta}{1 - 2\sum_{\beta=0}^{3} e_\beta c_\beta x_\beta + \sum_{\beta=0}^{3} e_\beta c^2_\beta \sum_{\gamma=0}^{3} e_\gamma x^2_\gamma} \quad (\alpha,\beta = 0\ 0,\ 1,\ 2,\ 3)$$

The special conformal transforamtions form a non affine abelian subgroup of the full conformal group which lies in the identity consponent; a Möbius transformation reduces to identity either by also adopting the 1st condition or by additionally demanding that it should always transform finite values of coördinates into finite ones.

In General Relativity Theory the quantities $g_{\alpha\beta}(g^{\alpha\beta})$ in the expressions (3.4) and (3.5) are not necessarily given by $\sum_{\gamma=0}^{3} e_\gamma \frac{\partial X'_\gamma}{\partial x'_\alpha} \frac{\partial X'_\gamma}{\partial x'_\beta}$ and $\sum_{\gamma=0}^{3} e_\gamma \frac{\partial X_\alpha}{\partial x_\gamma} \frac{\partial X_\beta}{\partial x_\gamma}$ respectively, but they are taken simply as functions of the coördinates x'_α ; however they are not arbitrary functions of x'_α as they are restricted to satisfy Einstein's equations, and also the inequalities $g_{oo} > 0$, $g_{ik}\, r_i r_k < 0$ $(i,k = 1, 2, 3,\ \ r_i \in \mathbb{R}$, in order to ensure that x'_o is of the nature of time whereas x'_i (i = 1, 2, 3) are of the nature of spatial coördinates.

It is readily seen that equations (3.18) and (3.19) are equivalent to the equations for the space-time geodesic curves (space-like, time-like and null with the appropriate choice of the constant in PR. 1, 3.1, as a negative, positive and zero number correspondingly),

By (3.6) $\frac{\partial}{\partial x_\alpha} (\delta^\beta_\alpha) = \frac{\partial}{\partial x_\alpha}\left(\sum_{\gamma=0}^{3} g_{\gamma\alpha}\, g^{\beta\gamma}\right) = \sum_{\beta=0}^{3} \frac{\partial g_{\gamma\alpha}}{\gamma x_\alpha} g^{\beta\gamma} + \sum_{\beta=0}^{3} g_{\gamma\alpha} \frac{\partial g^{\beta\gamma}}{\gamma x_\alpha} = 0$

By (3.18) $$y_\gamma = \sum_{\alpha=0}^{3} g_{\gamma\alpha} \frac{dx_\alpha}{dw} \qquad (3.22)$$

Relations (3.19) transform into

$$\frac{dy_\alpha}{dw} = -\tfrac{1}{2} \sum_{\alpha,\beta,\gamma=0}^{3} \frac{\partial g^{\beta\gamma}}{\partial x_\alpha} y_\beta\, g_{\gamma\alpha} \frac{dx_\alpha}{dw} = \tfrac{1}{2} \sum_{\alpha,\beta,\gamma=0}^{3} \frac{\partial g_{\gamma\alpha}}{\partial x_\alpha} g^{\beta\gamma} y_\beta \frac{dx_\alpha}{dw} \qquad \ldots\ldots \quad (3.23)$$

By using relations (3.22) and (3.18), relation (3.23) becomes

$$\frac{d}{dw}\left(\sum_{\beta=0}^{3} g_{\alpha\beta} \frac{dx_\beta}{dw}\right) = \tfrac{1}{2} \sum_{\alpha,\gamma=0}^{3} \frac{\partial g_{\gamma\alpha}}{\partial x_\alpha} \frac{dx_\gamma}{dw} \frac{dx_\alpha}{dw} \quad (\alpha = 0,\ 1,\ 2,\ 3) \qquad \ldots\ldots \quad (3.24)$$

hence

$$\sum_{\beta=0}^{3} g_{\alpha\beta} \frac{d^2 x_\beta}{dw^2} + \sum_{\beta,\gamma=0}^{3} \frac{\partial g_{\alpha\beta}}{\partial x_\gamma} \frac{dx_\gamma}{dw} \frac{dx_\beta}{dw} - \tfrac{1}{2} \sum_{\beta,\gamma=0}^{3} \frac{\partial g_{\beta\gamma}}{\partial x_\alpha} \frac{dx_\gamma}{dw} \frac{dx_\beta}{dw} = 0$$

and symmetrizing the coefficient of $\frac{dx_\beta}{dw} \frac{dx_\gamma}{dw}$ with respect to β and γ :

$$\sum_{\beta=0}^{3} g_{\alpha\beta} \frac{d^2 x_\beta}{dw^2} + \sum_{\beta,\gamma=0}^{3} \tfrac{1}{2} \left(\frac{\partial g_{\alpha\beta}}{\partial x_\gamma} + \frac{\partial g_{\alpha\gamma}}{\partial x_\beta} - \tfrac{1}{2} \frac{\partial g_{\beta\gamma}}{\partial x_\alpha}\right) \frac{dx_\beta}{dw} \frac{dx_\gamma}{dw} = 0 \qquad \ldots\ldots \quad (3.25)$$

Introducing the Christoffel symbols :

$$\tfrac{1}{2} \left(\frac{\partial g_{\alpha\beta}}{\partial x_\gamma} + \frac{\partial g_{\alpha\gamma}}{\partial x_\gamma} - \tfrac{1}{2} \frac{\partial g_{\beta\gamma}}{\partial x_\alpha}\right) = \{\beta\gamma,\alpha\}$$

and

$$\Gamma^\mu_{\beta\gamma} = \sum_{\alpha=0}^{3} g^{\mu\alpha} \{\beta\gamma,\alpha\}$$

(3.25) is written as

$$\sum_{\beta=0}^{3} g_{\alpha\beta} \frac{d^2 x_\beta}{dw^2} + \sum_{\beta,\gamma=0}^{3} \{\beta\gamma,\alpha\} \frac{dx_\beta}{dw} \frac{dx_\gamma}{dw} = 0$$

or

$$\frac{d^2x_\mu}{dw^2} + \Gamma^\mu_{\beta\gamma} \frac{dx_\beta}{dw} \frac{dx_\gamma}{dw} = 0 \tag{3.26}$$

A solution of equations (3.18) and (3.19) according to PR.1, 3.1, and relation (3.17) is given parametrically by :

$$x_\alpha = \xi_\alpha(w, u_i) \qquad \alpha = 0, 1, 2, 3$$

$$y_\alpha = \eta_\alpha(w, u_i) \qquad i = 1, 2, 3$$

$(\xi_\alpha, \eta_\alpha)$ $(\alpha = 0, 1, 2, 3)$ must therefore satisfy equations (3.15) and (3.16) i.e.

$$\frac{\partial \xi_\alpha}{\partial w} = H_{y_\alpha}(\xi_\beta, \eta_\beta) \tag{3.27}$$

$$(\alpha, = 0, 1, 2, 3)$$

$$\frac{\partial \eta_\alpha}{\partial w} = -H_{x_\alpha}(\xi_\beta, \eta_\beta) \tag{3.28}$$

Let $H(\xi_\alpha, \eta_\alpha) = \phi(w, u_i)$; we now investigate the derivatives of ϕ:

$\frac{\partial \phi}{\partial w} = \frac{\partial H}{\partial x_\alpha} \frac{\partial \xi_\alpha}{\partial w} + \frac{\partial H}{\partial y_\alpha} \frac{\partial \eta_\alpha}{\partial w} = 0$ by virtue of equations (3.27) and (3.28);

hence the function $H(x_\alpha, y_\alpha)$ is an integral of the differential equations (3.18) and (3.19) for the characteristics;

but $H(x_\alpha(w)),\ y_\alpha(w)) = \frac{1}{2} \sum_{\alpha,\beta=0}^{3} g^{\alpha\beta}(x_\alpha(w)) y_\alpha(w) y_\beta(w)$

$$= \frac{1}{2} \sum_{\alpha,\beta=0}^{3} g_{\alpha\beta}(x_\alpha(w)) \dot{x}_\alpha(w) \dot{x}_\beta(w) \qquad \text{by } (3.22)$$

determines the sign of the square of the infinitesimal space-time interval ds^2(see relation (3.4)) along a solution curve of equations (3.26)(of equations (3.18) and (3.19)equivalently), which, since H is constant along each characteristic curve , detemines its null, spacelike or timelike character.

$$\frac{\partial \phi}{\partial u_i} = H_{x_\alpha} \frac{\partial \xi_\alpha}{\partial u_i} + H_{y_\alpha} \frac{\partial \eta_\alpha}{\partial u_i} = - \frac{\partial \eta_\alpha}{\partial w} \frac{\partial \xi_\alpha}{\partial u_i} + \frac{\partial \xi_\alpha}{\partial w} \frac{\partial \eta_\alpha}{\partial u_i} \quad \text{by (3.27) and (3.28)}$$

$$= -[u_1, w] = 0 \qquad (i = 1, 2, 3)$$

and therefore the function $\phi(u_i) = H(\xi_\alpha, \eta_\alpha)$ must be constant everywhere. From the latter propositions, by choosing according to PR.1,3.1, this constant equal to 0, -1 and 1, we obtain for a characteristic a null, spacelike and timelike geodesic respectively.

The above presentation of geodesic curves in General Relativity i.e. via the theory of characteristics, has the advantage over the Lagrangian formulation that it includes null geodesics as well, which cannot be accounted for by a vanishing Lagrangian. Most significantly so, since the null geodesics can be regarded as invariantly determined gravitational rays i.e. trajectories of propagation of the gravitational wavefront.

Indeed, equation (3.5) is also the characteristic equation for Einstein's equations in empty space (B. FINZI). The (characteristic) hypersurface $f(x^\alpha) = 0$ of the field equations can be shown to be the hypersurface of discontinuity of the field functions (and/or their derivatives), known as the wave-front surface and the discontinuity as the Hadamard discontinuity in the solution of the equations on the hypersurface. See V.D. Zakharov (V.D. ZAKHAROV chapter 2 §2) for a bibliographical review. Furthermore the bicharacteristics of Einstein's equations are isotropic (null) geodesics.

Lastly we would like to mention the relation between a general integral of a system of canonical differential equations and the so called canonical transformations.

Let the relations

$$x'_i = X_i(x_j, y_j) \qquad (i,j = 1,\ldots,n) \qquad (3.29)$$

$$y'_i = Y_i(x_j, y_j) \qquad (3.30)$$

hold between n pairs of variables (x_j, y_j) and a similar set of variables (x'_i, y'_i) such that the functions X_i, Y_i are at least continuously differentiable and the relations

$$[x_j, x_k] = 0 \ , \quad [x_j, y_k] = \delta_{jk} \ , \quad [y_j, y_k] = 0$$

$$\ldots\ldots \qquad (i,j,k = 1,\ldots,n) \quad (4.31)$$

hold identically ($[x_j, y_k] = \frac{\partial X_i}{\partial x_j}\frac{\partial X_i}{\partial y_k} - \frac{\partial X_i}{\partial y_k}\frac{\partial X_i}{\partial x_j}$ etc)

A transformation of variables (x_i, y_i), $i = 1,\ldots n$, into the variables (x'_i, y'_i), $i = 1,\ldots,n$, defined by equations (3.29) and (3.30) will be then called *canonical*.

Equivalently , a necessary and sufficient condition for the transformation defined by (3.29) and (3.30) to be canonical is expressed by the simultaneous existence of the relations :

$$(X_i, X_j) = 0, \quad (Y_i, X_j) = \delta ij, \quad (Y_i, Y_j) = 0 \qquad (i,j = 1,\ldots,n) \qquad (3.32)$$

where the brackets are the *Poisson brackets*

$\left(\text{e.g. } (X_i, Y_j) = \frac{\partial X_i}{\partial y_k}\frac{\partial Y_j}{\partial x_k} - \frac{\partial X_i}{\partial x_k}\frac{\partial Y_j}{\partial y_k}\right)$. The Poisson bracket is an invariant formula under a canonical transformation (i.e. $(x,y) = (x',y')$, where X and Y are any twice continuously differentiable functions on 2n variables) and this additionally proves that the set of canonical transformations in 2n variables forms, a group for any n . A certain subgroup of the group of canonical transformations is that of the *elementary canonical transformations*; they are defined as products of both the following transformations

$$\begin{aligned} x'_i &= x_{a_i} \\ y'_i &= y_{a_i} \end{aligned} \qquad (i = 1,\ldots,n) \qquad (3.33)$$

(a_i) $i = 1,\ldots,n$ is any permutation of the first n integers

and

$$\begin{aligned} x'_\lambda &= y_\lambda & \lambda &= 1,\ldots,p) \quad 1 \le p \le n \\ x'_\mu &= x_\mu & \mu &= p+1,\ldots,n \\ y'_\lambda &= -x_\lambda \\ y'_\mu &= y_\mu \end{aligned} \qquad (3.34)$$

One can further prove that one can regard every canonical transformation as the product of an elementary canonical transformation and a transformation for which the functional determinant $\left\{\frac{\partial(x_1,\ldots,x_n)}{\partial(y_1,\ldots,y_n)} \neq 0\right\}$ of X 's with respect to y's is different from zero. One then can solve for $y_i = Z_i(x_j,x'_k)$ and prove that there is an at least twice continuously differentiable function $\Phi(x_j,x'_k)$, called *generating function*, satisfying the relations

$$Z_i(x_j,x'_k) = -\frac{\partial\Phi(x_j,x_k)}{\partial x_i} \quad \text{and } Y_i(x_j,Z_j) = \frac{\partial\Phi(x_j,x'_k)}{\partial x'_i} \tag{3.35}$$

from which Φ can also be calculated by integration.

Conversely, given an at least twice continuously differentiable function satisfying the condition det $\left\{\frac{\partial^2\Phi(x_j,x'_k)}{\partial x_j \partial x'_k}\right\} \neq 0$ one can calculate a canonical transformation (for which Φ is a generating function) from the relations (3.35)

$$y_i = -\frac{\partial\Phi(x_j,x'_k)}{\partial x_i} \quad \text{and } y'_i = \frac{\partial\Phi(x_j,x'_k)}{\partial x'_i} \quad .$$

Similar results are also true for the general case, although the form of Φ and the construction proofs are more complicated.

In one parameter family of canonical transformations

$$x'_i = X_i(w,x_j,y_j),\ y'_i = Y_i(w,x_j,y_j)(i,\ j=1,\ldots,n) \tag{3.36}$$

the functions X_i, Y_i can also be considered as the general solution of a system of differential equations

$$\dot{x}'_i = \phi_i(w,x'_j,y'_j)\ ,\ \dot{y}'_i = \psi(w,x'_j,y'_j)(i,\ j=1,\ldots,n) \tag{3.37}$$

where the dot means differentiation with respect to w. Therefore

$$\frac{\partial X_i}{\partial w} = \phi_i(w,X_j,Y_j) \text{ and } \quad \frac{\partial Y_i}{\partial w} = \psi_i(w,X_j,Y_j)$$

and the relations (3.32) result into

$$\left[\frac{\partial X_i}{\partial w}, X_j\right] = \frac{\partial\phi_i}{\partial y'_j} \qquad \left[\frac{\partial X_i}{\partial w}, Y_j\right] = -\frac{\partial\phi_i}{\partial x'_j}$$

$$\left[\frac{\partial Y_i}{\partial w}, X_j\right] \quad \frac{\partial\psi_i}{\partial y'_j} \qquad \left[\frac{\partial Y_i}{\partial w}, Y_j\right] = -\frac{\partial\psi_i}{\partial x'_j}$$

Differentiating relations (3.32) with respect to w and inserting the above we have

$$\frac{\partial\phi_j}{\partial y'_i} - \frac{\partial\phi_i}{\partial y'_j} = 0 \qquad \frac{\partial\psi_i}{\partial y'_j} + \frac{\partial\phi_j}{\partial x'_i} = 0 \qquad \frac{\partial\psi_i}{\partial x'_j} - \frac{\partial\psi_j}{\partial x'_i} = 0 \tag{3.38}$$

Equations (3.38) imply the existence of a function $H(x'_i, y'_i)$ such that the equations

$$\phi_i(w, x'_j, y'_j) = \frac{\partial H}{\partial y'_i} \quad , \quad \psi_i(w, x'_j, y'_j) = -\frac{\partial H}{\partial x'_i} \tag{3.39}$$

are valid. Therefore the equations (3.37) have the canonical forms

$$\dot{x}'_i = H_{y'_i}(w, x'_j, y'_j) \qquad \dot{y}'_i = -H_{x'_i}(w, x'_j, y'_j) \tag{3.40}$$

Conversely given an arbitrary system of canonical differential equations (3.40) as well as a general (2n-parameter family) of solutions (3.36) of these equations, which depend on the constants of integration (x_i, y_i), the Lagrange brackets $[x_i, x_j]$, $[x_i, y_j]$ and $[y_i, y_i]$ are independent of w as can be seen for example from the expression

$$\frac{\partial}{\partial w}\left(\frac{\partial X_k}{\partial x_i}\;\frac{\partial Y_k}{\partial x_j}\right) = -\sum_{k,\ell} H_{x'_k x'_\ell}\;\frac{\partial X_k}{\partial x_i}\;\frac{\partial X_\ell}{\partial x_j} + \sum_{k,\ell} H_{y'_k y'_\ell}\frac{\partial Y_k}{\partial x_i}\;\frac{\partial Y_\ell}{\partial x_j}$$

which is symmetric in i and j.

By choosing the initial values of x'_i, y'_i such that the functions X_i, Y_i satisfy the conditions (3.31), then these conditions are identically satisfied for all values of w and the equations (3.36) represent a family of canonical transformations. Hence the proposition :

PR.2 An arbitrary one-parameter family of canonical transformations and the general integral of a system of canonical differential equations whose constants of integration have been so chosen that the Lagrange brackets satisfy the relations (3.31), are identical concepts.

3.2 An invariant description

The aim of this section is to describe the structures related to non-linear first order partial differential equations, such as the equation (3.5), 3.1, in an invariant i.e. coördinate free way. Although we will be talking about spacetimes i.e. four-dimenional, connected, Hausdorff and C^∞-differentiable (smooth) manifolds with a globally defined, smooth, tensor field of type (0,2), which is nondegenerate and Loventzian (DFN.1,2.2), everything we say will also hold for any n-dimensional, smooth manifold.

We consider the ring of smooth further $C^\infty(M)$ on a smooth manifold, its maximal ideal $m_p, p \in M$, consisting of the functions that vanish at p and its powers m_p^k, $k > 0$, consisting of the functions that vanish at p together with their first k-derivatives, and the factor rings $J_p^k(M) = C^\infty(M)/m_p^{k+1}$ and $\tilde{J}_p^k(M) = {}^{m}p/_{m_p^{k+1}}$. We give both $\tilde{J}^k(M) = \bigcup_{p\in M} \tilde{J}_p^k(M)$ and $\tilde{J}^k(M) = \bigcup_{p\in M} \tilde{J}_p^k(M)$, $0 \leq k < \infty$, the structure of a smooth manifold : Let (x_i), $i = 1, \ldots, n$ be a system of local coördinates in a neighbourhood $U \subset M$; in $J_p^k(M)$, $p \in U$, regarded as a vector space, we introduce coördinates (u_α), $\alpha = (\alpha_1, \ldots, \alpha_n)$, $|\alpha| = \alpha_1 + \ldots + \alpha_n \leq k$, relative to the (vector space) basis :

$$j_k(\text{id.})_p, j_k(x_i - \overset{\circ}{x}_i)_p, \ldots, j_k((x_{i_1} - \overset{\circ}{x}_{i_1})^{\alpha_1} \ldots (x_{i_k} - \overset{\circ}{x}_{i_k})^{\alpha_n})_p$$

where $x_i^\circ = x_i(p)$ and $j_k(f)_p$ is the image of $f \in C^\infty(M)$ under the natural projection $C^\infty(M) \to J_p^k(M)$ called the *k-jet* of $f \in C^\infty(M)$ at $p \in M$; thus an element $j_k(f)_p$ can be written with respect to these coördinates as :

$$j_k(f)_p = (f_{(p)}, \left.\frac{\partial f}{\partial x_1}\right|_p, \ldots, \left.\frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}\right|_p) \quad |\alpha| \leq k$$

Let $\mathcal{U} = \bigcup_{p\in U} J_p^k(M)$ (open by definition) and every point $P \in \mathcal{U}$ is uniquely determined by the (n+1)-tuples $(x_1, \ldots, x_n, u_\alpha)$, $|\alpha| \leq k$, where x_i are the coordinates of $p \in U \subset M$, $P \in J_p^k(M)$ and u_α its coördinates in $J_p^k(u)$ relative to the basis introduced above. It is easily seen that if x_i and x_i' (coordinates of a point p in $U \cap U'$) are smoothly connected so are

(x_i, u_α) and (x'_i, u'_α).

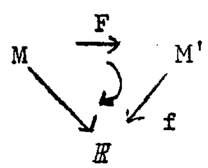
PR.3 The triples $(J^k(M), M, \pi_k)$ and $(\tilde{J}^k(M), M, \tilde{\pi}_k)$, where π_k and $\tilde{\pi}_k$ are defined by $\pi_k(j_k(f)_p)) = p$ and $\tilde{\pi}_k(j_k(f-f(0))_p) = p$, are vector bundles (Appendix III), called the *k-jet bundles* whose fibres at p are $J^k_p(M)$ and $\tilde{J}^k_p(M)$ respectively, with the natural vector space structure.

The map $\pi_{k,\ell} : J^k(M) \to J^\ell(M)$, $k \geq \ell$, defined by $\pi_{k,\ell}(j_k(f)_p) = j_\ell(f)_p$ is a morphism of vector bundles.

Let $J^k[M]$ and $\tilde{J}^k[M]$ be the $C^\infty(M)$-modules of smooth section of $(J^k(M), M, \pi_k)$ and $\tilde{J}^k(M), M, \tilde{\pi}_k)$ respectively; the vector bundle morphism $\pi_{k,\ell}$ defines a module hormorphine $\pi^{k,\ell} : J^k[M] \to J^\ell[M]$, $k \geq \ell$; $\tilde{J}^k[M]$ is the kernel of the homomorphism $\pi^{k,o}$. For example let $k = 0$; $J^\circ_p(M) = \mathbb{R}$ and $(J^\circ(M), M, \pi_o)$ is an one-dimensional vector bundle with trivializing section say, $j_o(\text{id.}) : M \to J^\circ(M)$ such that $j_o(\text{id.})(p) = p$ and thus $J^\circ(M) = M \times \mathbb{R}$; $J^\circ[M] = C^\infty(M)$. If $k = 1$ $\tilde{J}^1(M) = T^*(M)$ is the cotangent manifold to M and $\tilde{J}^1[M] = \Lambda^1(M)$ is the module of differentiable one-forms on M.

Given $F : M \to M'$, a smooth mapping we can define the ring homomorphism $F^* = C^\infty(M') \to C^\infty(M)$ defined by $F^*(f) \equiv f \circ F$, $f \in C^\infty(M')$.

$F^*(f) = f \circ F$



F^* in turn induces a homomorphism $J^k_{p'}(F)$ of the factor rings $J^k_{p'}(M') \to J^k_p(M)$, $p' = F(p)$, such that $J^k_{p'}(F)\ (j_k(f)_{p'}) \equiv j_k(F^*(f))_p$ and by taking the union of $J^k_{p'}(F)\ \forall\ p' \in M'$, a smooth mapping $J^k(F) : J^k(M') \to J^k(M)$.

$$\begin{array}{ccc} C^{\infty}(M) & \xleftarrow{F^{*}} & C^{\infty}(M') \\ {\scriptstyle j'_k(F^{*}(f))_p}\downarrow & \circlearrowleft & \downarrow{\scriptstyle j_k(f)_{p'}} \\ J^k_p(M) & \xleftarrow{J^k_{p'}(F)} & J^k_{p'}(M') \end{array}$$

A module homomorphism $J^k[F] : J^k[M'] \to J^k[M]$ over F^* can also be defined by:

$$J^k[F]\ (\theta)(p) \equiv J^k_{p'}(F)(\theta(p')),\ p' = F(p),\quad \theta \in J^*[M']$$

Similarly for $\tilde{J}^k[F]$.

$$\begin{array}{ccc} J^k(M') & \xrightarrow{J^k(F)} & J^k(M) \\ \theta\uparrow & \circlearrowleft & \uparrow{\scriptstyle J^k[F](\theta)} \\ p' \in M' & \xleftarrow{F} & M \ni p \end{array}$$

PR.4 From every smooth manifold M we have $J^k[M] = \tilde{J}^k[M] \oplus J^\circ[M]$;
if furthermore $F : M \to M'$ is smooth, then $J^k[F] = \tilde{J}^k[F] \oplus J^\circ[F]$

PROOF (V.V. LYCHAGIN) Define an $i_k = J^\circ[M] \to J^k[M]$ by $i_k(f)(p) = j_k(f)_p$, $p \in M$, $f \in C^\infty(M)$;

$$C^\infty(M) = J^\circ[M] \xrightarrow{i_k} J^k[M] \xrightarrow{\pi^{k,0}} C^\infty(M)$$

we have $\pi^{k,0} \circ i_k = \mathrm{id}$ and therefore $J^k[M] = \mathrm{Ker}\ \pi^{k,o} \oplus \mathrm{Im}\ i_k$.
The decomposition of the maps follows from the fact that $J^k[F]$ commutes both with $\pi^{k,o}$ and i_k.

As a result $J^1[M] = \Lambda^1(M) \oplus C^\infty(M)$ and therefore $J^1(M) = T^*(M) \times \mathbb{R}$; a projection $\pi : J^1(M) \to T^*(M)$ onto the first component and an injection $\alpha : T^*(M) \to J^1(M)$ onto the $T^*(M) \times \{0\}$ are defined in a natural way.

We also define two differential operators

DFN.1 $D_k = C^\infty(M) \to J^k[M]$, $k \geq 1$, is defined by $D_k(f)(p) = j_k(f)(p)$, $p \in M$, $f \in C^\infty(M)$ is a differential operator of order k.

For example $D_1 = C^\infty(M) \to J^1[M]$ in view of the direct decomposition $(\Lambda^1(M) \oplus C^\infty(M))$ of $J^1[M]$ can be written in the form $D_1(f) = (df, f)$.

DFN.2 $D = J^1[M] \to \Lambda^1(M)$ is defined by $D(\omega, f) = df - \omega$, $\omega \in \Lambda^1(M)$

PR.5 The operator $D = J^1[M] \to \Lambda^1(M)$ has the properties

1st Ker $D = \mathrm{Im}\, D_1$ i.e. the sequence

$$0 \to C^\infty(M) \xrightarrow{D_1} J^1 M \xrightarrow{D} \Lambda^1(M) \to 0$$

is exact on every smooth manifold.

2nd D is natural with respect to every smooth mapping $F = M \to M'$ i.e. $D \circ J^1[F] = F^* \circ D$ where $F^* = \Lambda^1(M') \to \Lambda^1(M)$

PROOF From the definitiin given, the decomposition $J^1[F] = \tilde{J}^1[F] \oplus J^\circ[F] = F^* \oplus F^*$ and the fact the operator of outer differentiation d is natural.

The first star indicates the transpose of the derivative map F_* and not a pull-back action; each interpretation is usually easily infered from the context.

DFN.3 A non-linear first order partial differential equation on M is a submanifold E of $J^1(M)$. A solution of $E \subset J^1(M)$ is a function $f \in C^\infty(M)$ such that $D_1(f)(M) \subset E$.

To see how this definition of equation and solution connect with the classical one let us introduce in $J^1(M)$ the local coördinates (x_i, u, u_j), $i,j = 1,..,n$, $u_\alpha = u$ if $\alpha = (0,\dots,0)$ and $u_\alpha = u_j$ if $\alpha = (0,\dots,1,\dots,0)$ with 1 in the jth position; for every point in E we can find a neighbourhood U of it, in which $E \cap U$ can be represented by the equations

$$E_i(x_1,\dots,x_u, u, u_1,\dots,u_n) = 0 \quad 1 \leq i \leq k$$

where codim $E = k$; if E considered as manifold has dimension $(n-k)$ codim E equals $\dim(T(M)/T(E))(T(M), T(E)$ are the tangent spaces to M and E respectively); in the coördinate system introduced, the section

$D_1(f)$ can be written as

$$D_1(f)(x_1,\ldots,x_u) = (x_1,\ldots,x_u,\ f(x_i),\ \frac{\partial f}{\partial x_1},\ldots,\frac{\partial f}{\partial x_n})$$

and the condition for the image of $D_1(f)(M)$ in U to lie in $E \cap U$ means

$$E_i(x_1,\ldots,x_u,f(x_i),\ \frac{\partial f}{\partial x_i},\ldots,\frac{\partial f}{\partial x_n}) = 0 \qquad 1 \leq i \leq k$$

which is the usual representation.

In a space-time the equation we are interested in, is equation (3.5) with a constant on the right-hand side (i.e. $g^{\alpha\beta}\ \frac{\partial f}{\partial x_\alpha}\ \frac{\partial f}{\partial x_\beta} = \text{const.}$)

The existence of the pseudo-Riemannian, non-degenerate, metric tensor field g with components $g_{\alpha\beta}(g^{\alpha\beta})$, $\alpha,\beta = 0, 1,2,3$, implies that there exists a vector bundle diffeomorphism $g_\flat = T(M) \to T^*(M)$ with inverse $g_\# = g_\flat^{-1}$ defined by $g_\flat(X)\ Y) = g(X,Y)\big|_p$, where $X,\ Y \in T_p(M)$, $\forall\, p_\in M$, $g_\flat$ being clearly a bijection and an isomorphian on each fiber, smoothness is proved by considering the local representative of $g_\flat$ with respect to the natural charts in $T^*(M)$ and the smoothness of g (R. ABRAHAM, J. MARSDEN, Proposition 13.7); $g_\flat$ and $g_\#$ are known as the lowering and raising indices operators and involve a summation with respect to one index of $g^{\alpha\beta}$ and $g_{\alpha\beta}$ respectively. g (as any other tensor field) is considered here and in many other places elsewhere as an $C^\infty(M)$-multilinear map from a cross-product of the appropriate combination of $C^\infty(M)$-modules of sections of the vector bundles $T(M)$ and $T^*(M)$ (vector and covector bundles) into the reals.

Then define the function $H = T^*(M) \to \mathbb{R}$ by $H(p,\omega) = \frac{1}{2}g(g_\#(\omega),g_\#(\omega))\big|_p$, $\forall\ \omega \in T^*(M)$, known as the *Hamiltonian function*. This is consistent with relation (3.5) or the general convention that a superscript denotes a contravariant component and a subscript a covariant one, i.e. the $\frac{\partial f}{\partial x_\alpha}$, $\alpha = 1,\ldots,n$, constitute in general a covariant vector, f_α, $\alpha=1,\ldots,n$, the components of the one-form df on M.

According to DFN. 3, 3.2, $H(p,\omega)$ = const·is a submanifold of $J^1(M)$ of codimension one, but the reason why we are considering $H(p,\omega)$ = const. as a submanifold of $J^1(M)$ rather than of $T^*(M)$, will not become evident till we further explore the geometric structures related to the concept of an equation and its solution.

We endeavour to extract geometrically from the set of all smooth sections $J^1[M]$ of $(J^1(M), M, \pi_1)$ those that are integrable, that is, have the form $D_1(f)$ for some $f \in C^\infty(M)$. This is done with the help of classifying (universal) elements $\omega \in \Lambda^1(T^*(M))$ and $\Omega \in \Lambda^1(J^1(M))$ defined as follows.

DFN.4 $\omega \in \Lambda^1(T^*(M))$ is defined by $\omega(X) = \Pi_3(X)\Pi_{1*}(X)$, $\forall\, X \in T(T^*(M))$, where Π_3 and Π_{1*} are shown in the commutative diagram

$$\begin{array}{ccc} T(T^*(M)) & \xrightarrow{\Pi_{1*}} & T(M) \\ \Pi_3 \downarrow & \circlearrowright & \downarrow \Pi_2 \\ T^*(M) & \xrightarrow{\Pi_1} & M \end{array}$$

with Π_1 , Π_2 , Π_3, all vector bundle projections.

Let (x_i, u_i), $i = 1,\dots,n$, be the coördinates of a chart in $T^*(M)$ so defined that u_i are the so called *conjugate* coördinates to x_i (dx_i are taken as a covector base for $T^*_{x_i}(M)$); then $\omega(x_i,u_i) = \sum_{i=1}^{n} u_i dx_i$;

$$dw(x_i,u_i) = \sum_{i=1}^{n} dx_i \wedge du_i$$ is a non degenerate closed two-form on $T^*(M)$

or a *symplectic* form ; thus the cotangent bundle of every smooth manifold carries a natural symplectic structure; ω and $d\omega$ are known as the *canonical* (or normal) forms on $T^*(M)$ and the coördinates chosen, the *canonical* coördinates.

PR.6 For an one-form $\theta \in \Lambda^1(M)$ a necessary and sufficient condition to be closed (i.e. $d\theta = 0$) is $\theta^*(d\omega) = 0$

PROOF θ and ω $(d\omega)$ are considered as sections of the contangent and the second cotangent bundle of M; the star indicates the covariant functor

$$M \xrightarrow{\theta} T^*(M) \xrightarrow{d\omega} T^*(T^*(M)) \xrightarrow{\theta^*} T^*(M)$$

Let $\theta(x_j) = \sum_{i=1}^{n} a_i(x_j)dx_i$ and therefore

$$d\theta(x_j) = \sum_{i=1}^{n} \frac{\partial a_i}{\partial x_j}(x_j)dx_j \wedge dx_i;$$

$d\theta = 0$ if and only if $\frac{\partial a_i}{\partial x_j} = \frac{\partial a_j}{\partial x_i}$; to evaluate $\theta^*(d\omega)$ we must compute $d\omega(X,X')$ for X, X' tangent to graph(θ); a base for the tangent space of graph (θ) is given by $X_i = \frac{\partial}{\partial x_i} + \sum_{j=1}^{n} \frac{\partial a_j}{\partial x_i}\frac{\partial}{\partial x_j}$ $(i = 1,\dots,n)$ and $\theta^*(d\omega) = 0$ iff $\frac{\partial a_i}{\partial x_j} = \frac{\partial a_j}{\partial x_i}$.

A closed form is by the Poincaré Lemma locally exact and therefore θ can be written as df for some $f \in C^\infty(M)$ defined by $f(x) = \int_\gamma \theta$; but this obviously makes it dependent on the homotopy class of the path γ and the criterion breaks down globally for non simply connected manifolds. By replacing $\Lambda^1(M)$ by $J^1[M]$ (i.e. working on the first jet bundle) we avoid this obstacle.

PR.7 A section $\Theta \in J^1[M]$ is integrable, that is $\Theta = D_1(f)$ for some $f \in C^\infty(M)$ iff $\Theta^*(\Omega) = 0$, where Ω is defined as follows:

DFN.5 $\Omega = D(\rho)$ (D in the differential operator of DFN.2, 3.2) and $\rho \in J^1[J^1(M)]$ is defined by the following proposition :

PR.8 There exists a unique element $\rho \in J^1[J^1(M)]$ such that $\forall\ \Theta \in J^1[M], J^1[\Theta](\rho) = \Theta$

$$J^1[J^1(M)] \xrightarrow{J^1[\Theta]} J^1[M]$$

PROOF Define $\rho\ \forall\ P \in J^1(M)$ as $\rho\big|_P = j_1(\pi_1^*(f))_P$ where P may be interpreted as the one-jet of some function $f \in C^\infty(M)$ at the point $p = \pi_1(P)$ i.e. $P = j_1(f)_p$

$$\mathbb{R} \xleftarrow{\pi_1^*(f) = f\circ\pi_1} J^1(M)$$

$$f \qquad \pi_1 \qquad \Theta$$

$$M$$

$$J^1[\Theta](\rho)\big|_P = J^1[\Theta](j_1(\pi_1^*(f))_P) \quad \text{by definition}$$

$$= J_p^1(\Theta)(j_1(\pi_1^*(f))_P) \quad \text{by definition of the } J^k[\ldots] \text{ map}$$

$$= j_1(\Theta^* \circ \pi_1^*(f))_p \qquad P = \Theta(p)$$

$$= j_1(f)_p \qquad \text{as } \Theta\circ\pi_1 = \text{id}.$$

$$= P = \Theta(p).$$

PROOF of PR.7, 3.2:

By PR.5, 3.2, $\Theta = D_1(f)$ if and only if $D(\Theta) = 0$; By PR.8, 3.2, $\Theta = J^1[\Theta](\rho)$ and $D(\Theta) = D(J^1[\Theta](\rho)) = \Theta^*(D(\rho))$ by PR.5, 3.2

$= \Theta^*(\Omega)$ by DFN.5,3.2.

In the special coordinate system (x_i, u, u_i), $i = 1,\ldots,n$, introduced in exemplifying DFN. 3, 3.2, we can choose for $f(P = j_1(f)_p)$ the function

$u^\circ + \sum_{i=1}^{n} u_i^\circ(x_1 - x_i^\circ)$ so that $(x_i^\circ, u^\circ, u_i^\circ)$ are the coördinate of P;

Therefore $\rho|_P = (u^\circ, \sum_{i=1}^{n} u_i^\circ dx_i)$ and in general $\rho = (u, \sum_{i=1}^{n} u_i dx_i)$ and

$\Omega = du - \sum_{i=1}^{n} u_i dx_i$. The injection $\alpha = T^*(M) \to J^1(M)$ induced map

$\alpha^* : \Lambda^1(J^1(M)) \to \Lambda^1(T^*(M))$ relates ω and Ω by $\omega = \alpha^*(\Omega)$

Note that the sequence :

$$0 \longrightarrow C^\infty(M) \xrightarrow{D_1} J^1[M] \xrightarrow{D} \Lambda^1(M) \longrightarrow 0$$

is exact for every smooth manifold M, in contrast to the de Rham sequence. The one-form Ω defined in DFN. 5,3.2, and PR.8, 3.2, endows $J^1(M)$ with a *contact structure*, since $\Omega \wedge (d\Omega)^n \neq 0$ everywhere on $J^1(M)$ (the exponent denotes the exterior power) (D.E. BLAIR, chapter 1, §1). In view of PR. 7, 3.2, it seems natural to restrict the class of diffeomorphisms of $J^1(M)$ to those which preserve the kernel of the form Ω and so must multiply Ω by some function; this motivates the following definition

DFN.6 A diffeomorphism ϕ: $J^1(M) \to J^1(M)$ is said to be *contact (strict contact) diffeomorphism* if and only if $\phi^*(\Omega) = f\cdot\Omega$, $f \in C^\infty(J^1(M))$ $(\phi^*(\Omega) = \Omega)$

The local picture of a contact manifold in general is much more suggestive of a geometric interpretion. A differentiable manifold M^{2n+1} is said to be a *contact manifold in the wider sense* iff it admits an atlas (O_α, f_α) $\alpha \in I$ such that (O_α) , $\alpha \in I$, is an open covering of M^{2n+1}, $f_\alpha : O_\alpha \to V_\alpha \subset \mathbb{R}^{2n+1}$ V_α open, and $f_\alpha \circ f_\beta^{-1} \in F(\mathbb{R}^{2n+1})$, where $F(\mathbb{R}^{2n+1})$ is a collection of diffeomorphisms of $\mathbb{R}^{2n+1}$ with $f^*\Omega = f\cdot\Omega$,

$\Omega = du - \sum_{i=1}^{n} u_i dx_i$, $f \in C^\infty(\mathbb{R}^{2n+1})$, $f \neq 0$ (i.e. contact diffeomorphisms)

and (x_i, u, u_i), $i = 1,\dots,n$, cartesian coordinates on $\mathbb{R}^{2n+1}$($F(\mathbb{R}^{2n+1})$ is a pseudogroup being closed under composition, formation of inverses and restriction to open subsets of $\mathbb{R}^{2n+1}$). An equivalence class of atlases ((O_α, f_α) and (O'_α, f'_α) are said to be equivalent iff $f'_\alpha \circ f_\alpha^{-1} \in F(\mathbb{R}^{2n+1})$, whenever defined) is called *a contact structure in the wider sense* on M^{2n+1}. By Darboux Theorem (S. STENBERG, Chapter III, Theorem 6.2) it is then easily seen that a contact manifold (i.e. one that carries a global differential one-form Ω such that $\Omega \wedge (d\Omega)^n \neq 0$ everywhere) is a contact manifold in the wider sense. The converse is true if M^{2n+1} is orientable and n even (D.E. BLAIR, chapter 1, §1); for the first-jet bundle of a space-time (n=4) the two concepts are therefore equivalent.

If M^{2n+1} is a contact manifold in the wider sense then there exists an open covering (O_α), $\alpha \in I$, of M^{2n+1} and locally defined contact forms Ω_α on O_α. We can define an 2n-dimensional subbundle D of the tangent bundle $T(M^{2n+1})$ of M^{2n+1} with fibres given by $D_p = \{X \in T_p(M^{2n+1}) : \Omega_\alpha(X) = 0, \alpha \in I, p \in O_\alpha\}$, the so called *contact distribution*. Since Ω_α is contact on O_α, $(d\Omega_\alpha)^n \neq 0$ on D_p and $d\Omega_\alpha$ is a non-degenerate, skew-symmetric , bilinear form on D_p uniquely determined to within a non-zero multiple, i.e. $\Omega_\alpha = m_{\alpha\beta}\Omega_\beta$ on $O_\alpha \cap O_\beta$; in the transformation law for the (2n+1)-form $\Omega_\alpha \wedge (d\Omega_\alpha)^n$, $\Omega_\alpha \wedge (d\Omega_\alpha)^n = m_{\alpha\beta}^{n+1} (\Omega_\beta \wedge (d\Omega_\beta)^n)$, $m_{\alpha\beta}^{n+1}$ is the Jacobian of the coördinate transformation which is always positive if M^{2n+1} is orientable; for n even, $m_{\alpha\beta}$ has to be positive and D is therefore orientable. The quotient bundle $T(M^{2n+1})/D$ is then an orientable real line bundle and therefore admits a cross-section without zeros. Thus M^{2n+1} admits a global non-vanishing vector field X_c such that :

$$\Omega(X_c) = 1 \text{ and } d\Omega(X_c, Y) = 0 \quad \forall Y \in T(M) \tag{3.41}$$

the so called *characteristic vector field* of the contact structure. As a result the local cross-sections X_α over O_α, defined by the equations $\Omega_\alpha(X_\alpha) = 1 \ \forall\, a \in I$, satisfy $X_\alpha = m_\alpha X_c$ where m_α's are non-vanishing functions of the same sign. By defining $\Omega = m_\alpha \Omega_\alpha$ on O_α we obtain a global one-form Ω such that $\Omega \wedge (d\Omega)^n \neq 0$, Q.E.D.

For example the one form $d\Omega = du - \sum_{i=1}^{n} u_i dx_i$ on $\mathbb{R}^{2n+1}$ with Cartesian coördinates (x_i, u, u_i) makes it into a contact manifold with characteristic vector field X_c given by $\frac{\partial}{\partial u}$ and the contact distribution D spanned by

$$X_i = \frac{\partial}{\partial x_i} + u_i \frac{\partial}{\partial u} \quad \text{and} \quad X_{n+i} = \frac{\partial}{\partial u_i} \qquad i = 1,\dots,n.$$

The condition $\Omega \wedge (d\Omega)^n \neq 0$ means that the contact distribution D is as far from being integrable as possible. The first thing to be noted is the fact that there do not exist integral submanifolds of the contact distribution on M^{2n+1} of dimension higher than n. For if $\{X_i\}$, $i = 1,\dots,m$, $m > n$, be m linearly independent vector fields which belong to the tangent space of such an m-dimensional integral submanifold we may extend these to a basis to the tangent space of the whole manifold by $X_{m+1},\dots,X_{2n}$, $X_{2n+1} = X_c$ and note that $\Omega(X_i) = 0$ and $d\Omega(X_i,X_j) = \frac{1}{2}(X_i\Omega(X_j) - X_j\,\Omega(X_i) - \Omega([X_i,X_j])) = 0$, $\forall\, i,j = 1,\dots,m$; then, since $m > n$, $\Omega \wedge (d\Omega)^n (X_1,\dots,X_{2n+1}) = 0$, a contradiction. The conditions that Ω and $d\Omega$ vanish when restricted to a submanifold are also sufficient for the submanifold to be an integral submanifold of D, because if $0 = d\Omega(X,Y) = -\frac{1}{2}\Omega([X,Y])$, $[X,Y]$ belongs to D (i.e. D is involutive) and by the Frobevnous theorem (C. CHEVALLEY, p. 94) unique and maximal integrability is ensured.

There is a large number of integral submanifolds of D, for the study of which an added difficulty is the fact that $d\Omega$ vanishes along the submanifold and so one does not have an induced structure. Indeed one can prove that the vanishing of Ω and $d\Omega$ on m linearly independent vectors $(m \leq n)$ at a point is sufficient for the existence of an m-dimensional integral submanifold tangent to them. Furthermore there exist always an m-dimensional integral submanifold of D $(1 \leq m \leq n)$ through a point p and such that a given vector field at p is tangent to it (S. SASAKI and D.E. BLAIR, chapter III)

PR. 9 $\mathcal{L}_{X_c}(\Omega) = 0$ and $\mathcal{L}_{X_c}(d\Omega) = 0$, where $\mathcal{L}_{X_c}$ denotes Lie differentiation with respect to the characteristic contact vector field X_c.

PROOF From $\mathcal{L}_X = d \circ i_X + i_X \circ d$, where i_X is the interior product with X, and relation (3.41).

Applying DFN. 6, 3.2, a local, one-parameter group of contact diffeomorphisms ϕ_w of $J^1(M)$, is defined by $\phi_w^*(\Omega) = f_w \cdot \Omega$, $f_w \in C^\infty(J^1(M))$ and by taking the Lie derivative with respect to the corresponding vector field

$$\mathcal{L}_Y(\Omega) = f \cdot \Omega \ , \text{ where } f = \left.\frac{df_w}{dw}\right|_{w=0} \tag{3.42}$$

On the basis of this remark we make the following definitions

DFN. 7 An *infinitesimal contact transformation* or a *contact vector field* is a vector field on $J^1(M)$ satisfying (3.42).

PR. 10 A vector field X on $J^1(M)$ is a contact vector field iff the group of translation along X is a one-parameter group of contact diffeomorphisms.

PROOF Condition (3.42) is equivalent to $\mathcal{L}_X(\Omega) \wedge \Omega = 0$, it can also be written in the form

$$\left.\frac{d}{dw}\right|_{w=0} (\phi_w^*(\Omega)) \wedge \Omega = 0$$

Since $\tau_{w+v} = \tau_w \circ \tau_v (\tau_w$ is the group of translations along X)

$$\left.\frac{d}{dw}\right|_{w=v} (\phi_w^*(\Omega)) \wedge \Omega = 0$$

Therefore $\phi_w^*(\Omega) \wedge \Omega = \phi_o^*(\Omega) \wedge \Omega = 0$ and ϕ_w are contact diffeomorphisms.

For example, the characteristic vector field X_c (defined by relations (3.41)) is contact with f_w = constant.

There exists an isomorphism between contact vector fields on $J^1(M)$ and smooth functions on $J^1(M)$, guaranteed by the following proposition:

PR.11 Every contact vector field X on $J^1(M)$ is uniquely determined by the function $f = \Omega(X)$. To every function $f \in C^\infty(J^1(M))$ there corresponds a unique contact vector field X_f such that

(1) $\Omega(X_f) = f$

(2) $\mathcal{L}_{X_f}(\Omega) = X_c(f)\cdot\Omega$

(3) $X_{f+g} = X_f + X_g$, $g \in C^\infty(J^1(M))$

(4) $X_{f\cdot g} = f\cdot X_g + g\cdot X_f - f\cdot g\cdot X_c$

(5) $X_f(f) = X_c(f)\cdot f$

PROOF (V.V. LYCHAGIN, theorem 1.4.3); see also the derivation of relation (3.44), 3.2, below.

DFN. 8 The function $f = \Omega(X)$ is called the *Hamiltonian* of the contact vector field X on $J^1(M)$

The question arises how to find X_f when $f \in C^\infty(J^1(M))$ is given; if Y is a contact vector field by DFN. 7 and PR. 10, 3.2, $\mathcal{L}_Y(\Omega) = g\cdot\Omega$; we take $Y = f\cdot X_c + Z$ where $\Omega(Z) = 0$; then

$$\begin{aligned}\mathcal{L}_Y(\Omega) &= g\cdot\Omega \\ &= \mathcal{L}_{f\cdot X_c}(\Omega) + \mathcal{L}_Z(\Omega) \quad \text{by substitution} \\ &= f\cdot\mathcal{L}_{X_c}(\Omega) + df \wedge i_{X_c}(\Omega) + \mathcal{L}_Z(\Omega) \quad \text{by the properties of Lie derivation} \\ &= df + d\circ i_Z(\Omega) + i_Z\circ d\Omega \quad \text{by PR. 9, 3.2, and relation (3.41)}\end{aligned}$$

hence

$$i_Z\circ d\Omega = g\cdot\Omega - df \tag{3.43}$$

By applying (3.43) on X_c we have

$$d\Omega(Z,X_c) = g\cdot\Omega(X_c) - X_c(f)$$

and by virtue of the relation (3.41) $g = X_c(f)$. It is evident that by taking $X_f = f\cdot X_c + Z$, where Z is defined by equation (3.43) with $g = X_c(f)$, we obtain the required field. In the local coördinates (x_i,u,u_i) $i = 1,\ldots,n$ on $J^1(M)$, X_c has the form $\frac{\partial}{\partial u}$ (the trajectories of X_c are the fibres of $(J^1(M),T^*(M),\pi))$, $g = \frac{\partial f}{\partial u}$ and the relation (3.43) reads :

$$- Z_{n+i}dx_i + Z_i, du_i = -\left(u_i \frac{\partial f}{\partial u} + \frac{\partial f}{\partial x_i}\right)dx_i - \frac{\partial f}{\partial u_i}du_i \qquad i = 1,\ldots,n$$

which determines all but the $\frac{\partial}{\partial u}$ components of Z i.e.

$$X_f = - \sum_{i=1}^{n} \frac{\partial f}{\partial u_i} \frac{\partial}{\partial x_i} + (f + ?) \frac{\partial}{\partial u} + \sum_{i=1}^{n}\left(u_i \frac{\partial f}{\partial u} + \frac{\partial f}{\partial x_i}\right) \frac{\partial}{\partial u_i}$$

Condition (5) of PR. 11, 3.2, implies that the missing term is $- u_i \frac{\partial f}{\partial u_i}$

and therefore

$$X_f = - \sum_{i=1}^{n} \frac{\partial f}{\partial u_i} \frac{\partial}{\partial x_i} + \left(f - \sum_{i=1}^{n} u_i \frac{\partial f}{\partial u_i}\right) \frac{\partial}{\partial u} + \sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i} + u_i \frac{\partial f}{\partial u}\right)\frac{\partial}{\partial u_i} \qquad \ldots\ldots \quad (3.44)$$

For example take the Hamiltonian function H defined on the cotangent bundle of a spacetime; every function in $C^\infty(T^*(M))$ can be regarded as a smooth function on $J^1(M)$ via the projection $\pi : J^1(M) \to T^*(M)$ and H on $T^*(M)$ induces $\pi^*(H)$ on $J^1(M)$. Relation (3.44) yields for the corresponding contact (Hamiltonian) vector field :

$$X_{\pi^*(H)} = - \sum_{\alpha,\beta=0}^{3} g^{\alpha\beta}(x_\alpha)u_\beta \frac{\partial}{\partial x_\alpha} - \left(\tfrac{1}{2} \sum_{\alpha,\beta=0}^{3} g^{\alpha\beta}(x_\alpha)u_\alpha u_\beta\right)\frac{\partial}{\partial u} + \tfrac{1}{2} \sum_{\alpha,\beta,\gamma=0}^{3} \frac{\partial g^{\alpha\beta}(x_\alpha)}{\partial x_\gamma} u_\alpha u_\beta \frac{\partial}{\partial u_\gamma}$$

and its integral curves $(x_\alpha(w),\ u(w),\ u_\alpha(w))$, $\alpha = 0, 1, 2, 3$, are given by the equations :

$$\frac{dx_\alpha}{dw} = - \sum_{\beta=0}^{3} g^{\alpha\beta}(x_\alpha(w))u_\beta(w) \qquad \alpha = 0, 1, 2, 3 \tag{3.45}$$

$$\frac{du_\gamma}{dw} = \tfrac{1}{2} \sum_{\alpha,\beta=0}^{3} \frac{\partial g^{\alpha\beta}(x_\alpha(w))}{\partial x_\gamma} u_\alpha(w) u_\beta(w) \qquad \gamma = 0, 1, 2, 3 \tag{3.46}$$

$$\frac{du}{dw} = -\tfrac{1}{2} \sum_{\alpha,\beta=0}^{3} g^{\alpha\beta}(x_\alpha(w)) u_\alpha(w) u_\beta(w) \tag{3.47}$$

Equations (3.45) and (3.46), are the canonical differential equations (3.18) and (3.19) (the sign difference is due to the definition of X_c in general and can be accomodated by a reflection :

$X_{\pi^*(H)}$ goes into $-X_{\pi^*(H)}$; all it really matters is the sign difference between the α and $n+\alpha$ terms which cannot be altered by any such transformation); a geometric interpretation of equation (3.47) appears in PR. 18, 3.3, below.

To conclude this section, we give the following definition for a generalized, many-valued, solution of a first-order partial differential equation

DFN. 9 A (many-valued) solution of an equation $E \subset J^1(M)$ is a submanifold $L \underset{i}{\hookrightarrow} J^1(M)$, $\dim L = \dim M$, $i^*(\Omega) = 0$ where Ω is given by DFN. 5, 3.2, and lying in E.

Similarly for equations in $T^*(M)$, with Ω in the definition above replaced by the standard (normal) canonical two-form $d\omega$, (DFN.4), 3.2. DFN. 9, 3.2 is consistent with DFN. 3, 3.2 for the ordinary (single-valued) solution determined by a single function $f \in C^\infty(M)$. The map $D_1(f) : M \to J^1(M)$ is an embedding; that is, the section $D_1(f) = (df, f)$ of $J^1(M)$ is an immersion, which implies that it has no critical points or equivalently that the derivative map $T_p(D_1(f))$ from $T_p(M)$ to $T_P(J^1(M))$, $\pi(P) = p$, $\forall\, p \in M$ is of maximal rank, i.e. n, as it can be seen from its matrix form.

$$\left(I_n \;\vdots\; \frac{\partial^2 f}{\partial x_i \partial x_j} \;\vdots\; \frac{\partial f}{\partial x_i} \right);$$

and also $D_1(f)$ being a section, i.e. $\pi_1 \circ D_1(f) = id_M$ is differentiable and it is homeomorphic onto its image $D_1(f)(M)$ because the inverse under $D_1(f)$ of each member of the subbase for the topology for $D_1(f)(M)\Big|_{D_1(f)(M)}$, induced by the manifold topology of $J^1(M)$ which in turn is defined by ϕ^{-1}, $\phi : \pi^{-1}(U) \to U \times \mathbb{R}^n \times \mathbb{R}$, U open in M, is U, i.e. open in M. Furthermore by PR.7, 3.2 $i^*(\Omega) = 0$ with $i : D_1(f)(M) \to J^1(M)$, since $D_1(f)$ is an integrable section. To a many--valued solution corresponds an ordinary one iff submanifold $L = \theta(M)$ for some section $\theta = M \to J^1(M)$, that is to say when the restriction $\pi_1\big|_L : L \to M$ of π_1 to L is a diffeomorphism.

We consider a differential equation of the form :

$$E = \{P \in J^1(M) : f(P) = 0.\ f \in C^\infty(J^1(M))\}$$

then

PR. 12 A solution L(DFN.9, 3.2) of E is invariant under X_f, the contact vector field corresponding to f and defined by PR. 11, 3.2.

PROOF: We must prove that $X_{f(P)} \in T_P(L)$ (where $T_P(L)$ is the tangent space to L at P); let V be the linear span of $T_P(L)$ and $X_{f(P)}$ in $T_P(J^1(M))$; V has the following two properties; first $\Omega(Y) = 0$, $\forall\, Y \in V$ and second $d\Omega(Y_1,Y_2) = 0$, $\forall\, Y_1, Y_2 \in V$; the first property is obvious from DFN. 9, 3.2, as $\Omega\big|_L = 0$ and $\Omega(X_f)(P) = f(P) = 0$ from the definition of the equation E ; to establish the second it is enough to take $Y_1 = X_{f(P)}$ and $Y_2 \in T_P(L)$ since the second property is true for Y_1 and Y_2, in $T_P(L)$ ($\pounds_{Y_1}(\Omega)(Y_2) = (d \circ i_{Y_1}\Omega + i_{Y_1} \circ d\Omega)(Y_2) = d\Omega(Y_1,Y_2)$ as $\Omega(Y_1) = 0$, and $\pounds_{Y_1}(\Omega)$ is Ω on P', P' still in L);

$\pounds_{X_f}(\Omega)(Y_2) = X_c(f)\Omega(Y_2)$ by property (2) of PR. 11, 3.2,

$= (d \circ i_{X_f}\Omega + i_{X_f} \circ d\Omega)(Y_2)$ by the Lie derivative relation to exterior and and interior differentiation

By the definition of interior differentiation $i_{X_f}\Omega = \Omega(X_f)$ (= f by property (1) of PR. 11, 3.2) and $(i_{X_f} d\Omega)(Y_2) = d\Omega(X_f, Y_2)$, therefore $d\Omega(X_f(P), Y_2(P)) = X_c(f)(P)\cdot\Omega(Y_2)(P) - df(Y_2)(P) = 0$; these conditions are sufficient for a submanifold to be an integral submanifold of the contact distribution and as such of dimension no greater than n, i.e. dim $V \leq n$, and since dim $T_P(L) = n$, $X_{f(P)} \in T_P(L)$.

3.3 Legendrian and Lagrangian mappings and their singularities

Solutions of first order partial differential equations (DFN. 3 and DFN. 9, 3.2) were identified as integral manifolds of the highest possible dimension of the contact distribution, determined by the universal contact form (DFN. 5, 3.2), on the first jet bundle (PR. 3, 3.2); this motivates the following definitions .

DFN. 10 An integral manifold of a contact distribution, on a contact manifold M^{2n+1} is called a *Legendrian submanifold* if and only if it has the higest possible dimension, namely, n.

DFN. 11 A *Lagrangian submanifold* of a symplectic manifold M^{2n} is one of dimension n, on which the symplectic form pulls back to zero.

Legendrian and Lagrangian submanifolds are better understood as leaves of foliated structures, the Lagrangian and Legendrian fibrations.

DFN. 12 A fibration $\pi : M^{2n} \to B^n$ $(\pi : M^{2n+1} \to B^{n+1})$ is said to be *Lagrangian (Legendrian)* if and only if its fibres $\pi^{-1}(B^n)$ $(\pi^{-1}(B^{n+1}))$ are Lagrange (Legendre) submanifolds

DFN. 13 *A Lagrangian (Legendrian) equivalence* is a diffeomorphism of M^{2n} (M^{2n+1}) which preserves the symplectic (contact) structure and the structure of the fibration.

Locally,

PR. 13 All Lagrangian fibrations are locally Lagrange equivalent

PROOF By Darboux's theorem.

Therefore, any Lagrange fibration is locally Lagrange equivalent to the standard fibration $\pi : T^*(\mathbb{R}^n) \to \mathbb{R}^n$ with coördinates $x_i \in \mathbb{R}^n$, $u_i \in T^*_x(\mathbb{R}^n)$, $\pi(x_i, u_i) = x_i$ and form $d\omega = dx_i \wedge du_i$.

Similarly

PR. 14 All Legendrian fibrations are locally Legendre equivalent

PROOF By the invariant and unique character of the contact form Ω (DFN. 5 and PR. 8, 3.2).

For example, a local model of a Legendre fibration is the fibration $\pi : J^1(\mathbb{R}^n) \to \mathbb{R}^{n+1}$, with coordinates $x_i \in \mathbb{R}^n$, $u_i \in T^*(\mathbb{R}^n)$, $u \in \mathbb{R}$, $\pi(x_i, u, u_i) = (x_i, u)$ and contact form $\Omega = du - \sum_{i=1}^{n} u_i dx_i$.

DFN. 14 A *Lagrange (Legendre) mapping* is the composite of the embedding of a Lagrangian (Legendrian) submanifold in the total space of a Langrangian (Legendrian) fibration and the projection onto its base.

DFN. 15 Lagrange (Legendre) mappings are said to be equivalent if and only if there exists a Lagrangian (Legendrian)equivalence which carries the corresponding Lagrangian (Legendrian) submanifolds into each other.

This is of course one of many definitions of equivalence and in fact a very strong one; for example all smooth curves in the plane are Lagrangian submanifolds, but the curves $(x^2 + y^2 - 1)((x - \lambda)^2 + y^2 - 1) = 0$, $0 \leq \lambda \leq 1$, for different λ's are Lagrange inequivalent because of the invariance of the affine structure on the fibre of a Lagrangian fibration (V.I. ARNOL'D, 1, §10)

To define the corresponding local concepts, it is only necessary to replace manifolds and maps by their germs everywhere. However, although Lagrangian equivalence implies equivalence in the sense of ordinary smooth maps (i.e. f_1, $f_2 : M \to N$, f_1 equivalent to f_2 iff $\exists$ diffeomorphisms $h : M \to M$ and $k : N \to N$ $k \circ f_1 = f_2 \circ f$) both locally and at large, the converse proposition is not true as it is shown by the example of smooth equivalent but Lagrange inequivalent germs at zero x^3 and x^3+x^4. This is so, as there is not a diffeomorphism of a neighborhood of zero of the plane which carries $L_1 = \{(x,y) : y=x^3\}$ into $L_2 = \{(x,y) : y = x^3+x^4\}$, while there are $k, h \in \mathrm{Diff}(\mathbb{R}^2)$ such that $k \circ f_1 = f_2 \circ h$ with $f_1 = x^3$ and $f_2 = x^3+x^4$; for h, a given diffeomorphism, $k = f_2 \circ h \circ f_1^{-1}$ is a diffeomorphism in a neighborhood of zero because $f_2 \circ h$ is invertible in a neighborhood of zero, since f_2 is one-to-one in a neighborhood of zero ($x_1^3+x_1^4 = x_2^3+x_2^4$, $x_1 \neq x_2$, with $x_1 < 1$ is only possible with $x_2 < -1$, because $-1 < x_2 < 0$ implies $x_2^3 + x_3^4 < 0$).

PR. 15 Any germ of a Lagrangian (Legendrian) submanifold is well projected onto at least one of the 2^n (2^{n+1}) m-dimensional ((n+1)-dimensional) coördinate subspaces (x_I, u_J) ((x_I, u, u_J)) where $I \cup J = \{1,\ldots,n\}$ and $I \cap J = \emptyset$. In this case there exists a unique function $F(x_I, u_J)$ such that the germ of the Lagrangian (Lengendrian) submanifold is defined by the equaitons

$$x_J = x_J(x_I,u_J) = \frac{\partial F(x_I,u_J)}{\partial u_J} \qquad I \cup J = \{1,\ldots,n\}$$

$$I \cap J = \emptyset$$

$$u_I = u_I(x_I,u_J) = -\frac{\partial F(x_I,u_J)}{\partial x_I} \tag{3.48}$$

$$\Big(u = F(x_I,u_J) - \sum_{j\in J} x_j u_j\Big)$$

PROOF (V.I. ARNOL'D, 2, LEMMAS 3.1.1 and 3.3.1)

PR. 15, 3.2, above is a local result and in view of PR. 13 and PR. 14, 3.3, it is enough to prove that every Lagrangian (n--dimensional) plane in $\mathbb{R}^{2n}$ (defined by the condition that the skew-scalar or wedge product of any two vectors of the plane equals zero) is transversal to one of the 2^n- cöordinate planes $L_1 = \{(x_i,u_i) : x_I = 0,\ u_J = 0,\ I \cup J = \{1,2,\dots,n\}$ $I \cap J = \emptyset\}$; in fact it is enough to prove that the Lagrangian plane L, such that $L \cap L_n$ be k-dimensional, where $L_n = \{(x_i,u_i): x_I = 0\ I = \{1,\dots,n\}\}$, is transversal to one of the $\binom{k}{n}$ cöordinate planes $L_k = \{(x_i,u_i) : x_I = 0,\ u_J = 0,\ I = \{1,\dots,k\},$ $I \cap J = \emptyset,\ I \cup J = \{1,\dots,n\}\}$. Let $L_o = L \cap L_n$ be k-dimensional and in L_n transversal to one of the $\binom{k}{n}$ (n-k)-dimensional planes $T = L_k \cap L_n$ for some k i.e. $L_o \cap L_k \cap L_n = 0$; we want to show that the plane L_k is transversal to L i.e. $L_k \cap L = 0$. $L_o \subset L$ and $T \subset L_k$ and since L and L_k are Lagrangian L and L_k are skew-orthogonal to L_o and T respectively; However $L_o + T = L_u$ which is itself a maximal skew-orthogonal plane and hence $L \cap L_k \subset L_n$; as a result $L \cap L_k \subset L \cap L_k \cap L_n = L_o \cap T = 0$.

The function F of PR. 15, 3.3, is called the *generating function* of the germ of the Lagrangian (Legendrian) submanifold defined by relation (3.48); one can further show that:

PR. 16 Any germ of a Lagrangian mapping is Lagrange equivalent to the germ of a gradient map, i.e. any germ of a Lagrangian submanifold in the neighborhood of each of its points, in which the tangent space to the submanifold is transversal to the u- space, is given by

$$u_i = \frac{\partial G}{\partial x_i} \qquad i = 1,2,\ldots,n \tag{3.49}$$

PROOF By PR. 15, 3.3, the germ of the Lagrangian submanifold is defined by relation (3.48); under the Lagrangian equivalence (DFN. 13, 3.3) $x'_I = x_I$, $x'_J = x_J + \lambda u_J$, $u'_I = u_I$, $u'_J = u_J$, $\lambda \in \mathbb{R}$, $I \cup J = \{1,\ldots,n\}$, $I \cap J = \emptyset$, relations (3.48) become

$$x'_J = \frac{\partial F}{\partial u'_J} + \lambda u'_J \qquad u'_I = - \frac{\partial F}{\partial x'_J}$$

which suggest that the new generating function will be $F'(x'_I,u'_J) = F(x'_I,u'_J) - \frac{1}{2}\lambda \sum_{j\in J} u'^2_j$;

$G = \sum_{j\in J} x_j u_j - F(x_I,u_J)$ and we can substitute for u'_J for almost all values of λ , since the condition of local solvability of $x'_J = \frac{\partial F}{\partial u'_J} + \lambda u'_J$ is guaranteed by det $(\frac{\partial^2 F}{\partial u_J^2} - \lambda I) \neq 0$;

the exceptional λ's are finite in number, they are in fact the eigenvalues of the Hessian of F with respect to u_J.

The geodesic curves, equation (3.26), of a specetime M were identified as the integral curves, equation (3.45) and (3.46) of the Hamiltonian vector field $X_{\pi^*(H)}$ on the first jet bundle $J^1(M)$ over M, i.e. the integral curves of the contact vector field determined (PR. 11, 3.2) by the Hamiltonian function $H(p,\omega) = \frac{1}{2}\, g(g_\#(\omega), g_\#(\omega))$. So, geodesic curves are considered as families of curves in the first jet bundle. Of special interest are those families of null and/or timelike character emanating from a hypersurface in spacetime; it is exactly this biased view of spacetime geometry which depicts the moving (evolutionary) character from a spacetime (world) geometrical picture. Hypersurfaces in the four-dimensional spacetime manifold, given locally by an equation $f(x_\alpha) = 0$, $\alpha = 0, 1, 2, 4$, relation (3.7), can be divided into two classes; two-surfaces in motion, ($(\nabla f)^2 \leq 0$, relation (3.8)) with a displacement velocity equal or less the velocity of light and the whole of infinite

space ($(\nabla f)^2 > 0$) the various points of which are all taken at different instants of time, the time at which a point is taken, being determined by the (time) equation in such a way that their four-dimensional interval (relation (3.4)) between any two of them is always spacelike (negative).

PR. 17 A hypersurface in a spacetime M, together with its normal covector at each of its points define a representative of a germ of a Lagrangian submanifold of the cotangent bundle $T^*(M)$.

PROOF Let the hypersurface in M be given locally by $f(x_\alpha) = 0$, α : 0, 1, 2, 3) with say $\frac{\partial f}{\partial x_o} \neq 0$ $\left(\frac{\partial f}{\partial x_1} \neq 0\right)$, then one can choose as coördinates of the embedded submanifold in $T^*(M)$ the set (x^1, x^2, x^3, dx^o) $((x^o, x^x, x^3, dx^1))$ and any such submanifold is locally the graph of a closed one-form, namely df; hence by PR. 6, 3.2, $(df)^*(d\omega) = 0$ i.e. the symplectic form $d\omega$ pulls back to zero on the graph (df); its dimension is clearly $n = 4$ and by DFN. 10, 3.3, is a Lagrangian submanifold.

PR. 18 The set of points defined for every fixed value of the affine parameter on the geodesics (considered as families of curves in the first jet bundle of a spacetime) conormal to a hypersurface of spacetime (specifying the zero value of the affine parameter) form a Legendrian submanifold.

PROOF: By DFN. 9, 3.2 and DFN. 10, 3.3, a solution of the equation $E = \{P \in J^1(M) : \pi^*(H) = \text{constant},\ H \in C^\infty(T^*(M))\}$ is a Legendre submanifold lying in E. By PR. 12, 3.2, such a submanifold is invariant under $X_{\pi^*(H)}$. Relations (3.45) and (3.46) however, identify the integral curves of $X_{\pi^*(H)}$ with the geodesics of the spacetime; relation (3.47) determines the character (null ot timelike) of the family according to the value of the constant (zero or one respectively) in the definition of the equation E.

PR.19 Let L_o be an (n-1)-dimensional, smooth, isotropic submanifold of the cotangent bundle $T^*(M)$ such that $H|_{L_o}=o$ and $X_H(P)\in T(L_o)$, $\forall P\in L_o$. Then the (local) flow L of L_o along the integral curves of X_H in $H=o$, is a Lagrange submanifold on which $H=o$ and apparently locally the only one containing L_o.

PROOF: The statement in PR.19 is equivalent to saying that the Cauchy problem for the equation $E=\{P\in T^*(M): H(P)=o,\ H\in C^\infty(T^*(M))\}$ is well posed locally (V.V. LYCHAGIN, 1.5.3, Proposition) (L_o isotropic means $d\omega(X,Y)=o$, $\forall X,Y\in T(L_o)$).

We can state our main objective in this chapter; that is to relate the image (critical values) of Legendre (Lagrange) mappings for Legendrian (Lagrangian) submanifolds, such as the ones identified in PR. 17 and PR. 18, 3.3, with wavefronts (focal (conjugate) points) in General Relativistic spacetimes. Relations(3.48) give locally the form of a germ of a Legendrian (Lagrangian) submanifold; by DFN. 14, 3.3, the form of the corresponding Legendre (Lagrange) mapping is

$$(x_I, u, u_J) \xrightarrow{\pi \circ i} \left(x_I,\ u = F - \sum_{j \in J} u_i \frac{\partial F}{\partial u_j},\ x_J = \frac{\partial F}{\partial u_J} \right)$$

$$((x_I, u_J) \xrightarrow{\pi \circ i_1} \left(x_I,\ x_J = \frac{\partial F}{\partial u_J} \right) \tag{3.50}$$

where $F = F\,(x_I, u_J)$. The image of $\pi \circ i$ is called a *wavefront* ; the set of critical values of $\pi \circ i_1$ is called a *caustic* ; in the context of the contangent bundle model, the singular set consists of those points at which the Lagrangian submanifold cannot be written, locally, as the graph of a (closed) one-form or equivalently those points at which the Lagrangian submanifold is not transverse to the fibres of the cotangent bundle.

Our aim next is to demonstrate how the study of focal or conjugate points to a hypersurface in specetime is the same with that of the critical values of certain Lagrangian mappings.

The propagation mechanism given by H, the Hamiltonian function $H(x,\omega) = \frac{1}{2}g(x)(g_{\#}(\omega), g_{\#}(\omega))$, can be described as follows. Consider a hypersurface N in M as the initial wavefront. Its conormal $\eta(x) \in T_x^*(M)$ for every $p = p(x) \in N$, specifies a geodesic through p, the projection of the maximal flow line of $X_{\pi^*(H)}$ through $(x, H(x,\eta(x)), \eta(x))$ in $J^1(M)$. Assuming that for some fixed value of the affine parameter w, the geodesics are still defined for every $p \in N$, we call $N_w = \{\widetilde{\exp}_p(w,\eta) \mid p \in N\}$ the *wavefront* at w , where $\widetilde{\exp} = \exp \circ g_{\#}$

and $\exp_p = T_p(M) \to M$ defined on those vectors in $T_p(M)$ for which the geodesic $\gamma(w)$ (with affine paramter w) with $\gamma(0) = p$ and $\dot{\gamma}(0) \in T_p(M)$ is defined on the interval containing the point one; $\exp(w,X) = \gamma_{wX}(1)$ is then defined in some open interval $(0,w)$, $w > 0$, since $\gamma_{aX}(w) = \gamma_X(aw)$, $a\in\mathbb{R}$, $\dot{\gamma}_X(aw)\big|_{w=0} = aX$, and the fact that $\gamma(w)$ is uniquely determined by $\gamma(0)$ and $\dot{\gamma}(0) = X$. We call the map $N \to N_w$, given by $x \to \widetilde{\exp}_x(w,\eta(x))$ the *geodesic map* at w. In the language of Lagrangian submanifolds introduced in PR. 17, 3.3, the corresponding objects to N and N_w in $T^*(M)$ are Lagrangian submanifolds L_N and L_{N_w} ; L_N is the canonical lifting of the normally oriented submanifold N in M and the projection $\pi = T^*(M) \to M$ onto M defines an embedding of L_N with image manifold $N \subset M$; however the projection of L_{N_w} need not be one and in case it is not, there is no well defined geodesic map from N to N_w. Therefore the image points under the geodesic map of those points in N, at which the geodesic map has rank $< n(n=4)$ are the critical values of the exponential map and they are the same with the points in N_w at which L_{N_w} fails to embed i.e. the critical values of the corresponding Lagrangian map.

To establish further the equivalence between conjugate points to N and critical points of the exponential map, let us assume that there is a focal point to N along $\gamma(w)$, where $\gamma(0) = p\in N$ and $\dot{\gamma}(0)$ is orthogonal to $T(N)$; then $\exp_p\big|_N$ is singular. In this result N need not be a hypersurface; it can be just a submanifold of M. Conjugancy implies the existence of a *Jacobi field* $X(w)$ associated with a *Jacobi variation* $\phi(w,\tau)$, $\tau \in I = (-\tau_o,\tau_o)$, of the geodesic $\gamma(w)$, (M.M. POSTNIKOV, chapter 4, §3 and Appendix) such that :

$$X(w) = \frac{\partial\phi}{\partial\tau}(w.0)$$

which vanishes at a point along $\gamma(w)$, and the variation is such that:

$$\phi(0,\tau) \in N \qquad \tau \in I \qquad\qquad (\phi(0,0) = \gamma(0))$$

$$\frac{\partial\phi}{\partial w}(0,\tau) \text{ is orthogonal to } T_{\phi(0,\tau)}(N) \qquad \left(\frac{\partial\phi}{\partial w}(0,0) = \gamma(0)\right)$$

For every fixed w_o, the Jacobi variation ϕ defines a (smooth) curve $\phi(w_o,\tau)$, $\tau \in I$, and a vector field $\frac{\partial\phi}{\partial w}(w_o,\tau), \tau \in I$, on $\phi(w_o,\tau)$, which uniquely determines a geodesic $\phi_\tau(w)(\phi_\tau(w) = \phi(w,\tau)$, for arbitrary, fixed $\tau \in I)$ of the variation, by definition. For arbitrary, $\tau \in I$, now

$$\phi(w,\tau) = \exp_{\phi(0,\tau)}(w, \frac{\partial\phi}{\partial w}(0,\tau)) \tag{3.50}$$

and differentiating with respect to τ and setting $\tau = 0$

$$X(w) = \frac{\partial\phi}{\partial\tau}(w,0) = \{D(\exp_p)\}^{\mu\nu}\left(\frac{\partial\phi}{\partial\tau}(0,0)\right)_\mu \left(\frac{D}{D\tau}\frac{\partial\phi}{\partial w}(0,0)\right)_\nu$$

But $\frac{D}{D\tau}\frac{\partial\phi}{\partial w}(w,\tau) = \frac{D}{Dw}\frac{\partial\phi}{\partial\tau}(w,\tau)$ (where D denotes the derivative map and $\frac{D}{D\tau}$, $\frac{D}{Dw}$ denote covariant differentiation with repect to τ, w respectively, on $\phi(\tau,w)$) and $\frac{D}{D\tau}\frac{\partial\phi}{\partial w}(0,0) = \frac{D}{Dw}\frac{\partial\phi}{\partial\tau}(0,0) =$

$= \frac{D}{Dw}X(0)$, $X(0) = \frac{\partial\phi}{\partial\tau}(0,0)$, and since the existence of $X(w)$ has been postulated, $X(0)$ and $\frac{D}{Dw}X(0)$ cannot be both zero, $D(\exp_p)$ is singular at p if $X(w)$ vanishes somewhere along $\gamma(w)$. Conversely let us assume that $\exp_p\big|_N$ is singular i.e. there is a non-zero rector $A \in T_p(M)$ in the kernel of $D(\exp_p)$ which is orthogonal to $T_p(N)$,then take a Jacobi variation of the type provided by relation (3.50) where $\frac{\partial\phi}{\partial w}(0,\tau)$ is the integral curve of an arbitrary vector field in $T(M)$ orthogonal to $T(N)$ at every point and equal A over p.

CHAPTER 4

4.1 Introduction

Our aim is to arrive at a classification of some of the objects we introduced in chapter 3; more specifically we endeavour to demonstrate how to arrive at normal forms of germs of certain Lagrange (Legendre) mappings; as a result some properties of topological nature of caustics and wavefronts are established.

The problem of classifying geometrical objects (or forms) becomes much simpler if one tries to classify only stable objects; furthermore stability is a natural condition to place upon forms or processes, understood as a system of forms in evolution, in nature [R. THOM]. Following R. Thom we understand a form in the following context: if E is a topological space and G a group (or pseudogroup) operating on E, then a *G-form* is defined to be an equivalence class of closed sets of E modulo the action of G. Then, a G-form A is called *structurally stable* iff any form B sufficiently close to A in E is G-equivalent to A; there may only be a finite (or, at most, enumerable) number of G-forms with the property of structural stability. We are interested in continuous families of geometrical objects A_s, each object A_s of the family parametrized by a point s of a space of parameters S; if A_s is the object corresponding to a given point s in S, it may happen that, for any point s' sufficiently close to s in S, the corresponding object $A_{s'}$ has the same form as A_s; in this case A_s is called a *structurally stable* object of the family; the set of points s in S for which A_s is structurally stable forms an open subset of S; its complement is called the set of *bifurcation points*; the question, whether the bifurcation set is nowhere dense is what is usually called the problem of structural stability and in most theories the object is to specify its topological structure and its singularities.

In many cases, though not in all, the stable objects themselves are *generic*, that is, they form an open and dense set; so in these cases almost every object is stable and every object is near to a stable one; the non-stable objects are exceptions (*nonforms*); in all cases however, structural stability is a *generic property* (e.g. DFN.30, 2.4). S. Smale has suggested, and has remained standard in the technical literature, that the adjective "generic" should be reserved for properties of a topological space and should never be applied to points of the space, a property being generic if the set of points possessing that property is dense in the space considered. Nevertheless, density of a property is not of itself a justification for assuming the property effectively true of all the points of the space; for example, the set of all irrational numbers between zero and one is dense, but it would be absurd to argue that the properties of irrationals are in some sense universal. Furthermore for any sets $A \subset B$ a topology can trivially be chosen to make membership of A generic. For a consideration of this and related issues in the context of Catastrophe Theory see: T. POSTON "On deducing the presense of catastrophies". Math. Sci. Hum. Vol. 16, no. 64, (1978) 71-99.

We are interested in geometric properties of C^{∞}-differentiable mappings, $C^{\infty}(M,N)$, from a smooth and compact manifold M into a smooth manifold N; the function space $C^{\infty}(M,N)$ (a Banach manifold) may be considered as the parameter space; two mapping f and g, in $C^{\infty}(M,N)$ have the same form or belong to the same topological (differentiable) equivalence class iff there are two homeomorphisms (or diffeomorphisms) μ, ν such that the following diagram:

$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ \mu \downarrow & \circlearrowleft & \downarrow \nu \\ M & \xrightarrow{g} & N \end{array}$$

is commutative.

J. Mather has given a characterization and classification of proper (drop the assumption on M), stable mappings. There are two points of interest in his work; first, the stability of a map is determined by its germ at each point and second, the criteria on the germs of a map which determine its stability are algebraic conditions on the jets of the germs. Both of the above statements with regards to the structural stability of dynamical systems are false (J. GUCKENHEIMER, 1). For smooth real valued functions (germs) the theory is rather trivial; such functions (germ) are stable iff they are Morse functions (i.e. all of their singularities are nondegenerate) with distinct critical values (M. GOLUBITSKY, V. GUILLEMIN, Chapter III, Proposition 2.2).

An answer to our question of stability and hence of classification of caustics has been provided for a variety of concrete and precisely formulated problems; we shall point out some ways in which they differ. In the language of the propagation mechanism described in 3.4, when the affine parameter is taken to be the time t, we define as *caustic points* at t, C_t, those points of the initial wavefront N (an one-codimensional oriented submanifold of M) on which the gerdesic map has rank $<n-1$; the set of all caustic points at all times t in the interval during which the propagation is considered is called the *caustic set* of the propagation. If we add to the assumptions for the existence of the geodesic map, the assumption that the map $\mathbb{R}_+ \times T^*(M) \to M \times M$ defined by $(\pi(\eta), ex\tilde{p}(t,\eta))$ is regular at (t, η), then $(\pi, ex\tilde{p})$ is a local diffeomorphism at (t_o, η_o)

and hence for suitable neighbourhoods U_x of $x_o=\pi(\eta_o)$ and U of $y_o=\widetilde{\exp}\,(t_o,\eta_o)$ we can choose for $(\pi, \widetilde{\exp})$ a local inverse $s:U_x\times U\to \mathbb{R}_+\times T^*(M)$ such that $s(x_o, y_o)=(t_o,\eta_o)$; the function we obtain if we follow s by the projection onto $\mathbb{R}_+$ is called a *geodesic length* function τ, associated to (t_o,η_o), as it gives the parameter length of some geodesic from x to y; note that we have not require (t,η_o) to be a regular point for every t and that for a given Hamiltonian the germ of a geodesic length function at (t_o,η_o) depends only on (t_o,η_o) and not on the choice of s. We then define a function F on N×U by:

$$F=(\tau-t_o)\,|\,N\times U$$

and two sets of points:

$$A\acute{=}\{y\in U:\exists x\in N \ni F(x,y)=t-t_o \text{ and } d_xF(x,y)=0\} \tag{4.1}$$

and

$$B\acute{=}\{y\in U:\exists x\in N \ni F(x,y)=t-t_o,\ d_xF(x,y)=0 \text{ and } d_x^2F(x,y) \text{ degenerate}\} \tag{4.2}$$

For everywhere positive and positively homogeneous of degree one Hamiltonian functions and under certain assumptions of smallness for U_x and U (iff $s(U_x\times U)$ never contains both (t,η) and $(t,-\eta)$) and N and ε, $\varepsilon>0$, (for a given s, $(t_o-\varepsilon,\ t_o+\varepsilon)\times T^*(N)\subset s(U_x\times U)$) we have, $\forall$ $t\in(t_o-\varepsilon,\ t_o+\varepsilon)$, $N_t=A$ and $C_t=B$. For this result see: K. JANICH, "Caustics and Catastrophes", Math. Ann., Vol. 209 (1974) 161-180, Theorem 1. This result establishes rigorously a relationship between the caustics of the propagation and the bifurcation set of a dynamical system (gradient model) see (Appendix III). Structural stability of caustics in the above model is being ensured locally under small perturbations of the initial wave front only: G. WASSERMANN, "Stability of Caustics", Math. Ann., Vol. 216 (1975) 43-50, Theorem 4.

One can actually solve for the successive wavefronts, in a direct sense, by introducing the Hamilton-Jacobi (partial) differential equation, $H(x,df)=0$, where the function $f:M\to\mathbb{R}$ is a solution such that $f|_N=0$; each level surface of f will be the set of points at a given geodesic distance from the initial wavefront N and will be one of the surfaces N_w; however the caustic points are the points where one cannot find a smooth solution with the given initial data. By PR. 19, 3.3, to find local solutions f of H=0 is equivalent to finding pieces of Lagrangian submanifolds transversal to the fibers on which H=0. A singularity of a solution is a point in M such that the corresponding Lagrangian submanifold is not transversal to the fiber of $\pi:T^*(M)\to M$ over the point in question. One would like to describe the local structure of the singularities of a generic set of solutions; in this context we mention a few facts from the work of J. GUCKENHEIMER, "Catastrophes and partial differential equations", Ann. Inst. Fourier (Grenoble), Vol. 23, No. 2 (1973), 31-59, and "Caustics and non-degenerate Hamiltonians", Topology, Vol, 13, (1974) 127-133. First, by introducing a parametrization of all Lagrangian submanifolds (planes) of $\mathbb{R}^{2n}$, we may identify the set of germs of Lagrangian submanifolds through a given point with a given tangent plane with the cube of the maximal ideal of the ring of germs of functions on $\mathbb{R}^n$. Let $T^*(\mathbb{R}^n)$ be identified with $\mathbb{C}^n$ via the map $(x,u)\to(x+iu)$ and let us define a hermitian scalar product in $\mathbb{C}^n$ by $\sum_{i=1}^{n}(x_i+iu_i)\cdot(x'_j-iu'_j)$ so that its real part $\sum_{i=1}^{n}(x_ix'_i+u_iu'_i)$ is the Euclidean scalar product in $\mathbb{R}^{2n}$ and its imaginary part $\sum_{i=1}^{n}(u_ix'_i-x_iu'_i)$ the standard symplectic form; let $L(n)$ be the set of all n-dimensional subspaces L in $\mathbb{C}$ on which the imaginary part of the hermitian scalar product vanishes indentically; this means that L and iL are orthogonal with respect to the Euclidean scalar product; the unitary group U(n) acts transitively on $L(n)$ and its isotropy subgroup at $\mathbb{R}^n$ (i.e. $U\mathbb{R}^n=\mathbb{R}^n$) is the orthogonal group O(n); we can therefore identify $L(n)$ with $U(n)/O(n)$ (Appendix III). An element $A+iB$ of $GL(\mathbb{C}^n)$ corresponds to an element $\begin{pmatrix} A & -B \\ B & A \end{pmatrix}$ of $GL(\mathbb{R}^{2n})$ and it is transversal to $\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ (PR.15, 3.3) if A is nonsingular; then $A^{-1}\cdot B$ is symmetric and the Lagrangian plane determined by $\begin{pmatrix} A & -B \\ B & A \end{pmatrix}$ is the span of the first n-column vectors or the span of $\begin{bmatrix} I \\ A^{-1}\cdot B \end{bmatrix}$ or the graph df, where $f\in C^\infty(\mathbb{R}^n)$ is the quadratic function determined by $A^{-1}\cdot B$; if graph df passes through the origin and has a horizontal tangent plane there, $f(o)=df(o)=d^2f(o)=o$ and $f\in m^3(n)$. To describe the set of germs of solutions of the first-order partial differential equation $H(x,u)=o$, choose coordinates so that $H(o,o)=o$ and let (o,o) be a regular point of H so that one can choose local canonical coordinates such that H becomes a coordinate function, i.e. $H(x,u)=x_n$ or $H(x,u)=u_n$; in the later case the solutions are functions f which do not depend upon x_n and the Lagrangian submanifolds consist of (n-1)-dimensional families of lines parallel to the x_n-axis, i.e. $m^3(n-1)$, which can be identified with germs of Lagrangian manifolds of $T^*(\mathbb{R}^{n-1})$; the non invariant choice of coordinates with respect to Lagrangian equivalences (DFN.13, 3.3) nonwithstanding, this is a nongeneric (in n-dimension) situation; the solutions of $H(x,u)=x_n$ are problematic too in that they are not transversal to the fiber of $\pi:T^*(\mathbb{R}^n)\to\mathbb{R}^n$ at any of its points; this is always the case for (positively) homogeneous (in the fibre coordinates) Hamiltonian functions which give rise to *conic Lagrangian submanifolds;* to define their singularities one cannot utilise their transversality to the fibers of $T^*(M)$ as they are nowhere so, and their singularities are the points at which $\pi\cdot i_1$ (relations 3.50) fails to be a local diffeomorphism. Second, one can construct a surjective map, σ, from a certain family of functions parametrized by M (Appendix III) into the germs of Lagrangian submanifolds of $T^*(M)$; σ corresponds to λ, of our construction below, and $\Sigma(F)$ to L_F (DFN.1, 4.3); J. Guckenheimer introduces three types of equivalence, and hence of stability, for germs of Lagrangian submanifolds, refered to as type I, II and III by using diffeomorphisms of the base M which map the critical values of the corresponding Lagrange mappings into each other, fiber preserving local diffeomorphisms of the bundle space T*(M) which map the Lagrangian submanifolds into each other and equivalent families of functions respectively. He then proves that the germ of the set of degenerate critical points of a family of functions in question occurs as the germ of the set of the critical values of the Lagrange mapping of the corresponding Lagrangian manifold if it is III-stable, and indicates how to prove that these sets describe the singularities of almost all solutions of H=o for certain H's; the conditions on the Hamiltonian functions are rather general: zero is not a critical point of H, its zero level is transverse to the fibers of T*(M), and its Hessian with respect to the fiber coordinates is not degenerate. Genericity of solutions is determined in terms of transversality to the stratification induced by the equivalence relation in question; this is possible utilizing a suitable version of Thom's transversality theorem (Appendix III). A certain relation to the stability of gradient vector fields is also treated by considering the family of functions determined up to a constant by the family of gradient vector fields, but no direct connection can be found to the model of the geodesic length function treated by K. Janich. However in the second of the above mentioned papers by J. Guckenheimer these results are further strengthened by proving II-stability for generic solutions of H through a given point, and the technique used is in a sense a generalisation of the initial wavefront method in that it constructs a total (n-parameter) solution without the homogeneity and positive definiteness assumptions on H, which can be interpreted as a wavefront propagation in a non-Riemannian setting.

The study of Lagrangian submanifolds and the study of the critical values of their corresponding Lagrange mappings was originated however in a slightly different context, that of integrals of rapidly oscillating functions and the theory of (pseudo-) differential operators. Studying (L. HORMANDER) general integral operators(A) of the following type

$$Af(x)=\int e^{iF(x,y,u)}G(x,y,u)f(y)dydu$$

the only contributions to the integral which are not rapidly decreasing come from the set of points where F is stationary as a function of u, i.e. $L_F \stackrel{\prime}{=} \{(x,u)\in\mathbb{R}^n\times\mathbb{R}^k:\ d_uF=o\}$; under certain conditions of regularity for F, i.e. $\operatorname{rank}(d_{(x,u)}d_uF)=k$, the map $\lambda:(x,u)\to(x,d_xF)$ has an injective differential and therefore defines locally an embedding of L_F in $T^*(\mathbb{R}^n)$ whose image turns out to be a Lagrangian submanifold (J.J. DUISTERMAAT, relations (1.1.12) and (1.2.9)). The asymptotic character of such integrals is thus related to the singularities of the corresponding Lagrange mapping; Lagrange mappings have more singularities than general mappings from an n-dimensional manifold into $\mathbb{R}^n$; a classification was originally given by V.I. Arnol'd (V.I. ARNOL'D 1,3) and a proof was suggested which was subsequently carried out in a somewhat sketchy fashion together with an extended classification by V.M. Zakalyukin (V.M. ZAKALYUKIN). The part of Arnold's paper (V.I. ARNOL'D 1) dealing with the classification of single singularities of smooth functions has been worked out by D. Rand (D. RAND). In the context of the study of oscillatory integrals, J.J. Duistermaat in his review article (J.J. DUISTERMAAT) develops a theoretical framework for a global study of Lagrangian submanifolds (Lagrange immersions i:

$L \rightarrow T^*(M)$); in the first instance generalizes L. Hormander's construction of conic Lagrangian submanifolds via a generating family of functions (the phase function F) homogenerus of degree one in the fiber variable (u) dropping the assumption of homogeneity; secondly in considering $L_F(x) \doteq \{u \in \mathbb{R}^k = (x,u) \in L_F\}$ he assumes the most general case of it containing more than one element i.e. assumes the function F(u,x) having both degenerate critical points as well as multiple critical values with respect to u. However the second case corresponds to shock phenomena and for our purpose one need not employ the rather cumbersome notation for multi jets.

Finally a note on the topology of the function (map) spaces we will be using; the Lagrangian (Legendrian) submanifolds which are of interest to us (PR.17 and PR.18, 3.3) are invariant under the action of the Hamiltonian X_H (contact $X_{\pi^*(H)}$) vector field (relations 3.45 to 3.47) which implies that close Lagrangian submanifolds can be obtained from each other along the solution curves of $X_H(X_{\pi^*(H)})$; as a result one can choose a neighbourhood in the space of inclusions of Lagrangian (Legendrian) manifolds with the Whitney C^∞-topology consisting of only one element; the next less fine topology of the same nature is the Weak C^∞-topology (Appendix III).

There has been an application to first-order WKB approximation of solutions of Einstein's equations near a conjugate point to a fixed spacelike surface, using certain normal forms for the phase function by Y. Manor: "Caustics in general relativity II. The WKB approximation", J. Phys. A, Vol. 10, No. 5, (1977) 765-76, but the classification itself (unpublished) based on the techniques developed by F. W. Warner: "The conjugate locus of a Remannian manifold" Am. J. Math, Vol. 87 (1965) 575-604, retrieves only one simple, stable form for the phase function.

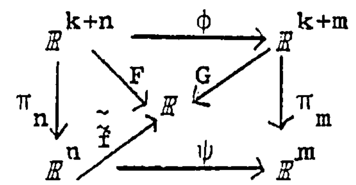
4.2 Versal deformations of functions

The remarks in the introductory section related to the definition of wavefronts and the caustic sets via the singularity (A) and the bifurcation (B) sets respectively of a local gradient model (Appendix III), together with relation, (3.48) and (3.50) defining locally a Lagrangian (Legendrian) submanifold and its Lagrange (Legendre) map, suggest that we will be interested in n-parameter families of germs of singularities, that is, germs at the origin of functions $F(u,x)\epsilon C^\infty(\mathbb{R}^{k+n})$ such that $F|_{\mathbb{R}^k\times\{0\}}=f$, where f has a germ in $m^2(k)$, the second power of the maximal ideal m(k) in the ring of differentiable germs $E(k)$ of functions on $\mathbb{R}^k$ at zero; such families are referred to as *deformations* or *unfoldings* of f. Let $E(n)$ denote the ring of germs of differentiable functions $C^\infty(\mathbb{R}^n)$ at zero and m(n) its maximal ideal (Appendix III); no misunderstanding will arise from the use of the same symbol in 3.2 for the maximal ideal of differentiable functions on a general manifold which vanish at a given point. Next we define morphisms between unfoldings; they are permitted to transform arbitrarily the parameter space $\mathbb{R}^n$ and the fibres of the fibration $\pi_n:\mathbb{R}^{k+n}\to\mathbb{R}^n, \forall x\epsilon\mathbb{R}^n$ other than the origin; in the range we only allow translations, the simplest transformations which allow us to ignore the constant terms of the restrictions of F to the fibres, as we are only being interested in germs of functions; hence we have the following definition;

DFN.1 Let $f\epsilon m(k)$, $F\epsilon E(k+n)$ and $F'\epsilon E(k+m)$ unfoldings of f. A *(right) morphism* from F to F' is a pair of map-germs (ϕ,ψ), $\phi\epsilon E(k+n, k+m)$ ψ $E(n, m)$, where $\phi|_{\mathbb{R}^k\times\{0\}}=id_{\mathbb{R}^k}$, and $\tilde{f}$ m(n) such that $F=F'\cdot\phi+\tilde{f}\cdot\pi_n$, $\pi_n\cdot\phi=\psi\cdot\pi_m$

We say that the unfolding (n,F) of f is induced by $(\phi,\tilde{f})$ from the unfolding (m,F').

Diagramatically:



The condition $\phi|_{\mathbb{R}^k\times\{0\}} = id_{\mathbb{R}^k}$ in fact implies that ϕ is a local diffeomorphism for those values of $u\in\mathbb{R}^k$ sufficiently close to zero; a morphism is an isomorphism exactly when ϕ is invertible; that is iff $n=m$ and ψ is the germ of a local diffeomorphism.

DFN.2 A deformation (n,F) of f is called versal iff any other deformation of f is induced from (n,F) by a suitable morphism; a versal deformation with minimal n is called *universal (miniversal)*.

The significance of the concept of versality consists in its being equivalent to stability. Stability is defined by requiring that for every sufficiently small perturbation F' of a representative of a germ of a deformation F, F and F' are G-equivalent in the sense of 4.1 ; but for a function F', defined in a neighbourhood of the origin in $\mathbb{R}^{k+n}$, close to F, where F is a representative of a germ of a singularity f, the germ of $F'|_{\mathbb{R}^k\times\{0\}}$ will not be equal to f, in fact it will not even be equivalent to f in the sense of mapping equivalence; however it is reasonable to expect that for F' sufficiently close to F, the germ at (u,x) of $F'|_{\mathbb{R}^k\times\{x\}}$ will be equivalent to the germ of f, considered as a germ in $m(k)$, that is, being made into a germ at the origin by prefixing a translation of $\mathbb{R}^{k+n}$ and then made into an element of $m(k)$ by ignoring the constant term. Consequently the definition of stability although modelled on the definition, DFN.1,4.2, of an (iso-) morphism, must involve functions defined on a neighbourhood of the origin as well as germs; the diffeomorphisms considered will always respect the fibration $\mathbb{R}^k \times \mathbb{R}^n$.

DFN.3 Let $F_i \in C^\infty(V_i, \mathbb{R})$, V_i open in $\mathbb{R}^{k+n}$ $i=1,2$; we shall say that F_1 at $(u_o, x_o) \in V_i$ is equivalent to F_2 at (u_o', x_o') if there exist some open nbd.s U_1 of u_o in $\mathbb{R}^k$ and U_2 of x_o in $\mathbb{R}^n$ such that $U_1 \times U_2 \subseteq V_1$ and there exist smooth functions $\tilde{\Phi}, \tilde{\Psi}$: $U_2 \to \mathbb{R}^n$ and $\tilde{f}: U_2 \to \mathbb{R}$ satisfying the following:

(i) $\tilde{\Phi}(u_o, x_o) = x_o'$ and $\tilde{\psi}(x_o) = x_o'$

(ii) $(\tilde{\Phi}(u,x), \tilde{\psi}(x)) \in V_2$, $u \in U_1$ and $x_\epsilon U_2$

(iii) $\tilde{\Phi}|_{U_1 \times \{x\}}$ is non-singular at (u_o, x_o), $\tilde{\psi}$ is non-singular at x_o and $\tilde{f}$ non-singular at x_o

(iv) $F_1 = F_2 \cdot \phi + f$, i.e. $F_1(u,x) = F_2(\tilde{\Phi}(u,x), \tilde{\psi}(x)) + \tilde{f}(x)$ $u \in U_2$ and $x \in U_2$, where $(\tilde{\Psi}, \tilde{\psi}) = \phi$

DFN.4 We shall say a deformation is stable iff for every open nbd. U of the origin in $\mathbb{R}^{k+n}$ and every representative F of the deformation, defined on U, there is a nbd. V of F' in $\mathbb{R}C^\infty(U, \mathbb{R})$, with the Weak C^∞-topology , such that for every $F' \in V$, there is a point $(u,x) \in V \subset \mathbb{R}^{k+n}$ such that F' at (u,x) is equivalent, DFN.2, 4.2, to F' at (o,o).

This is one of many definitions of stability given the definition of equivalence and the choice of topology, and the idea is to prove that stability does depend only on the equivalence class of the deformation; there is a definition of stability (infinitesimal stability) given by an algebraic condition which can be explicitly checked out, which, one can prove, is equivalent to versality for a certain class of singularities. This is what we will sketchily demonstrate in the rest of this section; to this purpose we need a few more technicalities.

Let G(k) be the group of germs of local diffeomorphisms of $\mathbb{R}^k$ at the origin; an element of G(k) induces an $\mathbb{R}$-algebra homomorphism on E(k) by action on the right. Let $G^r(k)$ denote the space of the r-jets, 3.2, of the elements of G(k); $G^r(k)$ is canomically isomorphic to the group of $\mathbb{R}$-algebra homomorphism of $J^r \equiv E(k)/m^{r+1}(k)$. Let $f \in E(k)$, then fG denotes the G-orbit of f, and fG^r the G^r-orbit of $j^r(f)$; G^r-orbits are embedded submanifolds of J^r since the action of $G^r(k)$ is smooth and algebraic. In general, for the right, smooth action of a separable finite dimensional Lie group G acting on a smooth manifold M we define:

DFN.5 A G-orbit is said to be *simple* if there exists a sufficiently small neighbourhood of any of its points which only meets a finite number of G-orbits.

If the group and action are algebraic, which is the case for the action of G^r on J^r, then every small neighbourhood of an orbit meets either a finite number of distinct orbits, or a continuous family of orbits; the parameters are called *moduli*.

DFN.6 A singularity is called *simple* iff there exists positive integers K,N such that the G^r-orbit of its r-jet is simple for all r>K and the number of *abutting* orbits $\rho(f,r) \leq N$, i.e. remains bounded as $x \to \infty$

An orbit A abuts an orbit B iff every neighbourhood of B meets A; $\rho(f,r)$ equals the number of G^r-orbits which meet a neighbourhood U of $j^r(f)$.

DFN.7 A singularity $f \in E(k)$ is said to be *r-determined* if its r-jet $j^r(f)$, is sufficient; an element ξ of J^r is said to be *sufficient* if $\xi = j^r(f) = j^r(f')$, $f' \in E(k)$, implies that f and f' are G(k)-equivalent, i.e. any two germs with ξ as r-jet are right equivalent.

Let $J(f)_{E(k)}$ be the *Jacobian ideal* of f, that is the E(k)-module generated by the *local ring* of f, i.e. $\left[\frac{df}{du_i}\ i=1,\dots,k\right] \subset E(k)$; we then define the dimension of the quotient $m(k)/J(f)_{E(k)}$ as an $\mathbb{R}$-vector space, as the *codimension* of f.

All the above concepts are related in the sense that simplicity implies determinary which in terms is equivalent to finiteness; first we need a criterion for determinary and the following necessary condition for right equivalence of jets

PR.1 Let $f \in E(k)$ and $j_r(f)$; then:

$$T_{j^r(f)}(fG^r) = m(k)\cdot J(f)_{E(k)} \bmod m^{r+1}(k)$$

where we identify the tangent spaces of J^r with J^r

PROOF: (D. RAND, Remark 9.3(1) and Lemma 9.4 and TH. BRÖCKER, chapter 11, Lemma 11.8)

PR.2 A germ $f \in E(k)$; such that $m^r(k) \subset m(k).J(f)_{E(k)}$, is r-determined. Conversely if $f \in E(k)$ is r-determined, then $m^{r+1}(k) \subset m(k).J(f)_{E(k)}$

PROOF: (TH. BROCKER, chapter 11, theorem 11.3 and Corollary 11.10)

Given the series of implications: $m^r \subset J \Rightarrow m^r \subset m\cdot J \Rightarrow m^{r+1} \subset m^2, m^{r+1} \subset m\cdot J$ $m^{r+1} \subset J$, this is the strongest result one can actually prove; for a proof of the suffiency when $m^{r+1} \subset m^2\cdot J$ see also D. RAND (D. RAND, Theorem 8.4 and Corollary 8.5) and E.C. Zeeman, "Classification of singularities of codim ≤5", Warwick University preprint; J.N. Mather in: "Stability of C∞ mappings, III: Finitely determined map-germs", I.H.E.S. Vol.35 (1968) 279-308, deals with the general case of map-germs and right-left equivalence.

The inclusion of relations in PR.1, 4.2, are between $E(k)$-modules; by taking the quotient with submodule $m_{(k)}^{r+1}$ we transform them into conditions on the r-jet of f, e.g. $m^r \subset m\cdot J+m^{r+1}$; such a condition written as $m^r \subset m\cdot J+m\cdot m^r$ implies by Nakayama Lemma (Appendix III) $m^r \subset m\cdot J$, i.e. our hypothesis. Our hypothesis is also equivalent to $E(k)/_{m(k)\cdot J(f)E(k)}$ being finite dimensional and as a result, generated by monomials of degree less than r as an $E(k)$-module. Indeed, let $A=E(k)/_{m(k)\cdot J(f)E(k)}$; then $m\cdot A \subsetneq A$ (contained but not equal) since by Nakayama lemma $A=m\cdot A$ would imply $A=o$; for every finite dimensional (r), vector space $V \supsetneq V_1 \supsetneq V_2 \ldots \supsetneq V_\ell$, V_i proper subspaces, there exist ℓ such that $V_\ell=o$ and $\ell \le r$; hence $0=m^\ell\cdot A \subsetneq m^{\ell-1}\cdot A \subsetneq \ldots \subsetneq m\cdot A \subsetneq A$ $\ell \le r$ (r=dim A); note that $\ell \ge 1$ in this case since $m^{\ell-1}\cdot A \subsetneq$); but $m\cdot J\cdot A=0_A$ implies $\alpha\cdot A=o \Rightarrow \alpha \in m\cdot J$, hence $m^\ell\cdot A=o \Rightarrow m^\ell \subset m\cdot J$ and since $\ell \le r$, $m^r \subset m\cdot J$. Furthermore $E/m\cdot J$ finite dimensional is equivalent to E/J being finite dimensional (Appendix III).

To say that f is r-determined is much more a statement about the r-jet, $j^r(f)$, of f than one about f; $j^r(f)$ is called *s-determined*, for some $s \le r$ if for every $\xi \in J^r$ such that $\pi_s^r(j^r(f))=\pi_s^r(\xi)$, $j^r(f)$ and ξ are G^r-equivalent. One can see that if $f \in m(k)$ is s-determined, its r-jet, $j^r(f)$, $s \le r$ is s-determined and conversely (although it is hard to prove it) if r is big enough ($r \ge s+1$); in general r-determinary of jets is weaker than r-determinary of germs which represent them. The non-determined jets become more scarce as the power increases. To be more precise let

$J_o^r = \{\xi \in J^r : \pi_o^r(\xi)=o\}$ be the set of r-jets at zero of germs in $E(k)$, whose zero-jet (value), is zero; J_o^r is a linear subspace of J^r of codimension one; for every $\xi \in J_o^r$ define $\tau(\xi)=\dim_{\mathbb{R}} E(k)/J(f)_{E(n)}$, where f is any germ in m(k) such that $j^r(f)=\xi$ and let $A_r=\{\xi \in J_o^r \mid \tau(\xi)\ge r\}$; if $\xi \notin A_r$, then ξ is r-determining in the sense that there exist $f \in m(k)$ such that $j^r(f)=\xi$ and f is r-determined;

conversely if $f\in m(k)$ is finitely determined, then for some r $j^r(f)\notin A_r$; A_r is an algebraic subset of the Euclidean space J_o^r and $\pi_s^r(A_s)\subseteq A_r$ $r\geq s$. For example, no zero-jet (values) determines its germ, the non-determined one-jets form the line $\mathbb{R}\times\{o\}$ in $\mathbb{R}^{n+1}=J^1(\mathbb{R}^n)$ and the non-determined two-jets form a thin set above a point on the line of non-determined one-jets.

PR.3 If $f\in m^2_{(k)}$ is simple, then f is finitely determined

PROOF (D. RAND, Proposition 13.1)

Suppose $f\in m^2_{(k)}$ is not finitely determined; by PR.2, 4.2, for every $r\geq 2$ $j^r(m^r)\not\subset j^r(m\cdot J_f)$ (i.e. $m^r(k)\not\subset m(k)\cdot J(f)_{E(k)}$ mod $m^{r+1}_{(k)}$ in full notation); thus for every $r\geq 2$ choose a homogeneous polynomial $\xi_r\in m^r/_{m^{r+1}}$ which does not lie in $j^r(m\cdot J_f)$; the map $w\to j^r(f)+w\xi_r$ is transverse to $j^r(m\cdot J_f)$, which equals $T_{j^r(f)}(fG^r)$ by PR.1, 4.2, and is thus transverse to the orbit fG^r. Hence for arbitrary small w_r, $(f+w_r\xi_r)G^r$ is distinct from fG^r; now the G^s orbits of $j^s(f+w_r\xi_r)$, $r=1,\ldots,s$ and $j^s(f)$ are all distinct and hence $\rho(f\ ,s)\geq r$, i.e. f is not simple by DFN.6, 4.2.

PR.4 Let $f\in m^2(k)$; f is finitely determined iff codim f is finite

PROOF: Let docim f $<\infty$; the sequence $m/_J\supset m^2+J/_J\supset\ldots$ must terminate and hence $m^r+J=J$ for some $r\geq 2$, i.e. $m^r\subset J$, and by PR.2, 4.2, f is (r+1)-determined.

Conversely by PR.2, 4.2, if f is finitely determined $m^r\subset J$ for some $r\geq 2$ and codim f $=\dim_{\mathbb{R}} m/_{m^r}<\infty$.

The following result relates the concept of codimension of a singularity to the codimensions of its G^r-orbits in the r-jet space; to this purpose one needs the following

PR.5 Let $f\in m^2(k)$ and codim $f<\infty$; then $\dim_{\mathbb{R}} J(f)_{E(k)}/_{m(k)\cdot J(f)_{E(k)}}$ equals n.

PROOF: (TH. BROCKER, 14.15 and D. RAND, Lemma 10.6)

If f is not finitely determined, then PR.5, 4.2, is false; for example $f=x^2$ and n=2, J=(x,o), $m\cdot J=(x^2,xy)$ and dimension of $^J/_{m\cdot J}=1\neq 2=n$.

PR.6 Let $f\in m^2(k)$ and c=codim f; then fG^r, $r\geq c+2$, is a submanifold of $m^2(k)/_{m^{r+1}(k)}$ of codimension c.

PROOF: By PR.1, 4.2, $T_{j^r(f)}(fG^r)=(m(k)J(f)_{E(k)})$ mod $m^{r+1}_{(k)}$; from the proof of PR.4, 4.2, it follows that if r is the least positive integer such that f is r-determined, then $r\leq$ codimf+2, hence by our assumption f is r-determined and by, PR.2, 4.2, $m^r\subset m\cdot J$. The codimension of fG^r in m^2/m^{r+1} is therefore equal to:

$$\dim_{\mathbb{R}} {}^{m^2}/_{m^{r+1}}-\dim_{\mathbb{R}} {}^{m\cdot J}/_{m^{r+1}}$$

$$= \dim_{\mathbb{R}}({}^{m^2}/_{m^{r+1}})/(mJ/m^{r+1})$$

$$= \dim_{\mathbb{R}} {}^{m^2}/_{mJ}$$

$$= \dim_{\mathbb{R}} {}^{m}/_{mJ}-\dim_{\mathbb{R}} {}^{m}/_{m^2} \text{ since } {}^{m}/_{mJ}={}^{m}/_{m^2}\oplus{}^{m^2}/_{mJ}$$

$$= \dim_{\mathbb{R}} {}^{m}/_{J}+\dim_{\mathbb{R}} {}^{J}/_{mJ}-\dim_{\mathbb{R}} {}^{m}/_{m^2} \text{ since } {}^{m}/_{mJ}={}^{m}/_{J}\oplus{}^{J}/_{mJ}$$

$$= c+n-n \qquad \text{by PR.5, 4.2}$$

One can further utilize this result in proving that for finitely determined singularities versality is equivalent to transversality (Appendix III) in the r-jet space of the orbits of G^r (the r-jets of the group of germs of local diffeomorphism of the parameter space of a deformation of a singularity, acting on the right).

Let $J^r_o=\pi_r(m(k))$, $\pi_r:E(k)\to J^r$ $(J^r={}^{E(k)}/_{m(k)^{r+1}})$, denote the linear subspace of J^k of codimension one of r-jets of germs of functions at zero, whose zero-jet (value) is zero, there exist a natural projection $P_r:J^r\to J^r_o$ defined by $P_r(j^r(f)(o))=j^r(f-f(o))(o)$

(swift to the origin) and a $\mathbb{R}$-linear map $i^*: J^r(k+n) \to J^r(k)$, induced by the canonical linear embedding $i = \mathbb{R}^k \to \mathbb{R}^{k+n}$ $(u \to i(u) = (u,o))$; define $j_o^r(f) = i^* \cdot P_r \cdot j^r(f)$; $j_o^r(f)$ sends (u,x) to the restriction on $\mathbb{R}^k \times \{o\}$ of the r-jet of $v \to (f(u+v,x) - f(u,x))$ at $v=o$; for a deformation of F of $f \in m^2(k)$ $(F|_{\mathbb{R}^k \times \{o\}} = f)$ $j_o^r(F)(o) = j^r(f)(o)$.

DFN.8 A deformation (n,F) of $f \in m^2(k)$ is said to be r-transversal iff $j_o^r(F)$ is transversal at the origin to the orbit fG^r of $j^r(f)$ in J^r.

The importance of transversality lies in the following result:

PR.7 If f is r-determined and (n,F) and (n,F') are r-transversal deformations of f, then (n,F) and (n,F') are isomorphic.

PROOF: (TH. BROCKER 16.3 and G. WASSERMANN, 3.16)

The proof is hard and the methods used are attributed to J. Mather.

There is an explicit (algebraic) criterion for transversality:

PR.9 A deformation (n,F) of $f \in m^2(k)$ is r-transversal iff

$$E(k) = J(f)_{E(k)} + V_F + m(k)^{r+1} \qquad (4.3)$$

where V_F is the linear subspace of $E(k)$ generated over $\mathbb{R}$ by the multiplicative identity of $E(k)$ and the elements of $J(F)_{\mathbb{R}}$, i.e. $\frac{\partial F}{\partial x_i}\Big|_{\mathbb{R}^k \times \{o\}} i=1,\ldots,n$

PROOF: (TH. BROCKER, 16.4 and G. WASSERMANN, 3.13)

$E(k) = m(r) \oplus \mathbb{R}$, where $f = (f - f(o)) \oplus f(o)$ and as a result $J_{\mathbb{R}} + m \cdot J_E = J_E$; the condition is therefore equivalent to $m(k) = J(f)_{E(k)} + J(F)_{\mathbb{R}} + m(k)^{r+1}$ where $J(F)_{\mathbb{R}}$ is spanned by $\left(\frac{\partial F}{\partial x_i}\Big|_{\mathbb{R}^k \times \{o\}} - \frac{\partial F}{\partial x_i}(o)\right)$, $i=1,\ldots n$

One can then show, using PR.9, 4.2, rather straightforwardly that if (n,F) is a versal deformation, then F is r-transversal for every r. Conversely transversality implies determinary and PR.7, 4.2 ensures versality. Moreover, for a finitely determined singularity (if $f\in m(k)$ and $f\notin m^2(k)$ then f is a universal deformation of itself) the minimal dimension of a universal deformation of f equals the codimension of f. Therefore, if $\{f_i(u)\}i=1,\ldots,n$ is a system of representatives for a basis of $m(k)/_{J(f)_{E(k)}}$, then the deformation F(u,x) of f(u), defined by

$$F(u,x)=f(u)+\sum_{j=1}^{n}x_j f_j(u) \qquad (4.4)$$

is universal.

Finally we introduce the concept of *infinitesimal stability*; to this purpose it is easier to consider equivalences of deformations under the combined action of the product group of germs of diffeomorphisms of the argument space of the parameters acting on the right with the general group of germs of diffeomorphisms of $\mathbb{R}$ acting on the left, instead of allowing only translations; this complicates things slightly but since a right morphism (DFN.1, 4.2) is a right isomorphism if and only if it is a right-left isomorphism the content of most of what has been said and especially PR.2 and PR.7 4.2, will remain the same but for suitable changes in the order of jets involved; we shall refer to all extended concepts as concepts "for levels"; a deformation (n,F) of a singularity $f\in m^2(k)$ is called *r-transversal for levels* iff $j_o^r(F)$ is transversal at the origin to the orbit $G^r(1)fG^r(k)$ of $j^r(f)$ in J^r (DFN.8, 4.2) under the action of $G^r(k)$ on the right and $G^r(1)$ on the left. Let f* be the induced algebra homomorphism $E(1)\to E(k)$ by action on the right; f* makes $E(k)$ into a module over $E(1)$; then (note that $J(F)_{\mathbb{R}}+f^*E(1)=V_F+f^*E(1)$ comparing to PR.9, 4.2)

PR.10 A deformation (n,F) of $f\in m^2(k)$ is r-transversal for levels iff

$$E(k)=J(f)_{E(k)}+J(F)_{\mathbb{R}+}+f^*E(1)+m^{r+1}(k) \qquad (4.5)$$

In view of PR.7, 4.2, for levels (G. WASSERMANN, 3.16), the relation (4.5) is also a condition for versality for levels; we will show that it is equivalent to infinitesimal stability; it is also referred to in the literature as *infinitesimal versality* (for levels). The idea of infinitesimal stability is that the notion of stability (DFN.4, 4.2) for levels, should hold infinitesimally at F, i.e. $\frac{d\tilde{F}}{dt}(u,x,o)=0, \forall (u,x)\in U\subset\mathbb{R}^{k+n}$, where $\tilde{F}(u,x,t)=\Phi_t(F_t(\phi_t(u,x),\psi(x)),x)$ $\tilde{F}(u,x)=(F(u,x),x)$, and ϕ_t, ψ_t, Φ_t, are germs of smooth paths; differentiating we have:

$$E(k+n)=J(F)_{E(k+n)}+J(F)_{E(n)}+F^*E(n+1) \qquad (4.6)$$

where $E(n)$ is considered as a subring of $E(k+n)$.

By applying $i^*:E(k+n)\to E(k)$ we have the condition in PR.10, 4.2. The proof of the converse is more involved and one needs the Malgrange Preparation theorem in the form of J. Mather (Appendix III). What PR.9 and PR.10, 4.2, mean is that for any representative g of a germ in m(k) there exists a decomposition

$$g=(\Phi\cdot f)+\sum_{j=1}^{k}\frac{\partial f}{\partial u_j}\cdot\psi_j+\sum_{i=1}^{n}\frac{\partial F}{\partial x_i}\Big|_{x=o}\cdot r_i \qquad (4.7)$$

if F, as a deformation of f, is versal (for levels), where $r_i\in\mathbb{R}$, and Φ, $\psi_j\in m(k)$.

Next comes the main result about deformations

PR.11 Let (n,F) be a deformation of $f\in m(k)$; F is infinitesimally stable iff F is stable

PROOF: G. WASSERMANN, 4.11 and M.F. LATOUR, Théorème)

G. Wassermann proves an equivalence of a whole variety of stability definitions, which seem to be relevant to the various applications of catastrophe theory; the important result however is that of their equivalence to infinitesimal stability. M.F. Latour's proof dealing with families of maps rather than simply families of functions, rest entirely on the work (and notation) of J. Mather. A readable and short proof of the fact that infinitesimal versality implies versality has also been given in: V.M. Zakalyukin, "The Versality theorem", Funct. Anal. Appl., vol.7, no.2 (1972) 110-112.

There are simple (polynomial) normal forms for functions near a critical point; for nondegenerate and isolated critical points they are Morse functions; for degeneracies of simple germs one can achieve an analogous classification (i.e. discrete); there exist classifications even when moduli are present. From the point of view of classification of degenerate singularities (i.e. germs of functions at zero, with critical value 0 and Hessians of not maximal rank) two facts will be of interest to us; first, that the problem of classifying the simple singularities (DFN.6, 4.2) is reduced to the problem of classifying the simple singularities in two variables; second, that the sets of all nonsimple germs of singularities of various types or degrees of degeneracy have large codimensions (in $\mathcal{E}$) (some infinite and certainly not less than six). Relations (4.4) allow us to compute normal forms for deformations. We will be interested in deformations of codimension four, because our families of functions are being parametrized by the points of a spacetime manifold. We state in the form of a proposition the relevant result from the classification theory.

PR.12 The following are the normal forms of simple functions of codimension ≤4 in the neighbourhood of a critical point (and their universal deformations):

Series A_k $:\pm u_1^{k+1} \pm u_2^2 + Q(X_{k-i}u_1^{k-1} + X_{k-2}u_1^{k-2} + \ldots + x_1 u_1 + x_o)$ $1 \le k \le 5$

Series D_k $:\pm u_1^2 u_2 \pm u_2^{k-1} + Q(x_{k-1}u_1 + x_{k-2}u_2^{k-2} + \ldots + x_1 u_2 + x_o)$ $4 \le k \le 5$

where Q denotes a standard quadratic form in the remaining u- variables

PROOF: (V.I. ARNOL'D, 1, Lemmas 4.3, 5.3, 5.4 and Corollary 8.4 and D. RAND Propositions 15.1 and 16.7)

4.3 A necessary and sufficient condition for stability of germs of Lagrangian manifolds

We start with two statements about the action of Lagrangian equivalences (DFN.13, 3.3) on the corresponding Lagrangian manifolds; we are actually talking about Lagrangian equivalence of germs of Lagrangian submanifolds (PR.13, PR.14 and DFN.15, 3.3).

PR.13 A Lagrangian equivalence Φ of the standard fibration $\pi: T^*(\mathbb{R}^n) \to \mathbb{R}^n$ is uniquely determined by a pair (ψ, f), where $f \in \mathcal{E}(n)$ is such that $\phi^*(\omega) = f \cdot \omega$, $\omega \in \Lambda^1(T^*(\mathbb{R}^n))$, and $\psi: \mathbb{R}^n \to \mathbb{R}^n$, $\psi \cdot \pi = \pi \cdot \phi$, is the induced diffeomorphism of the base.
If F is the generating function of the germ of a gradient map for the Lagrangian manifold L (PR.16, 3.3), then $F_\phi = (F+f) \cdot \psi^{-1}$ is the generating function of $\phi(L)$.

PROOF: Any fiber preserving symplectic diffeomorphism in $T^*(\mathbb{R}^n)$ is locally of the form
$(x,u) \to (\psi(x),\ {}^tD\psi(x)^{-1} \cdot (u+df(x)))$
for some ψ and f.

We now introduce the concept of a *family of generating functions* of a germ of a Lagrangian manifold.

DFN.9 A family of functions F_L of $u \in \mathbb{R}^k$, parametrized by $x \in \mathbb{R}^n$ satisfying (i) $\frac{\partial F_L}{\partial u} = 0$ and (ii) $\operatorname{rank}\left[\frac{\partial^2 F_L}{\partial u \partial u}, \frac{\partial^2 F_L}{\partial x \partial u}\right] = k$ is called the generating family of the germ of the Lagrangian manifold L, defined by the immersion $\lambda: L_F \to \mathbb{R}^{2n}$, where L_F is the submanifold of $\mathbb{R}^{k+n}$ determined by the equations in condition (i) and $\lambda(u,x) = \left(x, \frac{\partial F_L}{\partial x}\right)$; condition (ii) ensures that the equations in condition (i) are independent and therefore the dimension of L_F is n in $\mathbb{R}^{k+n}$.

In the model of the standard fibration $(T^*(\mathbb{R}^n), \mathbb{R}^n, \pi)$ and for a germ of a Lagrangian manifold given locally by equations (3.48)(PR.15, 3.3) a germ of a generating family is given by $F_L(u,x)=\sum_{j\in J} x_j u_j - F(u_j,x_i)$ $i\in I$, $I\cup J=\{1,2,\ldots,n\}$, $I\cap J=\emptyset$; the equations in condition (i), DFN.9, 4.3, above, provide the one set of the defining equations in 3.48 (i.e. $x_J=\frac{\partial F}{\partial u_J}$) and the map λ, the other (i.e. $u_I=\frac{\partial F}{\partial x_I}$). For an invariant description see A. Weinstein (A. WEINSTEIN).

It is natural to ask when two generating families over $\mathbb{R}^n$ (or B^n) generate equivalent Lagrangian submanifolds of $T^*(\mathbb{R}^n)$ (or M^{2n}). We describe two operations on families which do not change the Lagrangian submanifold being generated. Let $A_i \xrightarrow{\lambda_i} B_i$ be a differentiable submersion and F_i a function on B_i for i=1,2; we say that the families F_1 and F_2 are diffeomorphic iff there is a diffeomorphism $\Phi: A_1 \to A_2$ such that $\lambda_2 \cdot \Phi = \lambda_1$ and $F_2 \cdot \Phi = F_1$; diffeomorphic families obviously generate the same Lagrangian submanifold. Also, if F is a family defined on $A \xrightarrow{\lambda} B$ and r and s are nonnegative integers, we define the *(r,s) suspension* of F as the function $F_{r,s}$ on $B\times\mathbb{R}^{r+s}$ given by $F_{r,s}(u_j,v_1,\ldots,v_r,w_1,\ldots,w_s)=F(u_j)+\sum_{i=1}^{r} v_i^2-\sum_{i=1}^{s} w_i^2$; the conditions $\frac{\partial F_{r,s}}{\partial v_i}=o$ and $\frac{\partial F_{r,s}}{\partial w_i}=o$ make the definition of $F_{r,s}$ functionally dependent only on $B\times\{o,\ldots,o\}$ and therefore the Lagrangian submanifold generated by $F_{r,s}$ is the same as that generated by F.

Taking into account the diffeomorphisms of the base B and the fiber preserving property of a Lagrangian equivalence (DFN.13, 3.3), we arrive at the following definitions of equivalence for generating families of germs of Lagrangian manifolds.

DFN.10 Two generating families $F_1, F_2 \in C^\infty(\mathbb{R}^{k+n})$ are said to be equivalent iff there exist a diffeomorphism ϕ of $\mathbb{R}^{k+n}$, a diffeomorphism ψ of the base $\mathbb{R}^n$ and two functions $\tilde{f}\in C^\infty(\mathbb{R}^n)$ and $f\in C^\infty(\mathbb{R}^{k+n})$ such that

$$F_2\cdot\phi=F_1+f,\quad f=\tilde{f}\cdot\pi_n \quad \pi_n\cdot\Phi=\psi\cdot\pi_n$$

where $\pi_n:\mathbb{R}^{k+n}\to\mathbb{R}^n$

DFN.11 Two generating families $F_1\in C^\infty(\mathbb{R}^{k+n})$ and $F_2\in C^\infty(\mathbb{R}^{\ell+n})$ are said to be *stably equivalent* if there exist a family

$F_3 \in C^\infty(\mathbb{R}^{m+n})$, $m \le k, \ell$ such that F_1 and F_2 are equivalent to

appropriate suspensions of F_3 respectively

We now have available all the necessary concepts to state and prove

PR.14 All generating families of a germ of a Lagrangian manifold are mutually stably equivalent

PROOF: (V.M. ZAKALYUKIN, §1, Assertion 4 and L. HORMANDER §3.1)

We will set up a stable equivalence, in the sense of equivalence of generating families, between a representative of a germ of a generating family $F(u,x)$ at $(u_o,x_o)\in\mathbb{R}^{k+n}$ of L and its form in the standard fibration $F_L(\sum_I u_I x_I - \tilde{F}(u_I,x_J))$. There exists a diffeomorphism ϕ_1, which induces the identity on the base $(\mathbb{R}^n)$, such that $F\cdot\phi_1 = F_1 + Q$, where Q is a non-degenerate quadratic form and $\frac{\partial^2 F}{\partial u \partial u}(u_o,x_o)=o$; F_1 is stably equivalent (DFN.11, 4.3) to F. The immersion $\lambda: L_F \to \mathbb{R}^{2n}$, (DFN.9, 4.3), given an appropriate choice of I (such that $\frac{\partial^2 F}{\partial u \partial x_I}(u_o,x_o)\neq o$) defines a local diffeomorphism ϕ_2 and $F_2 = F_1\cdot\phi_2^{-1}$ is equivalent (DFN.10, 4.3) to F_1. Finally we consider the homotopy $F_t = F_L + t\cdot(F_2 - F_L)$, to $t\in[o,1]$; we will show that it defines a one-parameter family of diffeomorphisms, ϕ_t, of some nbd. of (u_o,x_o) which carry F_L into F_t; the composition $\phi_1\cdot\phi_2^{-1}\cdot\phi_t\big|_{t=1}$ is the desired equivalence. Indeed, $\frac{dF_t}{dt} = -(F_L - F_2)$, but $F_L - F_2$ belongs to the second power of the Jacobian ideal of F_L with respect to u, because the Lagrangian L_F is defined by graph $(d_u F)$, both for F_L and F_2; hence $\frac{\partial F_t}{\partial t} = \sum_I \frac{\partial F_L}{\partial u_I}\cdot f_I$, $f_I \in C^\infty(\mathbb{R}^{k+n})$; but $\frac{\partial F_L}{\partial u_I} = \sum_I \frac{\partial F_t}{\partial u_I}\cdot g_I$

$g_I \in C^\infty(\mathbb{R}^{k+n+1})$ because $\frac{\partial^2 F_2}{\partial u \partial u}(u_o,x_o)=o$; therefore

$\frac{\partial F_t}{\partial t} = \sum_I \frac{\partial F_t}{\partial u_I}\cdot F_I$, $F_I\big|_{L_F} = o$, $F_I \in C^\infty(U\subset\mathbb{R}^{k+n+1})$ with U a nbd. of (u_o,x_o),

which by a standard argument (G. WASSERMANN, Corollary 1.28) defines a one-parameter family of diffeomorphisms $\phi_t(\dot{u}_I=F_I,\dot{x}=o)$ in U.

The proof of PR.14, 4.3, actually ensures that the heuristic definitions of generating families (DFN.9, 4.3), and their equivalence (DFN.10 and DFN.11, 4.3), which was modelled to the equivalence of deformations, as families of functions (DFN.3, 4.2) or their germs (DFN.1, 4.2), are consistent and non-trivial. It is natural to ask next whether Lagrangian equivalence (DFN.13, 3.3) for germs of different Lagrangian submanifolds is also ensured by the equivalence of the corresponding generating families; the answer to this is yes.

PR.15 The germs of the Lagrangian submanifolds L_i, i=1,2, are Lagrange equivalent iff the corresponding generating families F_{Li} are stably equivalent.

PROOF: By DFN.10, PR.13 and PR.14, 4.3, the families F_{Li} lead into Lagrange equivalent submanifolds L_{Fi}; the construction of DFN.9, 4.3, is invariant under diffeomorphisms.

Comparing DFN.1 and DFN.2, 4.2, to DFN.10, 4.3, we can state, trivially

PR.16 Representations of germs of versal deformations are equivalent as generating families of Lagrangian manifolds

PROOF: Versal deformations are isomorphic, i.e. $\phi\epsilon E(k+n)$ and $\psi\epsilon E(n)$ are invertible and hence define local diffeomorphisms.

To define the notion of stability for germs of Lagrangian manifolds recall the general idea (see 4.1) that a given topological space and an equivalence relation on the space, an element of the space is called stable iff it is an interior point of its equivalence class. To this purpose we introduce to all spaces of functions and mappings the Weak C^∞-topology (Appendix III); Lagrangian submanifolds $L(n)$ of a symplectic manifold M are identified with their embeddings, i_1, in the total space M of some Lagrangian fibration; $i_1 \in C^\infty(L,M)$, L a fixed n-dimensional manifold and $\mathcal{L}(n) \stackrel{'}{=} \{f \in C^\infty(L,M)$, f a diffeomorphism and f(L) a Lagrangian submanifold of M$\}$; $i_1(\mathcal{L})$ as a subset of M, is an element of $\mathcal{L}(n)$, modulo a C^∞-diffeomorphism of L acting by composition on the right; the set $\Lambda(n) = {}^{\mathcal{L}(n)}/\mathrm{Diff}(L)$ has a topology induced from $C^\infty(L,M)$ as a subset of $C^\infty(L,M)/\mathrm{Diff}(L)$ and it is fibered over U(n)/O(n) with fiber $m^3(n)$ (see 4.1).

Then we define:

DFN.12 The Lagrangian submanifold $\mathcal{L}_{i_1}$ M is said to be *stable* iff every embedding i_1' in a nbd. of i_1 in the space, $C^\infty(L,M)$ of mappings from a fixed, n-dimensional, manifold L into M, is Lagrange equivalent (DFN.13, 3.3), by a Lagrangian equivalence close to the identity, to i_1, modulo a diffeo. of $\mathcal{L}$ close to the identify.

DFN.13 The germ $(\mathcal{L},P)$ of the Lagrangian submanifold $\mathcal{L}$ at P is said to be *stable* (or else $\mathcal{L}$ is said to be *stable at P*) iff for every other Lagrangian submanifold $\mathcal{L}'$ close to $\mathcal{L}$, there exist P' close to P, such that the germs (L,P) and $(\mathcal{L}',P')$ are Lagrange equivalent

One can easily see that in the case where i_1 is determined locally by a generating family $F_\mathcal{L}$ (DFN.9, 4.3), stability of i_1 (DFN.12, 4.3) corresponds to stability of $F_\mathcal{L}$ (DFN.4, 4.2).

PR.17 The germ of a Lagrangian submanifold $\mathcal{L}$ is stable iff its generating family $F_\mathcal{L}$ (DFN.9, 4.3) has a stable germ as a (versal) deformation of $F|_{\mathbb{R}^k \times \{0\}}$.

PROOF: By DFN.4, 4.3, stability of $F_\mathcal{L}$ ensures that in a nbd. of the origin in $\mathbb{R}^{k+n}$, and for every other function F' close to $F_\mathcal{L}$ one can find a point P' and maps $\tilde{\phi},\tilde{\psi},\tilde{f}$ as in DFN.3, 4.3, which therefore determine a fiber preserving (for $\pi_n:\mathbb{R}^{k+n}\to\mathbb{R}^n$) local diffeomorphism ϕ such that the conditions of DFN.10, 4.3 are fulfilled (with $f=\tilde{f}\circ\pi_n$) and the functions F' and $F_\mathcal{L}$ are equivalent as generating families of germs of Lagrangian submanifolds; by PR.15, 4.3, F' determines a germ of a Lagrangian manifold $\mathcal{L}'$ at P' equivalent (and close), to $\mathcal{L}$.

We would like to have the equivalent of PR.12, 4.3 for germs of Lagrange maps; we can only hope therefore to achieve a classification of stable and simple germs of Lagrange maps if and only if the germ of the corresponding generating family is related to a deformation of a simple singularity (DFN.6, 4.2); in rather general terms, a germ of a Lagrange map is simple if all nearby germs belong to finitely many equivalence classes; a simple germ can be non-stable, and a stable germ need not be simple.

PR.18 In the space of Lagrange maps $\pi\cdot i_1:\mathbb{R}^n\to\mathbb{R}^n$, $n<6$, with the Weak C^∞-topology, the maps which in a neighbourhood of each of their points, can be transformed to Lagrange maps with the following generating families, form an everywhere dense and open set:

Series $\mathcal{A}_k: \pm u_1^{k+1}+x_{k-1}\cdot u_1^{k-1}+x_{k-2}\cdot u_1^{k-2}+\ldots+x_2\cdot u_1^2$ $1\le k\le 5$

Series $\mathcal{D}_k: \pm u_1^2\cdot u_2 \pm u_2^{k-1}+x_{k-1}\cdot u_2^{k-2}+\ldots+x_3\cdot u_2^2$ $4\le k\le 5$

PROOF: Stability of germs of Lagrange maps is tautological to stability of germs of Lagrangian submanifolds and by PR.17, 4.3, is equivalent to stability of deformations of functions; by PR.12, PR.16, 4.3, and DFN.11, 4.3, one need only to rewrite the series $\mathcal{A}_k$ and $\mathcal{D}_k$ omitting the terms of power zero and one in u and the quadratic terms Q. The Lagrangian manifold is given in the space $(x,u)\in\mathbb{R}^{2n}$ by the equations $x_I=\frac{\partial F}{\partial u_I}$, $u_J=\frac{\partial F}{\partial x_J}$ where $I=\{1\}$ for the series $\mathcal{A}_k$ and $I=\{1,2\}$ for the series $\mathcal{D}_k$; the Lagrangian map is the projection onto the x-space.

APPENDIX I

Bases and subbases

Given a topological space (X, J), a family B of sets is a base for the topology J iff $B \subset J$ and $\forall x \in X$ and $\forall U$ (neighbourhood of x) $\exists V \in B$ such that $x \in V \subset U$.

PR. 1 B is a base for J. Then $\forall U \in J$ $U = \bigcup_{i \in I} V_i, V_i \in B$. The converse is also true.

THR. 1 A family of sets B is a base for some topology for the set $\bigcup\{V: V \in B\}$ iff $\forall U, V \in B$ and $\forall x \in U \cap V \exists W \in B$ such that $x \in W \subset U \cap V$.

Hence an arbitrary family of sets may fail to be a base for any topology. Still we may enquire whether there is a unique topology which is, in some sense, generated by our family of sets.

THR. 3 If S is any non-void family of sets, the family of all finite intersections of members of S is the base for a topology for the set $\bigcup\{A: A \in S\}$

A family of sets S is a subbase for a topology J iff the family of finite intersections of members of S is a base for J or equivalently:

PR. 2 Each member of J is the union of finite intersections of members of S.

PR. 3 Every non-void family of sets S is the subbase for some topology and this topology is uniquely determined by S. It is the smallest topology containing S.

Similarly one defines bases and subbases for the neighbourhood system of a point.

For a more detailed account see J.L. Kelley (J.L. KELLEY, pages 46-50).

APPENDIX II

On Lattices

DFN. 1 A lattice is a set L with two operations, called join ($\cup$) and meet ($\cap$), and satisfying the following axioms:

A1 $x \cup y = y \cup x$ and $x \cap y = y \cap x$ $\forall x, y \in L$ (commutativity)

A2 $x \cup (y \cup z) = (x \cup y) \cup z$ and $x \cap (y \cap z) = (x \cap y) \cap z$ $\forall x, y, z \in L$ (associativity)

A3 $x \cup (x \cap y) = x \cap (x \cup y) = x$ (absorption laws).

The above axioms are independent.

PR. 1 $x \cap x = x \cup x = x$ $\forall x \in L$, L a lattice (idempotent laws)

PROOF: $x \cup [x \cap (x \cup y)] = x \cup x$ and $x \cap [x \cup (x \cap y)] = x \cap x$ by applying A3 to the parentheses; again $x \cup [x \cap (x \cup y)] = x \cap [x \cup (x \cap y)] = x$ by applying A3 to the brackets.

DFN. 2 Two statements of lattice theory are called dual iff each can be obtained from the other by interchanging the two operations $\cup$ and $\cap$.

Axioms A1, A2 and A3 are self-dual, since any theorem that may be derived from these postulates can also be dualized within the theory (the proof being obtained from interchanging $\cup$ and $\cap$ at each step) lattice theory as a whole is a self-dual theory.

A lattice can be partially ordered:

DFN. 3 $x \leq y$ $x, y \in L$ iff $x \cup y = y$ iff $x \cap y = x$

The relation $\leq$ is uniquely defined by virtue of A3 as: $x \cap y = x \cap (x \cup y) = x$.

PR. 2 The relation $\leq$ is reflexive, anti-symmetric and transitive.

PROOF: By PR. 1, ($x \cup x = x$) and DFN. 3 $x \leq x$. Let $x \leq y$ and $y \leq x$; By DFN. 3 $x \cup y = y$ and $y \cup x = x$; by A1 $x \cup y = y \cup x$ and hence $x = y$. Let $x \leq y$ and $y \leq z$; by DFN. 3 $x \cup y = y$ and $y \cup z = z$; $x \cup z = x \cup (y \cup z)$ and by A2 $x \cup z = (x \cup y) \cup z = y \cup z = z$ i.e. $x \leq z$.

DFN.4 The converse (dual) relation ($\geq$) to the partial order $\leq$ is given by $x \geq y$ iff $y \leq x$ $x, y \in L$.

As a consequence of the partial order a number of derived concepts can be utilized in the study of lattices. Let P be a partially ordered set;

DFN.5 If $\forall x, y \in P$ either $x \leq y$ or $y \leq x$ is true, P is said to be linearly ordered (or simply or totally ordered) and it is called a chain.

DFN.6 An element $x \in P$ is called an upper (lower) bound of A, $A \subseteq P$ iff $a \leq x$ ($x \leq a$) $\forall a \in A$.

DFN.7 The sets of all upper (lower) bounds of a singleton set: $\{a\}$ $a \in P$, are called upper (lower) ideals of P.
We denote $\{x \in P: x \leq a\}$ by $\downarrow [a]$ and $\{x \in P: a \leq x\}$ by $\uparrow [a]$.
An upper ideal is often called the principal filter generated by a.

DFN.8 Generalizing DFN. 7 we denote $\uparrow A (\downarrow A)$, $A \subseteq P$, the sets $\{x \in P: a \leq x \ \forall a \in A\}$ ($\{x \in P: x \leq a \ \forall a \in A\}$).
Then $\uparrow A = \bigcap_{a \in A} \uparrow [a]$ and $\downarrow A = \bigcap_{a \in A} \downarrow [a]$ ($\bigcap$ denotes point set intersection here) and $\uparrow$, $\downarrow$ can be considered as maps from B(P) into B(P). (the power set of P).

DFN.9 An element $m \in P$ is called minimal (maximal) iff $x \leq m$ ($m \leq x$) $\Rightarrow x = m$.
A minimal (maximal) element is called least or a lower unit or zero (greater or an upper unit or universal element) iff $m \leq x$ ($x \leq m$) $\forall x \in P$.

DFN.10 A least upper bound (supremum, join, union) of A, $A \subseteq P$, is an element $l \in P$ such that 1st $a \leq l \ \forall a \in A$ and 2nd if $a \leq x \ \forall a \in A$, $x \in P \Rightarrow l \leq x$.
Dually a greatest lower bound (infimum, meet, intersection) of A, $A \subseteq P$, is an element $g \in P$ such that: 1st $g \leq a \ \forall a \in A$ and 2nd if $x \leq a \ \forall a \in A$, $x \in P \Rightarrow x \leq g$.

Generally in a partially ordered set least upper (and greatest lower) bounds need not exist; but if they do, they are unique. In a lattice these elements exist for finite subsets as can be seen from the following proposition and by induction.

PR. 3 $\forall$ x, y $\in$ L, L a lattice, $x \cup y$ and $x \cap y$ is the least upper bound and the greatest lower bound respectively of $\{x,y\}$ under the relation $\leq$ (DFN. 3).

PROOF: $x \cup (x \cup y) = (x \cup x) \cup y = x \cup y$ i.e. $x \leq x \cup y$; similarly $y \cup (x \cup y) = y \cup (y \cup x) = (y \cup y) \cup x = y \cup x = x \cup y$ i.e. $y \leq x \cup y$. Hence $x \cup y$ is an upper bound of $\{x,y\}$. Next let $x \leq z$ and $y \leq z$; then $x \cup z = z$ and $y \cup z = z$; therefore $(x \cup y) \cup z = x \cup (y \cup z) = x \cup z = z$ and $x \cup y < z$. Consequently $x \cup y$ is the least upper bound of $\{x,y\}$. The proof dualized gives $x \cap y$ as the greatest lower bound.

PR. 3 suggests that one can reverse the process of determining the algebraic structure of a lattice and make the order relation fundamental so that a lattice can be said to be a partially ordered set in which every pair of elements has a least upper and a greatest lower bound. As operations in general determine relations one could dispense with operations entirely and deal only with relations; however it is more convenient to work with operations separately. Considering consequently homomorphisms between lattices one should rather specify whether the homomorphism is preserving only the order relation (monotone map, isotone, order preserving and dually order-inverting, antitone map, dual homomorphism) or the operations as well; however every lattice homomorphism is a partial order homomorphism. It is worth noticing that in the special case of isomorphisms the converse is true i.e. if f and f^{-1} are order preserving isomorphisms then f preserves join and meets i.e. it is a lattice isomorphism; we will use this latter in the proof of PR. 6.

DFN. 11 Given a partially ordered set P and $Q \subseteq P$, Q is called a subpartially ordered set iff there exists a relation ρ on Q such that iff $x, y \in Q$ and $x \leq y \Rightarrow x \rho y$. A subset A of a lattice is called a sublattice iff A is dosed under the lattice operations.

Every sublattice is automatically a subpartially ordered set.

Completeness, closure operations, Galois connections, Dedekind cuts

DFN. 12 A partially ordered set P is called join-complete, meet complete or complete iff every subset of P has a least upper bound, a greatest lower bound or both, respectively.

It is possible to define a number of degrees of completeness between that of closure and completeness as in DFN. 12; these are defined by limiting the size of the sets (i.e. specifying their cardinality) which are required to have joins or meets.

An example of a complete lattice is the family of subalgebras of a given algebra (J.C. ABBOTT, Theorem 4.10).

A particular class of endomorphisms of some interest in the theory of partially ordered sets and lattices, are the closure operations.

DFN. 13 An operation $0: P \to P$ $(x \mapsto 0(x) = \bar{x})$ on a partially ordered set P is called a closure (interior) operation iff it satisfies:

a1 $x \leq y \Rightarrow \bar{x} \leq \bar{y}$ (isotone)

a2 $x \leq \bar{x}$ (extensive) ($x^{o} \leq x$ (intensive))

a3 $\bar{x} = \bar{\bar{x}}$ (idempotent)

Respectively an element $x \in P$ is called closed (open) iff $x = \bar{x}$ $(x = x^{o})$.

PR. 4 In a complete lattice with a closure operation the set of closed elements form a complete lattice under the operations.

$$\prod_{\alpha} x_{\alpha} = \bigcap_{\alpha} x_{\alpha} \qquad \bigsqcup_{\alpha} x_{\alpha} = \overline{\bigcup_{\alpha} x_{\alpha}} \qquad \alpha \in I \text{ an index set.}$$

PROOF: By a1 $\bigcap_{\alpha} x_{\alpha} \leq x_{\alpha} \Rightarrow \overline{\bigcap_{\alpha} x_{\alpha}} \leq \bar{x}_{\alpha} = x_{\alpha} \forall \alpha$; hence $\overline{\bigcap_{\alpha} x_{\alpha}} \leq \bigcap_{\alpha} x_{\alpha}$. But by a.2 $\bigcap_{\alpha} x_{\alpha} \leq \overline{\bigcap_{\alpha} x_{\alpha}}$; therefore $\bigcap_{\alpha} x_{\alpha} = \overline{\bigcap_{\alpha} x_{\alpha}}$ i.e. $\bigcap_{\alpha} x_{\alpha}$ is closed if the x_{α} are closed.

DFN.14 A closure operation in a lattice with a lower unit is called a Kuratowski closure operation iff it satisfies further:

a4 $\overline{x \cup y} = \overline{x} \cup \overline{y}$

a5 The lower unit is a closed element.

The most significant examples are the topological spaces. For another example see Zoltan Domotor (Z. DOMOTOR, theorem 1).

There is a simple way to construct closure operations from homomorphisms:

DFN15 Given two partially ordered sets $< P_1, \rho_1 >, < P_2, \rho_2 >$ and a pair of maps f_1, f_2, $f_1: P_1 \rightarrow P_2$ and $f_2: P_2 \rightarrow P_1$; we call (f_1, f_2) a Galois connection between P_1 and P_2 iff 1st f_1, f_2 are antitone, 2nd $x \; \rho_1 \; f_2 \cdot f_1(x) \; \forall x \in P_1$ and $y \rho_2 \; f_1 \cdot f_2 \; (y) \;\; \forall y \in P_2$.

DFN.16 A Galois connection induced by a relation ρ between two sets X and Y is the Galois connection between their power sets B(X) and B(Y) defined by: $f_1(A) = \{y \in Y: x \rho y \;\; \forall x \in A\} \subseteq B(Y)$, $A \in B(X)$ and $f_2(B) = \{x \in X: x \rho y \;\; \forall y \in B\} \subseteq B(X)$, $B \in B(Y)$.

The DFN.16 is consistent as (f_1, f_2) is indeed a Galois connection; the proof goes as follows.

Let A, $A' \in B(X)$, $A \subseteq A'$ and $y \in f_1(A')$ i.e. $x \rho y \;\; \forall \; x \in A'$ whence $\forall x \in A$; thus $y \in f_1(A)$ and $f_1(A') \subseteq f_1(A)$ i.e. f_1 is antitone. Similarly for f_2. Now let $x \in A$ and $B = f_1(A)$; then $x \rho y \;\; \forall y \in f_1(A) = B$; hence $x \in f_2(B) = f_2 \cdot f_1(A)$ i.e. $A \subseteq f_2 \cdot f_1(A)$. Similarly $B \subseteq f_1 \cdot f_2(B) \;\; \forall B \in B(Y)$.

PR. 5 If (f_1, f_2) is a Galois connection between P_1 and P_2, then $f_2 \cdot f_1$ is a closure operation in P_1 and $f_1 \cdot f_2$ is a closure operation in P_2.

PROOF: f_1, f_2 are antitone and consequently $f_1 \cdot f_2$ and $f_2 \cdot f_1$ are isotone. a2 of DFN. 13 is the 2nd property of DFN. 15. Finally let $f_2 \cdot f_1(x) = \overline{x}$, $x \in P_1$ and denote ρ_1, ρ_2 by $\leq$. $x \leq \overline{x} \Rightarrow \overline{x} \leq \overline{\overline{x}}$ and $f_1(\overline{x}) \leq f_1(x)$; but $f_1(x) \leq f_1 \cdot f_2 \cdot f_1(x) = f_1(\overline{x})$ and $f_2 \cdot f_1(x) \geq f_2 \cdot f_1(\overline{x})$ i.e. $\overline{x} \geq \overline{\overline{x}}$. Hence $\overline{x} = \overline{\overline{x}}$ and a3 is satisfied too.

As an example consider the pair $\uparrow, \downarrow$ of operations (endomorphisms) defined on B(P) as a Galois connection induced by a partial order relation ($\leq$) on P . The combined maps $\uparrow.\downarrow$ and $\downarrow.\uparrow$ are closure operations on B(P) by virtue of PR. 5; the closed elements C_p of B(P) are characterized by the property. $A = \uparrow.\downarrow A(\downarrow\uparrow A)\ A \in B(P)$. Such elements are called Dedekind cuts after Dedekind's method to define the real numbers out of the non-complete lattice of rationals. As can be seen from the following propositions this same method can be used to embed any partially ordered set to a complete lattice; in other words this lattice contains a subpartially ordered set isomorphic to the original set (H. MACNEILLE).

PR. 6 If L_1, L_2 are complete lattices, then the sets of closed elements of L_1, L_2: $C_{L_1} = \{x \in L_1: x = \bar{x} = f_2 \cdot f_1(x)\}$ and $C_{L_2} = \{x \in L_2: x = \bar{x} = f_1 \cdot f_2(x)\}$ are complete lattices and f_1, f_2 are dual isomorphisms between them.

PROOF: By PR. 4, C_{L_1} and C_{L_2} are complete lattices. Let $x \in C_{L_1}$, i.e. $x = f_2 \cdot f_1(x)\ (= \bar{x})$; if $y = f_1(x) \Rightarrow y = f_1 \cdot f_2 \cdot f_1(x) = f_1 \cdot f_2(y)$ i.e. $y \in C_{L_2}$. Therefore f_1 maps C_{L_1} into C_{L_2} ; similarly f_2 maps C_{L_2} into C_{L_1}. Further since $x = f_2 \cdot f_1(x)\ \forall x \in C_{L_1}$ and $y = f_1 \cdot f_2(y)\ \forall y \in C_{L_2}$, $f_2 \cdot f_1$ and $f_1 \cdot f_2$ are the identity maps on C_{L_1} and C_{L_2} respectively i.e. f_1, f_2 are one-to-one and onto. The fact that they are antitone means they are dual order isomorphisms. Consequently they are dual lattice isomorphisms.

PR. 7 If P is any partially ordered set then P can be embedded in the complete lattice of all its Dedekind cuts.

PROOF: By PR. 6 the set of Dedekind's cuts forms a complete lattice. Given that $\forall x \in P\ \downarrow[x] = \downarrow\uparrow[x]$ and dually $\uparrow[x] = \uparrow\downarrow[x] \Rightarrow \downarrow[x] = \downarrow\uparrow\downarrow[x]$; the map $\downarrow: P \rightarrow C_p \subseteq B(P)$, being isotone, is a partial order isomorphism; hence the set: $\{\downarrow[x] : \forall x \in P\}$, i.e. the set of principal lower ideals of P, is a subpartially ordered set of C_p isomorphic to P.

APPENDIX III

Concepts in Differential Geometry, Differential Topology and the Theory of Singularities

III.1 General Manifold theory

DFN.1 A *manifold* of dimension n is a Hausdorff space such that every point has an open nbd. homeomorphic to an open subset of $\mathbb{R}^n$.

For our purposes manifolds are assumed locally connected and paracompact

DFN.2 An *atlas* on a manifold M (dim. n) is a set of homeomorphisms $\phi_i = U_i \to V_i$ (i∈I) such that $M = \bigcup_{i \in I} U_i$, where U_i is open in M and V_i open in $\mathbb{R}^n$.

DFN.3 Let X be a topological space. A *pseudogroup* G on X is a collection of homeomorphisms $\phi: U \to V$, U, V open in X such that

(i) The identify belongs to G

(ii) $\phi, \psi \in G \Rightarrow \phi \cdot \psi \in G$, where $\phi: U \to V$, $\psi: U' - V'$ and $\phi \cdot \psi : \psi^{-1}(U \cap V') \to \phi(U \cap V')$

(This composition is associate)

(iii) $\phi \in G \Rightarrow \phi^{-1} \in G$

(iv) if $\phi: U \to V$ is a bijection, $U = \bigcup_{i \in I} U_i$ and $\phi_{|U_i} : U_i \to \phi(U_i)$ is in G for every i $\Rightarrow \phi \in G$

(v) if $\phi: U \to V$, $\phi \in G$ and W open and $W \subset U$ $\Rightarrow \phi_{|W} : W \to \phi(W)$ is in G.

DFN.4 An atlas (ϕ_i, U_i), i∈I, on a manifold M is a *G-atlas* iff $\phi_i \cdot \psi_j^{-1} \in G$ for all i,j∈I.

A *G-structure* on a manifold M is a maximal, G-atlas.

We will be mainly dealing with smooth manifolds, i.e. manifolds with a (Diff)∞-structure (the set of all local homeomorphisms on $\mathbb{R}^n$ which are differentiable to any power); differentiable class k usually means k at least one; $k=o,\omega$ are refered to as topological and analytic respectively.

DFN.5 Let M an n-dimensional C^∞-manifold and N a connected subset of M. N is said to be a *submanifold* of M of dimension $m(m<n)$ iff $\forall\, x\in N$ there exist an element of the differentiable structure (ϕ,U) such that $x\in U$ and $\phi(U\cap N)$ is an open subset of $\mathbb{R}^m\times\{0\}\subset\mathbb{R}^n$.

According to this definition submanifolds inherit the topology (induced topology) of the main manifold.

DFN.6 Let M be a C^k-differentiable manifold of dimension n

$f\in C^k(M)$ means $f:M\to\mathbb{R}$ and for every (chart) (U,ϕ),

$f\cdot\phi^{-1}:\phi(U)\to\mathbb{R}$ is C^k-differentiable

$\phi\in C^k(M,N)$ means $\phi:M\to N$ and for every C^k-differentiable function $f\in C^k(N)$ the pullback $f\cdot\phi$ is C^k-differentiable. f,ϕ are smooth iff f,ϕ are C^k-differentiable for every k.

DFN.7 Given $\phi\in C^k(M,N)$, the derivative map $D\phi$ at $p\in M$, is a linear map $\mathbb{R}^m\to\mathbb{R}^n$ (m=dim M and n=dim N) given by $(d\psi\cdot\phi\cdot\chi^{-1})$ at $\chi(p)$ where (U,ψ) and (V,χ) are charts for M and N, $p\in U$ and $\phi(U)\subset V$

DFN.8 Let $\phi\in C^k(M,N)$ and $D\phi$ at p has the maximum possible rank

(i) if dim M < dim N ϕ is said to be an *immersion at p*

(ii) if dim M > dim N ϕ is said to be a *submersion at p*

(iii) if ϕ is an immersion (submersion) at p, for every p, then ϕ is said to be an *immersion (submersion)*

(iv) An immersion which is a homeomorphism onto its image is called an *embedding*

DFN.9 A subset N of a manifold M is called an *immersed submanifold* if the inclusion map is an immersion; for a submanifold clearly the inclusion map is an embedding; and a *closed manifold* is a submanifold for which the inclusion map is proper, i.e. the preimage of every compact set is compact.

III.2 Topologies on function spaced

Let X, Y be topological spaces and $C^0(X,Y)$ the set of continuous maps from X to Y; the *fine* topology on $C^0(X,Y)$ is defined by considering as an open nbd. of $f\in C^0(X,Y)$ all $g\in C^0(X,Y)$ such that $\text{graph}(g)\subset U$ where U is an open nbd. of graph(f) in X×Y; if U is an open subset of X×Y, then $\mathcal{W}(U)\doteq\{f: \text{graph}(f)\subset U\}$ is a typical open set in the fine topology.

Given a continuous map $\pi:Y\to X$, a *section* of π is a map $s:X\to Y$ such that $\pi\cdot s=id_X$; one can topologize the space of sections by $S(V)\doteq\{s:\pi\cdot s=id_X$ and $s(X)\subset V$, V open subset of Y$\}$; one then has

PR. 1 The space of sections of $\pi:Y\to X$ form a subspace of $C^0(X,Y)$ with the fine topology.

PROOF: We will show that $S(V)=\mathcal{W}(X\times V)\cap$ sections and hence $S(V)$ is open in the fine topology on sections.

Let U open in X×Y and $f\in C^0(X,Y)$ such that $\text{graph}(f)\subset U$ (i.e. $f\in\mathcal{W}(U)$); $U=\cup A\times B$ where A open in X and B open in Y; let $V=\cup(\pi^{-1}(A)\cap B)$; $g\in S(V)$ implies $\forall x\in X$ $g(x)\in V$ and $\forall x\in X$ there exist A and B as above such that $g(x)\in\pi^{-1}(A)\cap B)$ i.e. $\forall x\in X$, $g(x)\in B$ and $\pi(g(x))\in A$ or $(x,g(x))\in A\times B\subset U$; hence $g\in S(V)$ implies $\text{graph}(g)\subset U$, i.e. $S(V)\subset\mathcal{W}(U)$.

Conversely: $g\in\mathcal{W}(U)\cap$ sections means $\forall x$ such that $(x,g(x))\in U$ that there are A and B such that $x\in A$, $g(x)\in B$; but $x=(\pi\cdot g)(x)$ and $g(x)\in\pi^{-1}(A)$, i.e., $g\in S(V)$.

DFN 10 The *fine C^k-topology* on $C^k(M,N)$, where M, N are finite dimensional manifolds, is the topology induced by the fine topology on the space of sections $M\to J^k(M,N)$, where $J^k(M,N)$ is the space of k-jets from M to N.

The *coarse C^k-topology* is generated by finite intersections of sets $\{f:j^k(f)(C)\subset V\}$ for some fixed, compact C in M and open V in $J^k(M,N)$.

For M compact, the two topologies coincide; one can see this when a basis of neighbourhoods is constructed for f:

PR.2 Let $f:M\to N$ be a C^k-differentiable map. Let $\{C_i\}$ be a locally finite family of compact subsets of M. For each $i\in I$ let B_i be an open subset of $J^k(M,N)$ (an open nbd. of $j^k(f)(C_i)$). Then

(i) $\{g: j^k(g)(C_i)\subset B_i,\ i\in I\}$ is an open nbd. of f in the fine C^k-topology

(ii) Fixing $\{C_i\}$ such that $M=\underset{i\in I}{\cup} U_i$, varying $\{B_i\}$ gives us a basis of nbd.s of f

PROOF: Cover M by coordinate nbd. s and choose C_i such that $f(C_i)\subset V_i$, some coordinate nbd. of N and $C_i\subset U_i$, some coordinate nbd. of M; specify B_i using two coordinate charts, i.e. $B_i\subset J^k(U_i,V_i)=U_i\times V_i\times$ vectorspace.

Let $\pi:J^k(M,N)\to M$. To prove (i) need only show that $\{g:j^k(g)(C_i)\subset B_i\}$ is a nbd. of f, i.e. need only to construct U open in $J^k(M,N)$ such that $j^k(g)(M)\subset U$ implies $j^k(g)(C_i)\subset B_i$, $\forall_i$; $\forall x\in C_i$ (not necessarily $\underset{i\in I}{\cup} C_i=M$) let $I(x)=\{i\in I:x\in C_i\}$; $j^k(f)(x)\in \underset{i\in I(x)}{\cap} B_i=B(x)$ is then open and there exist an open nbd. $A(x)$ of x in M such that $j^k(f)(A(x))\subset B(x)$ and $A(x)\cap C_i=\emptyset$ unless $i\in I(x)$; $U=\pi^{-1}(M\setminus\cup C_i)\cup(\underset{x\in\cup C_i}{\cup}(B(x)\cap\pi^{-1}(A(x))))$, hence $j^k(f)(M)\subset U$.

Conversely suppose $j^k(g)(M)\subset U$ and let $x'\in C_i$; then $j^k(g)(x')\in \pi^{-1}(A(x))\cap B(x)$ for some x in some C_k; projecting via π, $x'\in A(x)$, $A(x)\cap C_i\neq\emptyset$ and $i\in I(x)$; hence $B(x)\subseteq B_i$, $j^k(g)(x')\in B_i$, i.e. $j^k(g)(C_i)\subseteq B_i$.

To prove (ii) given U, open in $J^k(M,N)$ such that $j^k(f)(M)\subset U$, choose $\{C_i\}$ a covering, and put $B_i=U$ for each i; then $\{g:j^k(g)(C_i)\subset B_i,\ \forall i\in I\}=\{g:j^k(g)(M)\subset U\}$

One can utilize the fact that a countable intersection of open and dense subsets of a complete metric space with its metric induced topology, is dense, to prove that $C^\infty(M,N)$ is a *Baire space* in the fine C^∞-topology; the property is however true, but hard to prove since $C^\infty(M,N)$ is not even a topological vector space as scalar multiplication is not continuous, for M non-compact; the idea is to introduce a metric in $J^k(M,N)$ $k\geq o$, via $\rho_k(j^k(f), j^k(g))=\sup_{x\in M}\{\rho_k(j^k(f)(x), j^k(g)(x)\}$, by $\rho(f,g)=\sum_{k=1}^{\infty} 2^{-k} \frac{\rho_k}{1+\rho_k}$; $(C^\infty(M,N),\rho)$ is a complete metric space; the corresponding topology is the fine C^∞-topology.

For function spaces, such as $C^\infty(U,\mathbb{R})$, U open in M, a zero nbd. basis is given by the sets $\{f\in C^\infty(U,\mathbb{R}): \| f\|_{k,C} < \varepsilon, \forall k$, C compact in U$\}$ for the coarse $C^k(C^\infty)$-topology, where the norm $\| f\|_{k,C}=\sup\{|D_\alpha f(a)|, x\in C, \alpha=(\alpha_1,\dots,\alpha_n), |\alpha|=\sum_{i=1}^{n}\alpha_i\leq k\}$; a typical open set of $f\in C^\infty(U,\mathbb{R})$ is $\{g\in C^\infty(U,\mathbb{R}): j^k(f-g)(C)\subset U$, U open zero nbd. in $J^k(U,\mathbb{R})\}$.

It easily seen that:

PR. 3 Every sequence of functions converges in the $C^k(C^\infty)$-topology iff there exist a compact set such that the sequence together with its k first (all) derivatives converges uniformly on it, and equals the limit outside it, with the exception of at most finitely many of its elements.

III.3 The theorem of Darboux

The theorem of Darboux states that any non-singular closed 2-form in $\mathbb{R}^{2n}$ is locally isomorphic to the standard form, that is that in a suitable chart at a point, it has the standard expression $\omega(x,u)=\sum_{i=1}^{n} dx_i \wedge du_i$. The classical proof is lengthy; it is however a corollary of the following more general statement on Banach manifolds.

PR.4 Let M be a self-dual Banach space and ω a non-singular closed 2-form on an open set U of M. Let $x_o \in U$; then ω is locally isomorphic at x_o to the constant form $\omega(x_o)$.

PROOF: (S. LANG) Let $\omega: M\times M \to \mathbb{R}$ be continuous and bilinear and such that the induced mappings between M and M* are toplinear isomorphisms; then ω is called non-singular; if such an ω exist, then M is self-dual.

Let $\omega_o=\omega(x_o)$ and $\omega_t=\omega_o+t\cdot(\omega-\omega_o)$ $t\in[o,1]$; let us find a vector field X_t, locally at o such that $X_t(\omega_t)=\omega_o$; then the local isomorphism X_1 satisfies the requirements of the proposition. By Poincaré Lemma there exist a one-form θ, locally at o such that $\omega-\omega_o=d\theta$; assume $\theta(x_o)=o$; let X_t be such that $i_{X_t}(\omega_t)=-\theta$; by the existence theorem for flows, X_t can be integrated at least to t=1; but $X_o(\omega_o)=\omega_o$ $(X(o,x)=x$ and $DX(o,x)=id.)$ and $\frac{d}{dt}(X_t(\omega_t))=o$; indeed $d\omega_t=o$ implies $\pounds_{X_t}(\omega_t)=d(i_{X_t}(\omega_t))$ and $\frac{d}{dt}(X_t(\omega_t))=X_t(\frac{d}{dt}\omega_t)+X_t(\pounds_{X_t}(\omega_t))=X_t(\frac{d}{dt}\omega_t+d(i_{X_t}(\omega_t)))=X_t(\omega-\omega_o-d\theta)=o$. Q.E.D.

The theorem is also used in the following form

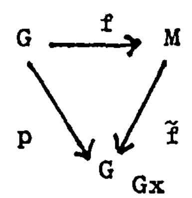
PR.5 If Ω is a globally defined one-form on a differentiable manifold M, such that $\Omega\,(d\Omega)^n\neq o$ everywhere on M, then M is a contact manifold in the wider sense, i.e. there exist a coordinate covering of M such that $\Omega=du-\sum_{i=1}^{n} u_i dx_i$

PROOF: The condition $\Omega\,(d\Omega)^n\neq o$ implies that the rank of $d\Omega$ is n:

$(d\Omega)^{n+1}=0$ since $(d\Omega)^{n+1}$ is a (2n+2) form in a (2n+1)-dimensional manifold and $(d\Omega)^n=o$ contradicts $\Omega\ (d\Omega)^n\neq o$; hence $d\Omega$ is non-singular and PR.1 applies, i.e. $d\Omega=\sum_{i=1}^{n} dx_i\wedge du_i$; Ω then is given by $du\pm x_i du_i$ and du is either a linear combination of the dx's and the du's or not; in terms of these coordinates $\Omega\wedge(d\Omega)^n=$ $du\wedge dx_1\wedge\ldots\wedge dx_n\wedge du_1\wedge\ldots\wedge du_n$ and since it is different from zero everywhere du is independent of the dx's and du's i.e. u can only be a function of the remaining coordinate and one can extend (u,x's,u's) to a new coordinate system such that $\Omega=du-\sum_{i=1}^{n} dx_i\wedge du_i$.

III.4 Homogeneous spaces

Given a Lie group G, a closed subgroup H of G and the quotien topology for G/H, the map $G\to G/H$ is continuous; any topological space homeomorphic to G/H for some G and H is called a *homogeneous space*; G/H can be given the structure of a differentiable manifold (C. CHEVALLEY); furthermore if G acts transitively on M (M locally compact), i.e. given $x,y\in M$ there exist $g\in G$ such that $g(x)=y$, then M is diffeomorphic to G/Gx, $\forall x\in M$, where $Gx\dot{=}\{g\in G: g(x)=x\}$ is the isotropy subgroup of G at x. For example, $G_{n,n}$, the set of n-dimensional linear subspaces of $\mathbb{R}^{2n}$, i.e. the Grassmannian space of n-planes, is the homogeneous space $O(2n)/O(n)\times O(n)$ where $O(2n)'\dot{=}\{A\in O(n): A(P)=P$ $A(P^{\perp})=P^{\perp}, P\in G_{n,n}\}$ and $O(2n)_P=O(n)\times O(n)$, since O(2n) acts transitively on $G_{n,n}$. To prove that M is homeomorphic to G/Gx, fix $x\in M$ and define $f:M\to G$ by $f(g)=g(x)$; the transitivity of the action of G implies that $f(G)=M$; therefore $f(g)=f(g')$, i.e. $g(x)=g'(x)$, is equivalent to $g'^{-1}g\in Gx$ and f factors through G/Gx; diagramatically:



$\tilde{f}$ is a continuous bijection; since p is continuous, it is sufficient to prove that f is open; let U be an open subset of $g\epsilon G$; choose a compact nbd. V of the identity e in G such that $V=V^{-1}$ and $gV^2 \subset U$ (where $V^{-1}=\{g\epsilon G: g\epsilon V \Rightarrow g^{-1}\epsilon V\}$ and $V^2=\{g\epsilon G: g\epsilon V \Rightarrow g^2 \epsilon V)$ the existence of V follows from the continuity of group action; there exists a sequence $g_n \epsilon G$ such that $G=\cup g_n V$ (G is a manifold and second-countable), hence $M=\cup(g_n V)(x)$; $g_n V$ is compact and so is therefore $(g_n V)(x)$ and we have expressed M as a union of countably many compact sets; at least one $(g_n V)(x)$ has an interior point (because if we choose a sequence $x_n \epsilon M$, open sets $W_n \subset M$ such that $\overline{W}_{n+1} \subset W_n$, W_n compact and $a_n \epsilon W_n \setminus (g_n V)(x)$, $\overline{W}_{n-1} \cap (g_n V)(x)=\emptyset$, then $\cap \overline{W}_n=\emptyset$, a contradiction), hence V(x) has an interior point say g(x); then x is an interior point of $(g^{-1}V)(x)$ $\subset V(x)$ (since $g^{-1}\epsilon V$ and $V^{-1}=V$) and $V(x) \subset (g^{-1}U)(x)$, i.e. g(x) is an interior point of U; the choice of U and g was arbitrary Q.E.D.

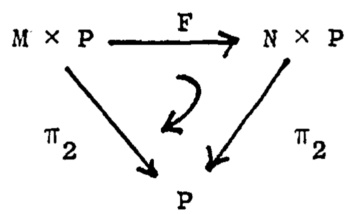
III. 5 Vector Bundles

By a C^∞-*bundle* we mean a triple (M,B,π) where M, B are smooth manifolds and π is a C^∞-map of M into B with the following property. There exists a manifold F such that for any $x\epsilon B$, there exists an open nbd. U of x in M and a diffeomorphism $\phi: \pi^{-1}(U) \to U \times F$ such that $\pi_1 \circ \phi = \pi$, where π_1 is the projection of $U \times F$ on its first factor.

By a C^∞-*vector bundle* we mean a bundle such that each fiber $M_x=\pi^{-1}(x)$ is provided with the structure of an $\mathbb{R}$-vector space so that the set of its sections over any open subset U of B is a $C^\infty(U)$-module.

III.6 Families of functions

Given M,N,P differentiable manifolds, by a *family of maps* from M to N, parametrized by P we mean the commutative diagram:



where π_2 is the projection onto the second factor.

An n-parameter *family of functions* in k-variables is a map $F : \mathbb{R}^n \to C^\infty(\mathbb{R}^k)$; it is a stable object iff $\tilde{F} : \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R} \times \mathbb{R}^n$ defined by $\tilde{F}(u,x)=(F(x)(u),x)$ is stable in the set of mappings fibered over $\mathbb{R}^n$; if $f \in C^\infty(\mathbb{R}^k)$ can be embedded in a stable finite dimensional family, we say that f has *finite codimension*; for example a polynomial of degree n, (in one variable (u)) $F : \mathbb{R}^{n-1} \to C^\infty(\mathbb{R})$, $F(x_1,\ldots,x_{n-1})= x_1u+\ldots+x_{n-1}u^{n-1}+u^{n+1}$, is a stable (n-1)-parameter family (of the coefficients $(x_1,\ldots,x_{n-1})$).

III.7 On transversality

The following considerations provide us with some insight into the definition of transversality to follow; two linear subspaces L_1 and L_2 of a linear (vector) space L are said to be transversal if $L_1+L_2=L$ (e.g. two planes in $\mathbb{R}^2$ meeting at a non-zero angle); next consider two manifolds M and N and two maps of f and g from M and N respectively into the same manifold A, i.e.

$$M \xrightarrow{f} A \xleftarrow{g} N$$

such that either $f(x)\neq g(y)$ or $f(x)=g(y)=a\in A$, $(x,y)\in M\times N$, and the images of the tangent planes to M and N at x and y under the action of f and g respectively are transversal in the tangent space of A at a:

$$f_*(T_x(M))+g_*(T_y(N))=T_a(A)$$

The two maps are said to be transversal at (x,y) or transversal if they are transversal at every point $(x,y)\in M\times N$ (e.g. two lines in R^3 are transversal only if they do not intersect)

DFN.11 Let A and M be smooth manifolds, B a submanifold C of A and f a smooth map from M into A; then f is said to be *transversal* to the submanifold B if it is transversal to the embedding $i=B\to A$.

Since the image of the tangent space and the tangent space to the image are not one and the same thing, note that a map of a line into a plane can fail to be transversal to a given line in the plane even when the image is normal to the given line.

Let M and N at least C^k-differentiable manifolds and let $J^\ell(M,N)$, $k>\ell\geq o$, the manifold of the ℓ-jets of C^k-differentiable mappings $C^k(M,N)$ from M into N; let B a $C^{(k-\ell)}$-differentiable submanifold of $J^\ell(M,N)$ of codimension c.

PR. 6 The set of maps in $C^k(M,N)$, whose ℓ-jet extensions are transversal to B, is everywhere dense, in the fine C^k-topology, if $(k-\ell)>\max\{(m-c),o\}$, m=dim M.

PROOF R. Thom and H. I. Levine in: Lecture Notes in Mathematics, Vol. 192, §7.1 (C.T.C. WALL).

The following is knows as the Weak transversality theorem:

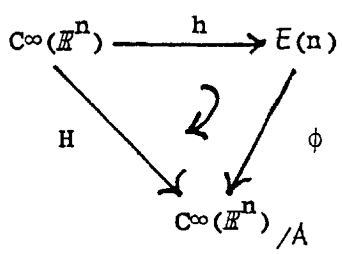
PR.7 Let $f:M\to A$ be a smooth map of a compact, smooth manifold M into a smooth manifold A and let B a compact submanifold of A. Then the maps f that are transversal to B form an open and everywhere dense subset of the function space of all maps from M into A with the C^k-topology $k>\max\{(\dim M-\dim A+\dim B),o\}$.

If B is not compact, "open" must be replaced by "intersection of a countable family of open sets"; if M is not compact one uses the fine topology.

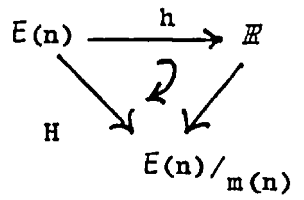
There is a generalization in the case where the space of maps from M into N is taken to be the space of sections of a fiber bundle with base M and fiber N or canonical or contact or volume preserving mappings (or transformations) (S.M. VISIK). However, since the whole purpose of this exercise was in fact to reduce the statements and proofs about Lagrangian (Legendrian) fibrations, i.e. canonical (contact) invariants, into statements about, rather general, families of functions, one only needs a rather simplified version of the PR. 6 above, on smooth manifolds and mappings and only for the Weak (coarse) C^∞-topology (G. WASSERMANN, §1.22 and 1.23).

III.8 The Local Ring

Given two topological spaces X and Y, a *map germ* f at x, $x \in X$, is an equivalence class $\tilde{f}$ of continuous maps $f:U \to Y$, U a nbd. of x in X, where f is equivalent to $f':V \to Y$ iff there exist a nbd. W of x in X such that $W \subset U \cap V$ and $f|_W = f'|_W$; a member f of an equivalence clan $\tilde{f}$ is called a representative of $\tilde{f}$; let $\mathcal{E}(n,m)$ ($\mathcal{E}(n)$) be the set of germs at zero of $C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ ($C^\infty(\mathbb{R}^n)$); $C^\infty(\mathbb{R}^n)$ is a ring and $A = \{f \in C^\infty(\mathbb{R}^n)$: f vanishes on a nbd. of zero$\}$ is an ideal; by the first isomorphism theorem there exist ϕ such that the following diagram:



commutes and kernel (ϕ)=kernel $(h)/_A$ (H is the natural homomorphism $f \to A+f$); A=kernel(h) and therefore kernel(ϕ) equals the zero subring of $C^\infty(\mathbb{R}^n)/_A$, hence ϕ is injective; since h is surjective $\mathcal{E}(n)$ is isomorphic to $C^\infty(\mathbb{R}^n)/_A$. Similarly if m(n) is the set of germs at zero of $C^\infty(\mathbb{R}^n)$ vanishing at zero, the following diagram commutes



and $E(n)/_{m(n)}$ is isomorphic to $\mathbb{R}$; hence $m(n)$ is a maximal ideal. We have therefore proved:

PR. 8 $E(n)$ is a local $\mathbb{R}$-algebra

PROOF: Need only to show that the unit germ (the germ of the identity map) belongs to the same ideal with $\tilde{f}, f(o)\neq o$; choose a nbd. U of zero such that $f\neq o$ on U and $U'\subset U$ and define $f'=\phi\cdot f+(1-\phi)$, where $\phi\in C^\infty(\mathbb{R}^n)$ is zero outside U, <1 outside U', equals 1 on U' and $\geq o$ everywhere else; $f'>o$ everywhere, $f'\big|_{U'}=f\big|_U$ and $\tilde{f}\cdot \tilde{1}/_{f'}=1$.

PR.9 $m^r(n)$ is a finitely generated $E(n)$-module; in particular is generated by all monomials of degree r.

PR.10 $J^r\equiv E(n)/_{m^{r+1}(n)}$ is a local $\mathbb{R}$-algebra with dimension $(n+r)!/_{n!r!}$ as an $\mathbb{R}$-vector space; J^r is isomorphic to $\mathbb{R}[x_1,\ldots,x_n]/(x_1,\ldots,x_n)^{r+1}$, i.e. the quotient of the ring of polynomials in $x_1,\ldots,x_n$, by its ideal generated by $(x_1,\ldots,x_n)$ raised to the power $(r+1)$.

PR.11 (Nakayama Lemma) Let R be a local ring and m its unique maximal ideal; let A be a finitely generated R-module. Then $m.A=A$ implies $A=o$

PROOF: (TH. BRÖCKER, 4.15)

Given A an R-module and B a submodule of A (i.e. an additive subgroup closed under the multiplication by every element of the ring R) an R-module structure can be defined on the quotient $A/_B$ by $r.(\alpha+B)=r.\alpha+B$, $\alpha\in A$, $\forall r\in R$ and $\forall$ coset $\alpha+B$.

PR. 12 Let B, C be R-modules such that $A, B\subset C$; then $A\subset B+m.A$ implies $A\subset B$

PROOF: By the remark above $(B+m.A)/_B = m\cdot({}^{A+B}/_B)$, but by the second isomorphism theorem ${}^{A+B}/_B = {}^{A}/_{A\cap B}$ and by Nakayama ${}^{A}/_{A\cap B} = o$, hence $A = A\cap B$ or $A\subset B$

Finally:

PR. 13 dim ${}^{E}/_{J} < \infty$ is equivalent to dim ${}^{E}/_{m\cdot J} < \infty$

PROOF: Since $m\cdot J\subset J$, ${}^{E}/_{J} \subset {}^{E}/_{m\cdot J}$; conversely from $E = m\oplus \mathbb{E}$ (direct sum) $m\cdot J + J_{\mathbb{E}} = J$ (vector-space sum) and ${}^{m\cdot J+J_{\mathbb{E}}}/_{m\cdot J} = {}^{J}/_{m\cdot J}$ is isomorphic ${}^{J_{\mathbb{E}}}/_{m.J\cap J_{\mathbb{E}}}$; but dim ${}^{J}/_{m\cdot J} =$ dim ${}^{J_{\mathbb{E}}}/_{m\cdot J\cap J_{\mathbb{E}}} \le n$, since dim $J_{\mathbb{E}} \le n$ and dim$(({}^{E}/_{m\cdot J})/(J/m\cdot J)) =$ dim ${}^{E}/_{J} < \infty$ implies dim ${}^{E}/_{m\cdot J} < \infty$.

All the algebraic material used is from: B. Hartley and T.O Hawkes, Rings, Modules and Linear Algebra, Chapman & Hall, 1970.

III.9 An interpretation scheme for the main classification theorem in the theory of singularities

Let $V_x \in C^\infty(V,\mathbb{R})$, V a nbd. of zero in $\mathbb{R}^k$, be a family of potential functions parametrized by the coordinates $x \in U$, U an open nbd. in $\mathbb{R}^n$ (external parameters of the model); when the extremal condition are fixed, at a fixed point $x \in U$, the system, subject to the dynamic described by the potential function V_x, will stay at a minimum of V_x; V_x are generically Morse functions (their minima are isolated and the quadratic form of the second derivatives is nondegenerate, i.e. V_x is 2-determined) at the singular point; however as x varies, one may ask which kinds of singularities can occur, generically, in the family of functions V_x; the local change in V_x around a point $x \in U$ and a singular point $u \in \mathbb{R}^k$ correspond to a deformation of the singularity V_x around x; the versal deformations give a description of all possible deformations; one wants to see those points in the *control space* (i.e. the space of deformation parameters x) which are most significant for the catastrophe i.e. the points where V_x has a singularity of order higher than two; in other words those points where a local extremum disappears; minima of V_x (unfolded) are called *local regimes* and a *process* for a germ V_x (unfolded) is a section s of the bundle $\mathbb{R}^k \times U \to U$ such that $(s(x),x)$ is either a local regime or at infinity; a *regular point* of a process is a point in the subset U where the section is locally defined and continuous on a nbd. of the point, which is equivalent to saying that over a nbd. of the point there exist a homermorphism of the bundle taking the section onto a constant section; a *catastrophe point* is a non-regular point; the *morphology* or the *catastrophe set* is the set of all catastrophe points; a *convention* assigns a process to the unfolded potential function and there are two basic ones to consider: the *Maxwell convention* states that s(x) is a point where V_x has its lowest minimum (as this may be at minus infinity, this convention is best used when V_x has only finite minima; clearly catastrophe points occur when V_x attains an absolute minimum in two places); the *perfect-delay convention* states that the section s will remain continuous for as long as possible, i.e. s(x) will follow a continuous family of minima until these minima disappear. Thus a particular catastrophe determines a germ $f \in m(k)^2$ which has a deformation $F \in m(k+n)$ with the points $D_F \equiv \pi(\Delta_F)$, important candidates for catastrophe points, where $\pi : \mathbb{R}^k \times U \to U$, $\Delta_F \equiv \{(u,x) \in \Sigma_F : d_u^2 F$ is a degenerate$\}$, $\Sigma_F \equiv \{(u,x) \in \mathbb{R}^k \times U : d_u F = o\}$; Σ_F, Δ_F and D_F are also known as the *catastrophe manifold*, the *singularity set* and the *bifurcation set* respectively. (R. THOM)

BIBLIOGRAPHY

PART I

J.C. ABBOTT — Sets, Lattices and Boolean Algebras
Allyn & Bacon, Inc., 1969

A.D. ALEXANDROV — A contribution to chronogeometry
Canad. Jour. Math., Vol. 19 (1967) 1119-28

A.D. ALEXANDROV and V.V. OVCHINNIKOVA — Notes on the foundations of relativity theory(Russian)
Vestnik Leningrad. Univ., Vol. 11 (1953) 95-110

J.L. ALONSO and F.J. YUDURÁIN — On the continuity of causal automorphisms of space-time
Commun. Math. Physc., Vol. 4 (1964) 349-51

R.W. BASS and L. WITTEN — Remark on Comological models
Rev. Mod. Phys., Vol. 29, No. 3 (1957) 452-3

G. BARUCCHI and G. TEPPATI — The causality group
Nuov. Cim., Vol. LII A, No. 1 (1967) 50-61

G. BARUCCHI — The group of timelike permutations
Nuov. Cim., Vol. LV A No. 3, (1968) 385-95

A.J. BRIGINSHAW — Causality and the group structure of spacetime
Inter. Jour. Theor. Phys., Vol. 19, No. 5 (1980) 329-45

R. BUDIC — R.K. SACHS and R. BUDIC

B. CARTER — Causal structure in spacetime
Gen. Rel. Grav., Vol. 1, No. 4 (1971) 349-91

Z. DOMOTOR — Causal models and spacetime geometries
in: Space, time and geometry (ed. P. Suppes)
D. Riedel, 1973

J. DUGUNDJI — Topology
Allyn & Bacon, 1966

R. GEROCH — 1. Spinor structure of space-times in General Relativity I
Jour. Math. Phys., Vol. 9, No. 1 (1968) 1739-44

2. Space-time structure from a global view point
in: General Relativity and Cosomology, Proceedings of International School in Physics 'Enrico Fermi', Course XLVII, (ed. R.K. Sachs), Academy Press, 1971, 71-103

3. Domain of Dependence
Jour. Math. Phys., Vol. 11, No. 2 (1970) 437-9

4. Singularities
in: Relativity, Proceedings of the Midwest Relativity Conference (eds. S. Flicker, M. Carmeli, L. Witten)
Plenum Press, 1970, 257-91

C. GHEORGHE and E. MIHUL — Causal groups of spacetime
Commun Math. Phys., Vol. 14 (1969) 165-70

P. HAVAS — 1. Causality requirements and the theory of Relat.
Synthése, vol. 18 (1968)75-102

2. Relativity and Causality
in: Proceedings of the 1964 International Congress

in Logic, Methodology and Philosophy of Science (ed. Y. Bar-Hillel), North Holland Publ. Co., 1965, 347-62

S.W. HAWKING and G.F.R. ELLIS — The large scale structure of space-time Cambridge Univ. Press, 1973

S.W. HAWKING and R.K. SACHS — Causally continuous spacetimes Commun. Math. Phys., Vol. 35 (1974) 237-96

S.W. HAWKING — 1. S.W. HAWKING and G.F.R. ELLIS

2. Stable and generic properties in Gen. Relat. Gen. Rel. Grav., Vol. 1, No.4 (1971) 398-400

3. The existence of cosmic time functions Proc. Roy. Soc. Lond., Vol. 308A (1968) 433-5

4. S.W. HAWKING and R.K. SACHS

D. HUME — An inquiry concerning human understanding The Library of Liberal Arts, Publ. Bobbs-Merrill, 1955

J.L. KELLEY — General Topology Van Nostrand Co., 1955

E.H. KRONHEIMER and R. PENROSE — On the structure of causal spaces Proc. Cambr. Phil. Soc., Vol. 63 (1967) 481-501

H.M. MACNEILLE — Partially ordered sets Transactions Amer. Math. Soc., Vol. 35 (1937) 416-460

L. MARKUS — Line element fields and Lorentz structures on differentiable manifolds Ann. Math., Vol. 62, No. 3 (1955) 411-7

E. MIHUL — C. GHEORGHE and E. MIHUL

S. NANDA and K. PANDA

1. Compact topologies on Minkowski space
Inter.Jour. Theor. Phys., Vol. 10, No. 3 (1974) 159-63

2. Minkowski space with order topology is simply connected
Inter. Jour. Theor. Phys., Vol. 12, No. 6 (1975) 393-400

S. NANDA

1. Topology for Minkowski space
Jour. Math. Phys., Vol. 12, No.3 (1971) 394-401

2. Weaker versions of Zeeman's conjectures on topologies for Minkowski space
Jour. Math. Phys., Vol. 13, No.1 (1972) 12-15

3. S. NANDA and K. PANDA, 1

4. S. NANDA and K. PANDA, 2

5. A geometrical proof that causality implies the Lorentz group
Math. Proc. Camb. Phil Soc., Vol. 79 (1976) 533-6

6. A - topology for Minkowski space
Jour. Austral. Math. Soc., Vol. 21 B (1979) 53-64

7. A note on maximal and minimal causal spaces
Jour. Austral. Math. Soc., Vol. 21 B (1980) 357-64

R.G. NEWTON — Relativity and the order of cause and effect in time
in: Causality and Physcial Theories, Proceedings of the AIP Conference No. 16 (ed. W.B. Rolnick), New York, 1974, 49-57

K. PANDA — S. NANDA and K. PANDA

R. PENROSE	1. E.H. KRONHEIMER and R. PENROSE
	2. Techniques of differential topology in Relativity Regional Conference Series in Applied Mathematics, No. 7, SIAM, Philadelphia, 1972
O.S. ROTHAUS	Order isomorphisms of cones Proc. Amer. Math. Soc., Vol. 17 (1966) 1284-88
R.K. SACHS	1. S.W. HAWKING and R.K. SACHS
	2. R.K. SACHS and R. BUDIC
R.K. SACHS and R. BUDIC	Causal boundaries for general relativistic spacetimes Jour. Math. Phys., Vol. 15, No. 8 (1974)
H.J. SEIFERT	The causal boundary of spacetimes Gen. Rel. Grav., Vol. 1, No.3 (1971) 247-59
N.E. STEENROD	The topology of fibre bundles Princeton Univ. Press, 1951
G. TEPPATI	An algebraic analogue of Zeeman's theorem Nuov. Cim.,Vol. LIV A(1968) 800-4
G.S. WHISTON	The Zeeman order topology on Minkowski space Inter. Jour. Theor. Phys., Vol. 6, No. 1 (1972) 75-76
G. WILLIAMS	1. The relation closer than in Minkowski space Amer. Jour. Phys., Vol. 41 (1973) 871-3
	2. Fine topologies for Minkowski space Proc. Camb. Phil. Soc., Vol. 76 (1974) 503-9
N.M.J. WOODHOUSE	1. Causal spaces and spacetimes Ph.D. Thesis, London Univ., 1971
	2. The completion of a chronological space Proc. Camb. Phil. Soc., Vol. 76 (1974) 531-44

F.J. YUDURÁIN	J.L. ALONSO and F.J. YUDURÁIN
E.C. ZEEMAN	1. Causality implies the Lorentz group
	Jour. Math. Phys., Vol. 5, No. 4 (1964) 490-3
	2. The topology of Minkowski space
	Topology, Vol. 6 (1967) 161-70

PART II

R. ABRAHAM and J. MARSDEN — Foundations of Mechanics
W. A. Benjamin, 1967

V.I. ARNOL'D — 1. Normal forms for functions near degenerate critical points, the Weyl group of A_k, D_k, E_k and Lagrangian singularities
Funct. Anal. Appl., Vol. 6, No. 4 (1972) 3-25

2. Characteristic class entering in quantization conditions
Funct. Anal. Appl., Vol. 1, No. 1 (1967) 1-13

3. Integrals of rapidly oscillating functions and singularities of projections of Lagrangian manifolds
Funct. Anal. Appl., Vol. 6, No. 3 (1972) 222-4

D.E. BLAIR — Contact manifolds in Riemannian geometry
Lecture Notes in Mathematics, Vol. 509
Springer-Verlag, 1976

TH. BRÖCKER — Differentiable germs and catastrophes
London Mathematical Society, Lecture Notes Series, No. 17
Cambridge Univ. Press, 1975

C. CARATHÉODORY — Calculus of variations and partial differential equations of the first order, Part I.
Holden-Day, 1965

C. CHEVALLEY — Theory of Lie groups
Princeton Univ. Press, 1946

J.J. DUISTERMAAT — Oscillatory integrals, Lagrange immersions and unfoldings of singularities

Comm. Pure Appl. Math., Vol. 27 (1974) 207-81

B. FINZI — Discontinuità sul fronte d'onda delle azioni gravitazionali
Atti della Acad. Naz. Lincei, Vol. 6 (1949) 18-25.

V. FOCK — The theory of space, time and gravitation
Pergamon Press, 1964

M. GOLUBITSKY and V. GUILLEMIN — Stable mappings and their singularities
Springer-Verlag, 1973

J. GUCKENHEIMER — Bifurcation and Catastrophe
in: Dynamical Systems
ed. M.M. Peixoto, 1973, 95-109

L. HORMANDER — Fourier integral operators. I
Acta Math., Vol. 127 (1971) 79-183

S. LANG — Differential manifolds
Addison Wesley Publ., 1972

M. F. LATOUR — Stabilité des champs d'application différentiables; généralisation d'un théorème de J. Mather
C. R. Acad. Sc. Paris, Vol. 268 (1969) 1331-4

V.V. LYCHAGIN — Local classification of non-linear first-order partial differential equations
Russ. Math. Surveys, Vol. 30, No. 1 (1975) 105-75

M.M POSTNIKOV — The variational theory of geodesics
W. B. Saunders Co., 1967

D. RAND — Arnold's classification of simple singularities of smooth functions
Mathematics Institute, preprint
Warwick Univ. 1977

S. SASAKI — A characterization of contact transformations
Tôhoku Math. Jour., Vol. 16 (1964) 285-90

S. STENBERG — Lectures on differential geometry
Prentice Hall, 1964

R. THOM — Structural stability and morphogenesis
W. A. Benjamin, 1975

S. M. VIŠIK — Some transversality theorems
Math. USSR Sbornik, Vol. 14, No. 1 (1971)
140-56

C.T.C. WALL — (ed.) Proceedings of the Liverpool Singularities Symposium I
Lecture Notes in Mathematics, Vol. 192
Springer-Verlag, 1971

G. WASSERMANN — Stability of unfoldings
Lecture Notes in Mathematics, Vol. 393
Springer-Verlag, 1974

A. WEISTEIN — Lectures on symplectic manifolds
Expository Lectures from the CBMS regional conference held at the University of North Carolina, March 8-12, 1976
Amer. Math. Soc., 1977

J. WESS — The conformal invariance in quantum field theory
Nuov. Cim., Vol. 18 (1960) 1086-1107

V.M. ZAKALYUKIN — Lagrangian and Legendrian singularities
Funct. Anal. Appl., Vol. 10, No. 1 (1976) 23-31

V.D. ZAKHAROV — Gravitational waves in Einstein's theory
Halsted Press, 1973