AXIOMATIC CAUSAL THEORY OF SPACE-TIME

Alexander K. Guts

Omsk State University, 55–A Prosp. Mira, Omsk 644077, Russia

Received 26 June 1995

Principles of construction of a causal space-time theory are discussed. A system of axioms for Special Relativity Theory, which postulates the macrocausality and continuity of time order, is considered. The possibilities of a topos-theoretic approach to the foundations of Relativity Theory are investigated.

Construction of a causal theory of space-time is one of the most attractive tasks of science in the 20th century. From the viewpoint of mathematics, partially ordered structures should be considered. The latter is commonly understood as a set V with a specified reflexive and transitive binary relation ≤. A primary notion is actually not that of causality but rather that of motion (interaction) of material objects. Causality is brought to the forefront since an observer detects changes of object motion or state. It is this detection that gives rise to the view of a particular significance of causes and effects for a phenomenon under study, along with the conviction that causal connections are non-symmetric. Causality is treated as such a relation in the material world that plays a key role in explaining the topological, metric and all other world structures.

1. Pre-physical and philosophic premises of causal axiomatics

This section is aimed at revealing the basic principles able to be a basis for an axiomatic building of special relativity, or, more precisely, the four-dimensional pseudo-Euclidean geometry of signature (+ − − −) known in physical theories as the Minkowski space-time.

Definition 1.1. The Minkowski space-time is the four-dimensional pseudo-Euclidean space \( \mathbb{E}^4 \) of signature (+ − − −), whose metric in a certain (orthonormal) coordinate frame \( x^0, x^1, x^2, x^3 \) has the form

\[ ds^2 = dx^0^2 - dx^1^2 - dx^2^2 - dx^3^2. \]  

(1)

The metric (1) is invariant under the 10-parameter group of transformations II, consisting of the groups of transitions \( T \), rotations and pseudo-rotations and calle the Poincaré group.

The group of transitions \( T \) may be identified with the 4-dimensional vector space \( v(\mathbb{E}^4) \), associated with the affine structure of the space \( \mathbb{E}^4 \).

As one can unambiguously put into correspondence to each point \( x \) of the pseudo-Euclidean space the vector \( c \bar{x} \in v(\mathbb{E}^4) \), where \( c \) is a fixed point of \( \mathbb{E}^4 \), the Minkowski space-time can be regarded as the vector space \( T \), being an Abelian topological group (with natural topology).

So we will operate on the premise that the space-time is an Abelian topological group. This is one of the formal premises of our axiomatics of the space-time of special relativity.

It is also necessary to enumerate the pre-physical and philosophic premises. They are inevitable since our goal is an axiomatization of geometry, which maybe would not deserve attention by itself but has become an object of thorough study just due to its connection with one of the most significant physical theories. Actually the fact that it is pseudo-Euclidean geometry rather than relativity that is under study, conceals such an important circumstance as the very fact of geometrization of the relativity theory. The latter has been carried out by Minkowski. Geometrization of a physical theory is essentially a step to its axiomatization since geometry has always been a historical example of perfect formulation of a theory. However, axiomatization is not only a matter of system nature and strictness; it is as well a tool for maintaining clearness and purity. Making a physical theory axiomatic, we reveal its essence. And here the philosophic and pre-physical premises of an axiomatic theory are of great importance. They help one to correctly choose the primary notions and axioms.

The axiomatics to be presented below is based on the space-time conception developed in the papers by Minkowski [1], Robb [2] and A.D. Alexandrov [3].

A starting image of space-time as a world or manifold of elementary (atomic) events. An elementary event is a phenomenon whose extension in both space and time may be neglected. It is assumed that all phenomena consist of elementary events. An event is like a point in Euclidean geometry: it is indivisible, or primary. Such an approach allows us to repeat Euclid’s way and to arrive at a geometric theory of space-time.

The manifold of events should represent the ma-
terial world around us. The matter exists in no way than in motion that manifests itself in bodies’ influence upon each other. So events also affect each other. Attempting to follow the process of influence, we simplify the interaction picture, concentrate out attention on changing states and thus distinguish causes and effects. So the world appears before our eyes as a full set of most diverse cause-and-effect connections among events.

Thus, if one treats the world in its space and time manifestations as the full set of interacting phenomena, one can try to present the space-time as a set of elementary events $\mathcal{M}$, connected by cause-and-effect relations. In other words, the space-time structure of the world is nothing more than its cause-and-effect structure taken in a proper abstractness [3]. Consequently, the metric structure of space-time, appearing in Definition 1.1, is secondary and “derivable” from its causal structure.

We have formulated the main ideas of a causal theory of space-time. The cause-and-effect relation is assumed to be non-symmetric, thus reflecting the observed essence of time. Therefore, in its formalization, the causal relation is modelled by a (partial) ordering relation.

Evidently, the existence of a causal connection between the events $x$ and $y$ says nothing on how it has formed, i.e., in which way $x$ affected $y$. Were there any intermediate events, where the action of $x$ upon $y$ “rolled over”? It would be closer to tradition to believe that there were. At any rate, the classical presentation of special relativity agrees with such a view (the influence domain $P_x$ of the event $x$, i.e., the set of events subject or able to be subject to an influence of $x$, forms the so-called future light cone). But is such an assumption necessary for building the relativity theory? Is it so necessary to think that $P_x$ is a set for which $x$ is a limit point?

It can be asserted that the true core of relativity theory is the Poincaré group for which the metric (1) is just an invariant, in the same way as one of its invariants is the system of influence domains $\{P_x : x \in \mathcal{M}\}$. On the other hand, if it is possible to obtain the Poincaré group itself as a group whose invariant is the postulated system of influence domains $\{P_x : x \in \mathcal{M}\}$, which is not a priori necessarily a system of (elliptic) null cones, then we will obtain as consequences both the metric (1) and the light cones. Moreover, if therewith $x$ is not a limit point for $P_x$, that means that the theory does not assume any intermediate events in the transition of influences in sufficiently small domains of the world of events (space-time). In other words, an action is transferred by jumps and events which are close enough to each other in space and time, are not connected by any cause-and-effect relations. This does not exclude the possibility of their interaction but means that this interaction cannot be analyzed in terms of cause and effect, due to their “roughness”, inability to “see” more subtle details of interactions taking place at small distances and within small time intervals.

Is it possible to build such an axiomatic theory, “rejecting” microcausality? Yes, it is. The role of an influence domain is played by the set

$$P_x = \left\{(y^0, y^1, y^2, y^3) : (y^0 - x^0)^2 - \sum_{i=1}^{3} (y^i - x^i)^2 \geq l_0^2 \& y^0 \geq x^0 \right\} \cup \{x\},$$

where $l_0$ is a certain fixed number determining the spatial scale of domains deprived of causal connections (the corresponding temporal scale is $l_0/c$ where $c$ is the speed of light).

Postulating the existence of a transformation group leaving invariant the influence domains system, one requires in essence a sort of “likeness” in the action of the cause-and-effect mechanism in different part of the space-time. A logical completion of such a statement is the requirement of space-time homogeneity under the action of the above group. Formally, this is equivalent (according to what was said in the beginning of the section about the group of translations $T$) to the assumption that the set $\mathcal{M}$ has a(n) (algebraic) group structure. The group structure, corresponding to the translation group $T$, is Abelian, and this should be reflected in the axiomatics. It would be perfect to obtain the Abelian nature of the group structure as a consequence of the space-time causal structure. It is possible in principle, but at the expense of losing the “self-evidence” of the corresponding conditions to be imposed on the causal structure. In the following Section 2 it is shown that the commutativity follows from the causal structure but not to the desired extent.

One more important circumstance is to be noted. Apart from the world homogeneity conditions, it is possible to impose conditions connected with symmetries with respect to interaction propagation in the space-time. These are actually conditions on the nature of cause-and-effect relations. Some studies have shown that the causal relation homogeneity conditions and the homogeneity of the space-time itself are connected with each other by means of restrictions on the left-invariant affine structure admitted by a homogeneous world of events [4, 5].

2. Axiomatics based on the macrocausality assumption

Thus, in building an axiomatic theory of space-time we will take into consideration the causal structure, the group structure (space-time homogeneity) and the homogeneity of causal relations.

The primary notions are: the set $\mathcal{M}$ (the manifold of events, or the space-time); elements $x, y, \ldots$ of the set $\mathcal{M}$ (events); a system of subsets $\mathcal{P} = \{P_x : x \in \mathcal{M}\}$ of the set $\mathcal{M}$ (influence domains); a group of transformations $Aut(\mathcal{P})$, consisting of the bijections $f : \mathcal{M} \rightarrow \mathcal{M}$ such that $f(P_x) = P_{f(x)}$ for each $x \in \mathcal{M}$.
The axiomatics meant is consistent rather than formal in the understanding of Hilbert and Bernays [6, c.24]. The latter emerge as a result of abandoning a concrete content and are formulated in an existential form.

One adds to the primary notions the set of real numbers \( \mathbb{R} \). It is a very strong assumption, which is in general not so necessary [7]. Although real numbers are defined using a large number of axioms, all of them are created by mathematical intelligence and therefore cannot interfere with our main goal in the axiomatization of relativity, namely, to reveal a minimal set of philosophical and pre-physical premises which act as a basis of the postulated statements (axioms).

Let us enumerate the axioms of (disconnected) causal space-time theory, dividing them into groups.

Pre-order axiom.

A1. The system \( \mathcal{P} = \{ P_x : x \in \mathcal{M} \} \) specifies on \( \mathcal{M} \) a partial ordering, i.e.,

\[
x \in P_x;
y \in P_x \Rightarrow P_y \subset P_x.
\]

Recall that an ordering is obtained if a pre-order satisfies the additional condition (the relation non-symmetric condition)

\[
x \neq y \Rightarrow P_x \neq P_y.
\]

Consequently, Axiom A1 does not require that actions between the events \( x \) and \( y \) be non-symmetric. The asymmetry follows from Axiom A6 below (Proposition 2.2 from [10]).

Introduce the notations

\[
Q_x = P_x \setminus \{ x \}, \quad Q_x^- = P_x^- \setminus \{ x \},
\]

where \( P_x^- = \{ y \in \mathcal{M} : x \in P_y \} \).

Topological and group structure axiom

Consider \( \mathcal{B} \), the family of all subsets of the form \( Q_x \cap Q_y^- \), where \( y \in P_x \).

Let \( \mathcal{T}_\preceq \) be a topology on \( \mathcal{M} \) with a pre-base \( \mathcal{B} \).

Then

A2. \( \langle \mathcal{M}, \mathcal{T}_\preceq \rangle \) is a connected, Hausdorff, locally compact 4-dimensional topological group.

It should be noted that it is unnecessary to postulate a finite dimensionality: it follows from the axioms below. Thus Axiom A2 just makes certain the number of dimensions.

Group structure commutativity axioms

Definition 2.1. A subsemigroup \( H \) is called maximal if from \( H \subset L \subset \mathcal{M} \), where \( L \) is a subsemigroup, it follows \( L = \mathcal{M} \), and is called normal if for each \( x \in \mathcal{M} \) one has \( xH = Hx \).

Definition 2.2. A set \( K \subset \mathcal{M} \) is called convex if it can be presented in the form

\[
K = \bigcap_{\alpha \in A} x_\alpha H_\alpha,
\]

where \( x_\alpha \in \mathcal{M}, H_\alpha \) is a maximal subsemigroup and \( A \) is a certain set of indices.

A3. There exists a closed maximal normal proper subsemigroup.

A4. The topology \( \langle \mathcal{M}, \mathcal{T}_\preceq \rangle \) is convex, i.e., there is a base of neighbourhoods of the group unity consisting of neighbourhoods with closures which are convex sets.

Axiom A3 is independent of the causal structure, it is a property of the group structure. However, the two structures are connected by Axiom A4 as well as by the following group of axioms.

Axiom of connection between causal and group structures

A5. The pre-order \( \preceq \) is bi-invariant, i.e., \( xP_e = P_x x = P_x \) for each \( x \in \mathcal{M} \) where \( e \) is the unity of the group \( \mathcal{M} \).

It is clear that \( P_x^- = P_x^{-1} \) and that \( P_x \) is a normal subsemigroup of the group \( \mathcal{M} \).

Axioms of connection between order and topology

A6. For each \( y \in P_x, x \in \mathcal{M} \), the set \( \overline{P_x} \cap \overline{P_y} \) is compact.

Here \( \overline{A} \) denotes the closure of the set \( A \) with respect to the topology \( \mathcal{T}_\preceq \).

A7. \( x \notin \overline{Q_x}, Q_x \neq \emptyset \).

Space isotropy axiom.

A8. The stabilizer Aut(\( \mathcal{P} \)) at the point \( x \) acts transitively on the border \( \partial Q_x \) of the set \( Q_x \).

This axiom characterizes the degree of homogeneity of causal connections.

Let

\[
\sigma(\overline{P_x}) = \bigcap_{x \in \overline{Q_y}} Q_y.
\]

The family \( \sigma(\mathcal{P}) = \{ \sigma(\overline{P_x}) : x \in \mathcal{M} \} \), as it is easily verified, specifies a pre-order on \( \mathcal{M} \). Although this pre-order is created by the causal structure, it has another meaning. The set \( \sigma(\overline{P_x}) \) consists of events which are causally created (or might be causally created) by all those events which are the cause of the event \( x \). If \( y \in \sigma(\overline{P_x}) \), then \( y \) is not necessarily a consequence of \( x \), although this is the case for the events from \( P_x \subset \sigma(\overline{P_x}) \). The events which entered into \( \sigma(\overline{P_x}) \) are nothing more than events occurring later than \( x \). In
other words, the pre-order $\sigma(\mathcal{P})$ is the temporal order in the space-time. In this axiomatics it is determined by the causal order; it worth noting, however, that the relation between the causal order and the temporal one is a subject of numerous philosophic arguments.

One can now consider continuous mappings of the set of nonnegative real numbers $\mathbb{R}_+$ into $\mathcal{M}$ which are monotone with respect to the natural order in $\mathbb{R}$ and the pre-order $\sigma(\mathcal{P})$, i.e., $\text{Rusc}(\mathcal{P})$ be the set of all the above mappings $\mathbb{R}_+ \rightarrow \mathcal{M}$, continuous and monotone with respect to $\sigma(\mathcal{P})$, such that 0 is mapped into the unity $e$ of the group $\mathcal{M}$. Such a familiar property of the temporal order as continuity is postulated with the aid of the axiom

Axiom of continuity of the temporal order.

A9. $P_e \subset \text{Rusc}(\mathcal{P})$.

Finally, the following axiom excludes pseudo-Finslerian geometries:

A10. The set $\sigma(\mathcal{P})$ is not a direct product of subsemigroups.

One of the results of the above axiomatics is

Theorem [8]. Let the set of Axioms A1–A10 be valid. Then $\mathcal{M}, \mathcal{T}_\sigma >$ may be equipped with such a pseudo-Euclidean structure $\mathcal{E}^4$ that in some orthonormal coordinates $x^0, x^1, x^2, x^3$ we will have:

$$\begin{align*}
\partial Q_\sigma &= \left\{ y \in \mathbb{E}^4 : (y^0 - x^0)^2 - 3 \sum_{i=1}^3 (y^i - x^i)^2 = l_0^2, \\
& \quad l_0 = \text{const} \neq 0 \& y^0 > x^0 \right\}; \\
\sigma(\mathcal{P}_\sigma) &= \left\{ y \in \mathbb{E}^4 : \\
& \quad (y^0 - x^0)^2 \geq 3 \sum_{i=1}^3 (y^i - x^i)^2 \& y^0 \geq x^0 \right\};
\end{align*}$$

and lastly, $\text{Aut}(\mathcal{P})$ is the Poincaré group.

Proof. As follows from Axioms A2–A4 and Theorem 2 from [9], the group $\mathcal{M}$ is isomorphic to the four-dimensional vector space $\mathbb{V}^4$ with natural topology, that is an Abelian group. Therefore $\mathbb{V}^4$ can be equipped with an affine structure $\mathcal{A}$, whose associated vector space is $\mathbb{V}^4$. Then $\mathcal{M}$ is the affine space $\mathcal{A}$, where the so-called unconnected pre-order $\mathcal{P}$ (Axiom A7), invariant with respect to parallel transport (Axiom A5), is defined. By A9, the pre-order $\sigma(\mathcal{P})$ is connected, i.e., $x \in \sigma(\mathcal{P}_\sigma) \setminus \{x\}$. Consequently, due to Axiom A6, Lemma 2.3 and Proposition 1.2 from [10], the pre-order $\sigma(\mathcal{P})$ satisfies the conditions of Theorem A from [10]. Therefore, $\text{Rusc}(\mathcal{P}_\sigma)$ is a convex cone with a vertex $x$, called contingency [7].

Then by Proposition 3.3 [10] (the interior $\text{int}(\mathcal{P}_\sigma) \neq \emptyset$ due to A7 and Theorem 4.1 [10]) the order $\mathcal{P}$ is maximally lined, hence by Theorem 4.2 [10], $\text{Aut}(\mathcal{P})$ consists of affine transformations of the space $\mathcal{V}^4$, while $\sigma(\mathcal{P}_\sigma)$ is a convex closed cone with the vertex $x$, coinciding with the closure of the convex hull of the set $\mathcal{P}_\sigma$. Moreover, the cone $C = \text{int}(\sigma(\mathcal{P}_\sigma)) \cup \{e\}$, where $\text{int}(\mathcal{A})$ is the interior of the set $\mathcal{A}$, is homogeneous. As known, in this case the cone $C$ is one of the following three: $L_1 \times L_2 \times L_3 \times L_4$, $L \times K^3$, $K^4$, where $K^i$ is an $i$-dimensional elliptic cone with the vertex $e$. The first two cases are excluded by Axiom A10.

Thus $\sigma(\mathcal{P}_\sigma)$ is an elliptic cone (the light cone). One can (Theorem 4.2 from [10], Theorem 5.2 and a conclusion from it on page 56 from [8]) introduce the affine coordinates $x^0, x^1, x^2, x^3$ with the origin $e$ so that the cone $\sigma(\mathcal{P}_\sigma)$ be specified by the inequalities

$$(y^0 - x^0)^2 \geq 3 \sum_{i=1}^3 (y^i - x^i)^2 \& y^0 \geq x^0,$$

the set

$$\partial Q_x = \left\{ y \in \mathbb{V}^4 : (y^0 - x^0)^2 - 3 \sum_{i=1}^3 (y^i - x^i)^2 = l_0^2, \\
& \quad l_0 = \text{const} \neq 0 \& y^0 > x^0 \right\},$$

and the group $\text{Aut}(\mathcal{P})$ is the Poincaré group.

The scalar product in these coordinates has the form

$$\langle u, v \rangle = u^0 v^0 - 3 \sum_{i=1}^3 u^i v^i,$$

where $u$ and $v$ are vectors of the vector space associated with $\mathbb{V}^4$.

Thus $\mathbb{V}^4$ is equipped with the pseudo-Euclidean structure $\mathcal{E}^4$ with the signature $< + - - - >$.

The theorem is proved. $ullet$

Thus the space-time geometry $\mathcal{E}^4$ is an agreement of the causal and group structures, under the condition of temporal order continuity (Axiom A9). The presence of a group structure is obligatory as long as geometry is concerned. This conforms with the spirit of Felix Klein’s Erlangen programme.

It has been repeatedly claimed that the pseudo-Euclidean geometry follows from axioms like A1 - A8, A10 (or their analogues), i.e., it is unnecessary to postulate the microcausality and temporal order continuity. However, up to now a proof of this statement, satisfactory in all respects, did not appear.

The time continuity requirement can be replaced by the causal order continuity condition, that means a denial of the requirement $x \notin \partial Q_x$ in Axiom A7. The corresponding axiomatics is presented in [11]. In the case under consideration, that of microcausality denial, one can avoid the time and causality continuity requirement, but only at the expense of admitting an affine structure of the world of events [7].

3. A transition to topos theory. New possibilities

In Section 2 we interpreted the world of events as a set. This means that the mathematical modelling of
the physical space-time was based on set theory. This theory has been in the 20th century not only the language used by the mathematicians to formulate and realize their ideas, but also in essence their ideology. Evidently Nature is not forced to be confined in the set-theoretic ideological frames of mathematical abstractions. A transition beyond the frames of set theory brings new possibilities for describing the real space-time properties [12]. An event was so far treated as an indivisible phenomenon. This is, however, an evident simplification. Time loops, appearing in general relativity, clearly demonstrate the deficiency of such an approach. A theory should admit the possibility of automatically complicating the elementary (atomic) event structure depending on situation. The structure of (causal) interaction of events should herewith accordingly complicate, as well as the space-time topological and metric structures.

Ideally, it would be necessary to have such a formal theory of space-time that would be able acquire a most sudden appearance relevant to a concrete model. This approach may be exemplified by obtaining in [12] from the same set of axioms, only at the expense of model (topos) choice, either the flat Minkowski space-time, or by the scheme used in [12]. However, the resulting space-time theory will be non-classical, different from that of Minkowski space-time. This is a new theory of space-time, created in a purely logical manner. It will reflect the real space-time properties to the same extent as the development of mathematical abstractions accompanies the development of the real world.

References