5-1-2010

Metrization Theorem for Space-Times: From Urysohn's Problem Towards Physically Useful Constructive Mathematics

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Metrization Theorem for Space-Times: From Urysohn's Problem Towards Physically Useful Constructive Mathematics

Vladik Kreinovich

Department of Computer Science, University of Texas, El Paso, `vladik@utep.edu`

To Yuri Gurevich, in honor of his enthusiastic longtime quest for efficiency and constructivity.

Abstract. In the early 1920s, Pavel Urysohn proved his famous lemma (sometimes referred to as "first non-trivial result of point set topology"). Among other applications, this lemma was instrumental in proving that under reasonable conditions, every topological space can be metrized.

A few years before that, in 1919, a complex mathematical theory was experimentally proven to be extremely useful in the description of real world phenomena: namely, during a solar eclipse, General Relativity theory – that uses pseudo-Riemann spaces to describe space-time – was (spectacularly) experimentally confirmed. Motivated by this success, Urysohn started working on an extension of his lemma and of the metrization theorem to (causality-)ordered topological spaces and corresponding pseudo-metrics. After Urysohn's early death in 1924, this activity was continued in Russia by his student Vadim Efremovich, Efremovich's student Revolt Pimenov, and by Pimenov's students (and also by H. Busemann in the US and by E. Kronheimer and R. Penrose in the UK). By the 1970s, reasonably general space-time versions of Urysohn's lemma and metrization theorem have been proven.

However, these 1970s results are not constructive. Since one of the main objectives of this activity is to come up with useful applications to physics, we definitely need constructive versions of these theorems – versions in which we not only claim the theoretical existence of a pseudo-metric, but we also provide an algorithm enabling the physicist to generate such a metric based on empirical data about the causality relation. An additional difficulty here is that for this algorithm to be useful, we need a physically relevant constructive description of a causality-type ordering relation.

In this paper, we propose such a description and show that, for this description, a combination of the existing constructive ideas with the known (non-constructive) proof leads to successful constructive space-time versions of the Urysohn's lemma and of the metrization theorem.

Key words: Urysohn's lemma, metrization theorem, space-times, constructive mathematics

1 Introduction

Urysohn's lemma. In the early 1920s, Pavel Urysohn proved his famous lemma (sometimes referred to as "first non-trivial result of point set topology"). This lemma deals with *normal* topological spaces, i.e., spaces in which every two disjoint closed sets have disjoint open neighborhoods; see, e.g., [12]. As the very term "normal" indicates, most usual topological spaces are normal, including the n-dimensional Euclidean space.

Urysohn's lemma states the following:

Lemma 1. *If X is a normal topological space, and A and B are disjoint closed sets in X, then there exists a continuous function $f : X \to [0,1]$ for which $f(a) = 0$ for all $a \in A$ and $f(b) = 1$ for all $b \in B$.*

Resulting metrization theorem. Urysohn's lemma has many interesting applications. Among other applications, this lemma was instrumental in proving that under reasonable conditions, every topological space X can be metrized, i.e., there exist a *metric* – a function $\rho : X \times X \to \mathbb{R}_0^+$ to the set $\mathbb{R}_0^+$ of all non-negative real numbers for which the following three conditions are satisfied

$$\rho(a,b) = 0 \Leftrightarrow a = b; \quad \rho(a,b) = \rho(b,a); \quad \rho(a,c) \le \rho(a,b) + \rho(b,c);$$

and for which the original topology on X coincides with the topology generated by the open balls $B_r(x) = \{y : \rho(x,y) < r\}$.

Specifically, from Urysohn's lemma, we can easily conclude that:

Theorem 1. *Every normal space X with countable base is metrizable.*

Comment. It is worth mentioning that the normality condition is too strong for the theorem: actually, it is sufficient to require that the space is:

- *regular*, i.e., for every closed set A and every point $b \notin A$ can be separated by disjoint open neighborhoods, and
- *Hausdorff*, i.e., every two different points have disjoint open neighborhoods.

Space-time geometry and how it inspired Urysohn. A few years before Urysohn's lemma, in 1919, a complex mathematical theory was experimentally proven to be extremely useful in the description of real world phenomena. Specifically, during a solar eclipse, General Relativity theory – that uses pseudo-Riemann spaces to describe space-time – was (spectacularly) experimentally confirmed; see, e.g., [20].

From the mathematical viewpoint, the basic structure behind space-time geometry is not simply a topological space, but a topological space with an *order* $a \preceq b$ whose physical meaning is that the event a can causally influence the event b.

For example, in the simplest case of the Special Relativity theory (see Fig. 1), the event $a = (a_0, a_1, a_2, a_3)$ can influence the event $b = (b_0, b_1, b_2, b_3)$ if we can

get from the spatial point (a_1, a_2, a_3) at the moment a_0 to the point (b_1, b_2, b_3) at the moment $b_0 > a_0$ while traveling with a speed which is smaller than or equal to the speed of light c:

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2} \leq c \cdot (b_0 - a_0).$$
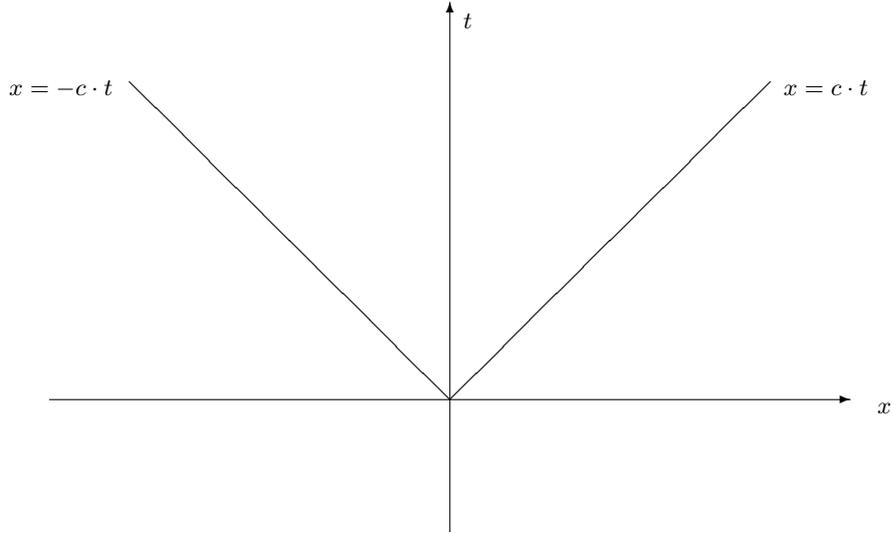

Fig. 1. Causality relation of the Special Relativity theory

Motivated by this practical usefulness of ordered topological spaces, Urysohn started working on an extension of his lemma and of the metrization theorem to (causality-)ordered topological spaces and corresponding pseudo-metrics.

Space-time metrization after Urysohn. P. S. Urysohn did not have time to work on the space-time extension of his results, since he died in 1924 at an early age of 26.

After Urysohn's early death, this activity was continued in Russia by his student Vadim Efremovich, by Efremovich's student Revolt Pimenov, and by Pimenov's students – and also by H. Busemann in the US and by E. Kronheimer and R. Penrose in the UK [10, 13, 22] (see also [16]).

This research actively used the general theory of ordered topological spaces; see, e.g., [21].

By the 1970s, reasonably general space-time versions of Urysohn's lemma and metrization theorem have been proven; see, e.g., [14, 15].

Space-time metrization results: main challenge. One of the main objectives of the space-time metrization activity is to come up with useful applications to physics.

From this viewpoint, we definitely need *constructive* versions of these theorems – versions in which we not only claim the theoretical existence of a (pseudo)metric, but we also provide an algorithm enabling a physicist to generate such a metric based on empirical data about the causality relation.

The original 1970s space-time metrization results are not constructive. It is therefore necessary to make them constructive.

An additional difficulty here is that for this algorithm to be useful, we need a physically relevant constructive description of a causality-type ordering relation.

What we do in this paper. In this paper,

- we propose a physically relevant constructive description of a causality-type ordering relation, and
- we show that for this description, a combination of the existing constructive ideas with the known (non-constructive) proof leads to successful constructive space-time versions of the Urysohn's lemma and of the metrization theorem.

2 Known Space-Time Metrization Results: Reminder

Causality relation: the original description. The current formalization of space-time geometry starts with a transitive relation $a \preceq b$ on a topological space X.

The physical meaning of this relation is causality – that an event a can influence the event b. This meaning explains transitivity requirement: if a can influence b and b can influence c, this means that a can therefore (indirectly) influence the event c.

Need for a more practice-oriented definition. On the theoretical level, the causality relation $\preceq$ is all we need to know about the geometry of space-time.

However, from the practical viewpoint, we face an additional problem – that measurements are never 100% accurate and, therefore, we cannot locate events exactly. When we are trying to locate an event a in space and time, then, due to measurement uncertainty, the resulting location $\widetilde{a}$ is only approximately equal to the actual one: $\widetilde{a} \approx a$.

From this viewpoint, when we observe that an event a influences the event b, we record it as a relation between the corresponding approximations – i.e., we conclude that $\widetilde{a} \preceq \widetilde{b}$; see Fig. 2. However, this may be a wrong conclusion: for example, if an event b is at the border of the future cone $F_a \overset{\text{def}}{=} \{b : a \preceq b\}$ of the event a, then

- we have $a \preceq b$, but
- the approximate location $\widetilde{b}$ may be outside the cone,

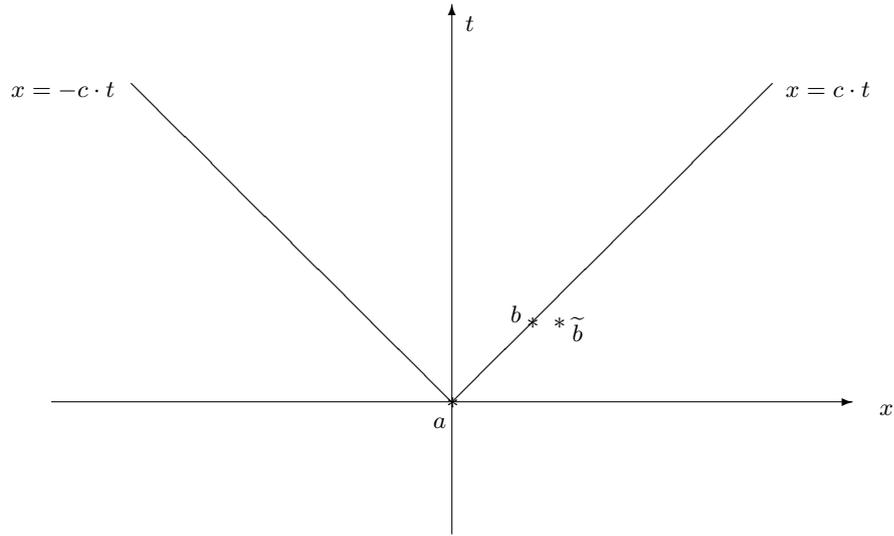so the conclusion $a \preceq \widetilde{b}$ is wrong.

Fig. 2. Need for a more practice-oriented definition of causality

Kinematic causality: a practice-oriented causality relation. To take into account measurement uncertainty, researchers use a different causality relation $a \prec b$, meaning that every event in some small neighborhood of b causally follows a.

Comment. Since a can only be measured with uncertainty, it is also reasonable to consider a more complex relation: every event in some small neighborhood of b causally follows every element from some small neighborhood of a. Under certain reasonable conditions, however, this more complex definition is equivalent to the above simpler one. Thus, in the following text, we will consider the above simpler definition.

In precise terms, the above definition $a \prec b$ means that b belongs to the interior $\text{Int}(F_a)$ of the future cone F_a.

In the simplest space-time of special relativity, this means that we are excluding the border of the future cones (that corresponds to influencing by photons and other particles traveling at a speed of light c) and only allow causality by particle whose speed is smaller than c. The motion of such particles is known as *kinematics*, hence this new practice-oriented causality relation is called *kinematic causality*.

This definition implies, e.g., that the kinematic casuality relation is transitive, as well as several other reasonable properties. These properties lead to the following formal definition of the kinematic causality relation.

Definition 1. *A relation $\prec$ is called a* kinematic causality *if it is transitive and satisfies the following properties:*

$$a \not\prec a; \quad \forall a \,\exists \underline{a}, \overline{a}\, (\underline{a} \prec a \prec \overline{a}); \quad a \prec b \Rightarrow \exists c \,(a \prec c \prec b);$$

$$a \prec b, c \Rightarrow \exists d \,(a \prec d \prec b, c); \quad b, c \prec a \Rightarrow \exists d \,(b, c \prec d \prec a).$$

Topology and causality can be defined in terms of kinematic causality relation. We started our description with a pre-ordered topological space X with a causality relation $\preceq$. Based on topology and on the causality relation, we defined the kinematic causality $\prec$.

It turns out that in many reasonable cases, it is sufficient to know the kinematic causality relation $\prec$. Based on this relation, we can uniquely reconstruct both the topology and and the original causality relation $\preceq$.

Indeed, as a topology, we can take a so-called *Alexandrov topology* in which intervals

$$(a, b) \overset{\text{def}}{=} \{c : a \prec c \prec b\}$$

form the base.

Once the topology is defined, we can now describe causality as

$$a \preceq b \overset{\text{def}}{=} b \in \overline{a^+},$$

where $a^+ \overset{\text{def}}{=} \{b : a \prec b\}$ and $\overline{S}$ denotes the closure of the set S.

Comment. In principle, we can use a dual definition $a \preceq b \overset{\text{def}}{=} a \in \overline{b^-}$, where $b^- \overset{\text{def}}{=} \{c : c \prec b\}$. To make sure that these two definitions lead to the same result, the following additional property is usually required:

$$b \in \overline{a^+} \Leftrightarrow a \in \overline{b^-}.$$

Towards a space-time analog of a metric. Traditional metric is defined as a function $\rho : X \times X \to \mathbb{R}_0^+$ to the set to the set $\mathbb{R}_0^+$ of all non-negative real numbers for which the following properties are satisfied:

$$\rho(a, b) = 0 \Leftrightarrow a = b; \quad \rho(a, b) = \rho(b, a); \quad \rho(a, c) \le \rho(a, b) + \rho(b, c).$$

The usual physical meaning of this definition is that $\rho(a, b)$ is the length of the shortest path between a and b. This meaning leads to a natural explanation for the triangle inequality $\rho(a, c) \le \rho(a, b) + \rho(b, c)$. Indeed, the shortest path from a to b (of length $\rho(a, b)$) can be combined with the shortest path from b to c (of length $\rho(b, c)$) into a single combined path from a to c of length $\rho(a, b) + \rho(b, c)$. Thus, the length $\rho(a, c)$ of the shortest possible path between a and c must be smaller than or equal to this combined length: $\rho(a, c) \le \rho(a, b) + \rho(b, c)$.

In space-time, we do not directly measure distances and lengths. The only thing we directly measure is (proper) time along a path. So, in space-time geometry, we talk about times and not lengths.

It is well known that if we travel with a speed close to the speed of light, then the proper travel time (i.e., the time measured by a clock that travels with us) goes to 0. Thus, in space-time, the smallest time does not make sense: it is always 0. What makes sense is the *largest* time. In view of this, we can define a "kinematic metric" $\tau(a, b)$ as the longest (= proper) time along any path from event a to event b.

Of course, such a path is only possible if a kinematically precedes b, i.e., if $a \prec b$.

If $a \prec b$ and $b \prec c$, then the longest path from a to b (of length $\tau(a, b)$) can be combined with the longest path from b to c (of length $\tau(b, c)$) into a single combined path from a to c of length $\tau(a, b) + \tau(b, c)$. Thus, the length $\tau(a, c)$ of the longest possible path between a and c must be larger than or equal to this combined length: $\tau(a, c) \geq \tau(a, b) + \tau(b, c)$. This inequality is sometimes called the *anti-triangle inequality*.

These two properties constitute a formal definition of a kinematic metric.

Definition 2. *By a* kinematic metric *on a set X with a kinematic causality relation $\prec$, we mean a function $\tau : X \times X \to \mathbb{R}_0^+$ to the set $\mathbb{R}_0^+$ of all non-negative real numbers that satisfies the following two properties:*

$$\tau(a, b) > 0 \Leftrightarrow a \prec b; \quad a \prec b \prec c \Rightarrow \tau(a, c) \geq \tau(a, b) + \tau(b, c).$$

Space-time analog of Urysohn's lemma. The main condition under which the space-time analog of Urysohn's lemma is proven is that the space X is *separable*, i.e., there exists a countable dense set $\{x_1, x_2, \ldots, x_n, \ldots\}$.

Lemma 2. *If $\prec$ is a kinematic causality relation on a separable space X, and $a \prec b$, then there exists a continuous $\preceq$-monotonic function $f_{(a,b)} : X \to [0, 1]$ for which:*

- $f_{(a,b)}(x) = 0$ *for all x for which $a \not\prec x$, and*
- $f_{(a,b)}(x) = 1$ *for all x for which $b \preceq x$.*

This lemma is *similar* to the original Urysohn's lemma, because it proves the existence of a function $f_{(a,b)}$ that separates two disjoint closed sets:

- the complement $-a^+$ to the set a^+ and
- the set $\overline{b^+}$.

The new statement is *different* from the original Urysohn's lemma, because:

- first, it only considers special closed sets, and
- second, in contrast to the original Urysohn's lemma, the new lemma also requires that the separating function f be monotonic.

Space-time analog of the metrization theorem. Based on the space-time analog of the Urysohn lemma, one can prove the following results:

Theorem 2. *If X is a separable topological space with a kinematic causality relation $\prec$, then there exists a continuous kinematic metric τ which generates the corresponding kinematic causality relation $\prec$ — in the sense that $a \prec b \Leftrightarrow \tau(a, b) > 0$.*

Comment. Since, as we have mentioned, the kinematic causality relation $\prec$ also generates the topology, we can conclude that the kinematic metric τ also determines the corresponding topology.

Proof. First, we prove that for every x, there exists a $\preceq$-monotonic function $f_x : X \to [0,1]$ for which $f_x(b) > 0 \Leftrightarrow x \prec b$.

The proof of this statement is reasonably straightforward.

The only technically cumbersome part of this proof is to show that if a space with a kinematic causality is separable, i.e., if there exists an everywhere dense sequences $\{x_1, \dots, x_n, \dots\}$, then there exists a decreasing sequence y_i that converges to x. Moreover, we can select this sequence in such a way that for every i, if $x \prec x_i$ then $y_i \prec x_i$.

Indeed, since the relation $\prec$ is a kinematic causality, there exists a point $\overline{x}$ for which $x \prec \overline{x}$. We then take $y_0 \overset{\text{def}}{=} \overline{x}$. By our choice of y_0, we thus have $x \prec y_0$.

Let us assume that we have already selected points $y_0, \dots, y_k$ for which $x \prec y_k \prec y_{k-1} \prec \dots \prec y_0$. Let us construct a point y_{k+1} for which, first, $x \prec y_{k+1} \prec y_k$ and, second, if $x \prec x_{k+1}$, then $y_{k+1} \prec x_{k+1}$.

If $x \not\prec x_{k+1}$, then, due to the properties of the kinematic causality, there exists a point c for which $x \prec c \prec y_k$. We will then take $y_{k+1} = c$.

If $x \prec x_{k+1}$, then, due to the properties of the kinematic causality, from $x \prec x_{k+1}$ and $x \prec y_k$, we can conclude that there exists a point d for which $x \prec d \prec x_{k+1}, y_k$. We can then take $y_{k+1} = d$.

Let us now prove that $y_n \to x$, i.e., that for every open neighborhood U of the point x, there exists an index n_0 for which $y_n \in U$ for all $n \geq n_0$. Indeed, let U be such a neighborhood. Since open intervals form a base, there exists an open interval $(a,b) \subseteq U$ that contains the point x. By definition of the interval, $x \in (a,b)$ means that $a \prec x$ and $x \prec b$. By definition of the kinematic causality, there exists a point c for which $x \prec c \prec b$. Thus, the open interval (x,b) is non-empty. Since the sequence $\{x_n\}$ is everywhere dense, it has a point x_{n_0} in this interval, for which $x \prec x_{n_0} \prec b$. By the properties of the sequence y_i, this implies that $x \prec y_{n_0} \prec x_{n_0} \prec b$. Since the sequence $\{y_n\}$ is decreasing, we thus conclude that $x \prec y_n \prec b$ for all $n \geq n_0$. From $a \prec x \prec y_n$, we then deduce that $a \prec y_n$. Hence, $y_n \in (a,b) \subseteq U$ and so, $y_n \in U$ for all $n \geq n_0$. The statement is proven.

Once a decreasing sequence y_i that converges to x is constructed, we can take $f_x(b) = \sum\limits_{i=1}^{\infty} 2^{-i} \cdot f_{(x,y_i)}(b)$.

Next, we prove that for every x, there exists a $\preceq$-decreasing function

$$g_x : X \to [0,1]$$

for which $g_x(a) > 0 \Leftrightarrow a \prec x$. The proof of this second statement is similar to the proof of the first statement.

Once these two auxiliary statements are proven, we can use the countable everywhere dense sequence $\{x_1, x_2, \dots, x_n, \dots\}$ to construct the desired kinematic

metric as

$$\tau(a, b) = \sum_{i=1}^{\infty} 2^{-i} \cdot \min(g_{x_i}(a), f_{x_i}(b)).$$

It is reasonably easy to prove that thus defined function is indeed a kinematic metric. □

3 Towards a Physically Reasonable Constructive Definition of Causality

Need for a constructive definition of causality. As we have mentioned, in order to provide a physically meaningful constructive version of the space-time metrization theorem, we must come up with a physically meaningful constructive definition of causality.

Towards a constructive definition of casuality: analysis of the physical situation. To come up with a physically meaningful constructive definition of causality, let us recall how causality can be physically detected.

In the ideal world, detecting whether an event a is causally related to the event b (i.e., whether $a \preceq b$) is straightforward. We send a signal at event a in all directions and at all possible speeds, and we check whether this signal was detected at b:

 − if this signal is detected at b, we conclude that $a \preceq b$;
 − if this signal is not detected at b, we conclude that $a \npreceq b$.

In practice, we can only locate an event with a certain accuracy. As a result, when we try to detect whether $a \preceq b$, then, instead of two, we now have *three* possible options:

 − if the signal is detected in the entire vicinity of b, then we conclude that $a \prec b$;
 − if no signal is detected in the entire vicinity of b, then we conclude that $a \npreceq b$;
 − in all other cases, we cannot make any conclusion.

As we increase the location accuracy, we can get more and more information about the casuality. In general, if $a \prec b$, this means that the event a affects all the events in some vicinity of b. Thus, when the location inaccuracy is sufficiently small, we will able to detect that $a \prec b$. In other words, $a \prec b$ if and only we can detect this causality for an appropriate (sufficiently high) level of accuracy.

We can describe this situation by saying that we have a sequence of decidable relations $\prec_n$ corresponding to increasing location accuracy, and

$$a \prec b \Leftrightarrow \exists n \, (a \prec_n b).$$

To detect whether $a \prec b$, we repeat the above experiments with increasing accuracy. If in all these experiments, we do not detect the effect of a on b, this

means that a is not in kinematic causality relation with b: $a \not\prec b$. It seems reasonable to argue that if this negative phenomenon does not occur, this means that for some accuracy level n, we will be able to detect the causality. In other words, we require that $\neg(a \not\prec b) \Rightarrow a \prec b$, i.e., that the "Markov principle" $\neg\neg(a \prec b) \Rightarrow a \prec b$ holds for the constructive kinematic causality relation.

As a result, we arrive at the following constructive version of kinematic causality.

Definition 3. *A relation $\prec$ on a set X is called a* constructive kinematic causality *if it satisfies the following properties:*

- *$\prec$ is transitive: $(a \prec b \,\&\, b \prec c) \Rightarrow a \prec c$.*
- *$\prec$ satisfies the formula $\neg\neg(a \prec b) \Rightarrow a \prec b$.*
- *$\prec$ satisfies the following properties:*

$$a \not\prec a; \quad \forall a \, \exists \underline{a}, \overline{a} \, (\underline{a} \prec a \prec \overline{a}); \quad a \prec b \Rightarrow \exists c \, (a \prec c \prec b);$$

$$a \prec b, c \Rightarrow \exists d \, (a \prec d \prec b, c); \quad b, c \prec a \Rightarrow \exists d \, (b, c \prec d \prec a).$$

- *If $a \prec b$, then $\forall c \, (a \prec c \lor b \not\preceq c)$.*
- *There exists a sequence $\{x_i\}$ for which $a \prec b \Rightarrow \exists i \, (a \prec x_i \prec b)$.*
- *There exists a decidable ternary relation $x_i \prec_n x_j$ for which*

$$x_i \prec x_j \Leftrightarrow \exists n \, (x_i \prec_n x_j).$$

A set X with a constructive causality relation is called a constructive space-time.

Constructive meaning: reminder. The main difference between this new definition and the original definition of the kinematic causality is that the existential quantifier $\exists$ (and the disjunction $\lor$) are understood constructively: as the existence of an algorithm that provides the corresponding objects; see, e.g., [1–4, 9, 18, 19]. In these terms:

- The formula $\forall a \, \exists \underline{a}, \overline{a} \, (\underline{a} \prec a \prec \overline{a})$ means that there exists an algorithm that, given an event a, returns events $\underline{a}$ and $\overline{a}$ for which $\underline{a} \prec a \prec \overline{a}$.
- The formula $a \prec b \Rightarrow \exists c \, (a \prec c \prec b)$ means that there exists an algorithm that, given two events a and b for which $a \prec b$, returns an event c for which $a \prec c \prec b$.
- The formula $a \prec b, c \Rightarrow \exists d \, (a \prec d \prec b, c)$ means that there exists an algorithm that, given events a, b, and c for which $a \prec b, c$, returns an event d for which $a \prec d \prec b, c$.
- The formula $b, c \prec a \Rightarrow \exists d \, (b, c \prec d \prec a)$ means that there exists an algorithm that, given events a, b, and c for which $b, c \prec a$, returns an event d for which $b, c \prec d \prec a$.
- The formula $a \prec b \Rightarrow \forall c \, (a \prec c \lor b \not\preceq c)$ means that there exists an algorithm that, given events a, b, and c for which $a \prec b$, returns either a true statement $a \prec c$ or a true statement $b \not\preceq c$.

- The formula $a \prec b \Rightarrow \exists i\,(a \prec x_i \prec b)$ means that there exists an algorithm that, given events a and b for which $a \prec b$, returns a natural number i for which $a \prec x_i \prec b$.
- The formula $x_i \prec x_j \Leftrightarrow \exists n\,(x_i \prec_n x_j)$ means that there exists an algorithm that, given natural numbers i and j for which $x_i \prec x_j$, returns a natural number n for which $x_i \prec_n x_j$.
- Finally, the fact that the ternary relation $x_i \prec_n x_j$ is decidable can be described as

$$\forall i\,\forall j\,\forall n\,(x_i \prec_n x_j \vee x_i \nprec_n x_j).$$

Comment. In strictly constructive terms, we can say that points x_i are simply natural numbers, $x_i \prec_n x_j$ is a ternary relation between natural numbers, and an arbitrary constructive event a can be described by two constructive sequences m_i and M_i for which $x_{m_i} \prec a \prec x_{M_i}$, $x_{m_i} \to x$, and $x_{M_i} \to x$.

In these terms, if an event a is described by sequences m_i and M_i and an event b is described by sequences n_i and N_i, then $a \prec b$ means that there exist i and j for which $x_{M_i} \prec x_{n_j}$.

4   Constructive Space-Time Version of Urysohn's Lemma

Lemma 3. *For every constructive kinematic casuality relation, for every $a \prec b$, there exists a constructive $\preceq$-monotonic function $f_{(a,b)} : X \to [0,1]$ for which $f_{(a,b)}(-a^+) = 0$ and $f_{(a,b)}(b^+) = 1$.*

Comment. This formulation is interpreted constructively.

A real number x is given constructively if we have an algorithm that, given accuracy k, returns a rational number r_k with $|r_k - x| \leq 2^{-k}$. A function $f(x)$ is given constructively if, given an input x – i.e., a black box that, given k, returns a 2^{-k}-approximation r_k to the number x – we can compute the value $f(x)$ with a given accuracy.

In this sense, the above formulation means the existence of an algorithm, that, given a, b, x, and a given accuracy k, computes the rational number which is 2^{-k}-close to $f_{(a,b)}(x)$.

Proof.

Part 1. Let us define $\prec$-monotonic values $\gamma(p/2^q)$ for all natural numbers p and q for which $p \leq 2^q$. We will define them inductively, first for $q = 0$, then for $q = 1$, etc.

For $q = 0$: We take $\gamma(0) = a$ and $\gamma(1) = b$.

From q to $q+1$: Since $(2p)/2^{q+1} = p/2^q$, the values $\gamma((2p)/2^{q+1})$ are already defined. We just need to define the values $\gamma((2p+1)/2^{q+1})$ corresponding to a midpoint

$$(2p+1)/2^{q+1} = \frac{p/2^q + (p+1)/2^q}{2}$$

between $p/2^q = (2p)/2^{q+1}$ and $(p+1)/2^q = (2p+2)/2^{q+1}$. For each p, since $\gamma(p/2^q) \prec \gamma((p+1)/2^q)$, we can run the algorithm that, given $a \prec b$, returns i for which $a \prec x_i \prec b$; this algorithm is a part of the description of the constructive kinematic causality relation.

By applying this algorithm to $\gamma(p/2^q)$ and $\gamma((p+1)/2^q)$, we get an integer i for which $\gamma(p/2^q) \prec x_i \prec \gamma((p+1)/2^q)$. We then take $\gamma((2p+1)/2^{q+1}) = x_i$.

Comment. In this paper, we operate within an algorithmic approach to constructive mathematics, where existence means the existence of an algorithm. For readers who are more familiar with more general axiomatic approach to constructive mathematics, it is worth mentioning that this construction requires dependent choice.

Part 2. For every x, we now define $f_{(a,b)}(x) \overset{\text{def}}{=} \sup\{r : \gamma(r) \prec x\}$. Let us explain how this function can be computed.

Due to the properties of the constructive kinematic causality relation, for each x and for each p and q, we have $\gamma(p/2^q) \prec x \vee \gamma((p+1)/2^q) \not\prec x$, and hence

$$f_{(a,b)}(x) > p/2^q \vee f_{(a,b)}(x) \le (p+1)/2^q.$$

In other words, there exist an algorithm that, given x, p, and q, tells us whether $f_{(a,b)}(x) > p/2^q$ or $f_{(a,b)}(x) \le (p+1)/2^q$. For each q, by applying this algorithm for different $p \le 2^q$, we can compute the value $f_{(a,b)}(x)$ with accuracy 2^{-q}.

So, the function $f_{(a,b)}(x)$ is indeed computable. $\qquad\square$

5 Constructive Space-Time Metrization Theorem

Theorem 3. *For every constructive space-time X with a constructive kinematic causality relation $\prec$, there exists a constructive kinematic metric $\tau(a,b)$ which generates the corresponding kinematic causality relation $\prec$ — in the sense that $a \prec b \Leftrightarrow \tau(a,b) > 0$.*

Comment. This formulation is meant as a constructive one: that there exists an *algorithm* that computes the values of $\tau(a,b)$.

Proof. In this proof, we use the above lemma, that for all $a \prec b$, there exists a $\preceq$-monotonic function $f_{(a,b)} : X \to [0,1]$ for which

$$f_{(a,b)}(-a^+) = 0 \text{ and } f_{(a,b)}(b^+) = 1.$$

Let us define, for every i, the following auxiliary function $f_{x_i} : X \to [0, 1]$:

$$f_{x_i}(b) \stackrel{\text{def}}{=} \sum_{j,n:\ x_i \prec_n x_j} 2^{-j} \cdot 2^{-n} \cdot f_{(x_i, x_j)}(b).$$

Since the relation $x_i \prec_n x_j$ is decidable, this function is computable: to compute it with accuracy 2^{-p}, it is sufficient to consider finitely many terms (j, n).

From the $\preceq$-monotonicity of the functions $f_{(x_i, x_j)}(x)$, one can conclude that their linear combination $f_{x_i}(b)$ is also $\preceq$-monotonic.

It is also possible to prove that $f_{x_i}(b) > 0 \Leftrightarrow x_i \prec b$. Indeed:

– If $f_{x_i}(b) > 0$, this means that $f_{(x_i, x_j)}(b) > 0$ for some j. Since $f_{(x_i, x_j)}(-x_i^+) = 0$, this means that b cannot belong to the complement $-x_i^+$, i.e., that

$$\neg\neg(x_i \prec b).$$

Thus, we have $x_i \prec b$.

– Vice versa, if $x_i \prec b$, then there exists a j for which $x_i \prec x_j \prec b$ and thus, $f_{(x_i, x_j)}(b) = 1$. Since $x_i \prec x_j$, there exists an n for which $x_i \prec_n x_j$ and thus, $f_{x_i}(b) \geq 2^{-j} \cdot 2^{-n} > 0$.

Similarly, we define functions $g_{x_i}(a)$ which are $\preceq$-decreasing and for which $g_{x_i}(a) > 0 \Leftrightarrow a \prec x_i$.

Now, we can define the kinematic metric in the same way as in the non-constructive proof:

$$\tau(a, b) = \sum_{i=1}^{\infty} 2^{-i} \cdot \min(g_{x_i}(a), f_{x_i}(b)).$$

Since $0 \leq g_{x_i}(a) \leq 1$ and $0 \leq f_{x_i}(b) \leq 1$, we have $0 \leq \min(g_{x_i}(a), f_{x_i}(b)) \leq 1$. One can easily check that this formula defines a computable function: to compute it with accuracy 2^{-p}, it is sufficient to compute the sum of the terms $i = 1, \ldots, p$, the remaining terms are bounded from above by the sum

$$2^{-(p+1)} + 2^{-(p+2)} + \ldots = 2^{-p}.$$

So, to complete the proof, we need to prove:

– that the function $\tau(a, b)$ is in correct relation with the kinematic causality relation, and
– that this function satisfies the anti-triangle inequality.

Let us first prove that $a \prec b \Leftrightarrow \tau(a, b) > 0$:

– If $a \prec b$, then there exists i for which $a \prec x_i \prec b$. Thus, by the properties of the functions f_{x_i} and g_{x_i}, we have $g_{x_i}(a) > 0$ and $f_{x_i}(b) > 0$ and thus, $\min(g_{x_i}(a), f_{x_i}(b)) > 0$. Hence, we have $\tau(a, b) > 0$.

– Vice versa, if $\tau(a, b) > 0$, this means that there exists an i for which $\min(g_{x_i}(a), f_{x_i}(b)) > 0$, i.e., for which $g_{x_i}(a) > 0$ and $f_{x_i}(b) > 0$. By the properties of the functions f_{x_i} and g_{x_i}, this means that $a \prec x_i$ and $x_i \prec b$. By transitivity, we can now conclude that $a \prec b$.

To prove the anti-triangle inequality, let us prove that a similar anti-triangle inequality holds for each of the expressions $\min(g_{x_i}(a), f_{x_i}(b))$, i.e., that

$$a \prec b \prec c$$

implies that

$$\min(g_{x_i}(a), f_{x_i}(b)) + \min(g_{x_i}(b), f_{x_i}(c)) \leq \min(g_{x_i}(a), f_{x_i}(c)).$$

Once we prove this, the desired anti-triangle inequality can be obtained by simply multiplying each of these inequalities by 2^{-i} and adding them.

To prove the above inequality, let us take into account that for every real number x, it not possible not to have $x > 0 \vee x \leq 0$: $\neg\neg(x > 0 \vee x \leq 0)$. Thus, we can consider separately

– situations when $\min(g_{x_i}(a), f_{x_i}(b)) > 0$ and
– situations when $\min(g_{x_i}(a), f_{x_i}(b)) = 0$,

and conclude that the double negation of the desired inequality holds. Since for constructive real numbers, $\neg\neg(p \leq q)$ is constructively equivalent to $p \leq q$, we get the desired inequality.

If $\min(g_{x_i}(a), f_{x_i}(b)) > 0$, this means that $x_i \in (a, b)$. Since $a \prec b \prec c$, this implies that we cannot have $x_i \in (b, c)$, and hence, that $\min(g_{x_i}(b), f_{x_i}(c)) = 0$. Since the function $f_{x_i}(b)$ is $\preceq$-monotonic and $b \prec c$, we have $f_{x_i}(b) \leq f_{x_i}(c)$ and thus, $\min(g_{x_i}(a), f_{x_i}(b)) \leq \min(g_{x_i}(a), f_{x_i}(c))$. Due to $\min(g_{x_i}(b), f_{x_i}(c)) = 0$, we have $\min(g_{x_i}(a), f_{x_i}(b)) + \min(g_{x_i}(b), f_{x_i}(c)) = \min(g_{x_i}(a), f_{x_i}(b))$ and thus, we get the desired inequality

$$\min(g_{x_i}(a), f_{x_i}(b)) + \min(g_{x_i}(b), f_{x_i}(c)) \leq \min(g_{x_i}(a), f_{x_i}(c)).$$

If $\min(g_{x_i}(b), f_{x_i}(c)) > 0$, this means that $x_i \in (b, c)$. Since $a \prec b \prec c$, this implies that we cannot have $x_i \in (a, b)$, and hence, that $\min(g_{x_i}(a), f_{x_i}(b)) = 0$. Since the function $g_{x_i}(b)$ is $\preceq$-decreasing and $a \prec b$, we have $g_{x_i}(b) \leq g_{x_i}(a)$ and thus, $\min(g_{x_i}(b), f_{x_i}(c)) \leq \min(g_{x_i}(a), f_{x_i}(c))$. Due to $\min(g_{x_i}(a), f_{x_i}(b)) = 0$, we have $\min(g_{x_i}(a), f_{x_i}(b)) + \min(g_{x_i}(b), f_{x_i}(c)) = \min(g_{x_i}(b), f_{x_i}(c))$ and thus, we get the desired inequality

$$\min(g_{x_i}(a), f_{x_i}(b)) + \min(g_{x_i}(b), f_{x_i}(c)) \leq \min(g_{x_i}(a), f_{x_i}(c)).$$

Finally, if $\min(g_{x_i}(a), f_{x_i}(b)) = 0$ and $\min(g_{x_i}(b), f_{x_i}(c)) = 0$, then

$$\min(g_{x_i}(a), f_{x_i}(b)) + \min(g_{x_i}(b), f_{x_i}(c)) = 0$$

and hence, since $\min(g_{x_i}(a), f_{x_i}(c)) \geq 0$, we also have the desired anti-triangle inequality. $\square$

6 Additional Results

Similar techniques enable us to prove constructive versions of other results about space-time models.

Definition 4. *By a* time coordinate t *on a space* X *with a kinematic causality relation* $\prec$, *we mean a function* $t : X \to \mathbb{R}$ *for which:*

- $a \prec b \Rightarrow t(a) < t(b)$; *and*
- $a \preceq b \Rightarrow t(a) \leq t(b)$.

Proposition 1. *On every constructive space-time X with a constructive kinematic causality relation $\prec$, there exists a constructive time coordinate.*

Proof. The desired constructive version of a time coordinate can be designed as follows:

$$t(b) \overset{\text{def}}{=} \sum_{i=1}^{\infty} 2^{-i} \cdot f_{x_i}(b).$$

Since $f_{x_i}(b) \in [0, 1]$, this is constructively defined (computable): to compute $t(b)$ with accuracy 2^{-p}, it is sufficient to add first p terms in the sum.

Let us prove that this function is indeed the time coordinate. Indeed, since each of the functions $f_{x_i}(b)$ is $\preceq$-monotonic, their convex combination $t(b)$ is also $\preceq$-monotonic.

To prove that the function $t(b)$ is $\prec$-monotonic, we can use the fact that $a \prec b$ implies the existence of a natural number i for which $a \prec x_i \prec b$. For this i, we have $f_{x_i}(a) = 0$ and $f_{x_i}(b) > 0$, hence $f_{x_i}(a) < f_{x_i}(b)$. For all other $j \neq i$, due to $a \prec b \Rightarrow a \preceq b$ and $\preceq$-monotonicity of f_{x_j}, we have $f_{x_j}(a) \leq f_{x_j}(b)$. Thus, by adding these inequalities, we get $t(a) < t(b)$. □

Comment. This result is similar to the constructive existence of a *utility function* $u(x)$, i.e., a function for which $a \prec b$ implies $u(a) < u(b)$, where $\prec$ is a preference relation; see, e.g., [5–8].

All possible time coordinates determine the causality relation: non-constructive case. In Newtonian physics, time $t(a)$ is absolute, and

$$a \preceq b \Leftrightarrow t(a) \leq t(b).$$

One of the main discoveries that led Einstein to his Special Relativity theory is the discovery that time is relative: a time coordinate corresponding to a moving body is different from the time coordinate corresponding to the stationary one. In general, there are many possible time coordinates t, each of which has the same property:

$$a \preceq b \Rightarrow \forall t \, (t(a) \leq t(b)).$$

For each of these time coordinates t, the mere fact that $t(a) \leq t(b)$ does not necessarily mean that a causally precedes b: it may happen that in some other

time coordinate, we have $t(a) > t(b)$. What *is* true is that if a is *not* causally preceding b, then there exist a time coordinate for which $t(a) > t(b)$:

$$a \not\preceq b \Rightarrow \exists t \, (t(a) > t(b)).$$

In non-constructive space-time geometry, the above two statements simply mean that

$$a \preceq b \Leftrightarrow \forall t \, (t(a) \leq t(b)),$$

i.e., that the causality relation is uniquely determined by the class of all possible time coordinates.

All possible time coordinates determine the causality relation: constructive case. Let us show that in the constructive case, under reasonable conditions, we also have the implication

$$a \not\preceq b \Rightarrow \exists t \, (t(a) > t(b)).$$

For that, we will need to impose an additional physically reasonable requirement.

For every event b, the *past cone* $P_b \stackrel{\text{def}}{=} \{c : c \preceq b\}$ is a closed set; thus, classically, its complement $-P_b = \{c : c \preceq b\}$ is an open set. The point a belongs to this set; thus, a whole open neighborhood of a belongs to this set as well. Since the topology is the Alexandrov topology, with intervals as a base, this means that there exist values $\underline{a}$ and $\bar{a}$ which which $\underline{a} \prec a \prec \bar{a}$ and the whole interval $(\underline{a}, \bar{a})$ belongs to the complement $-P_b$.

Since the sequence $\{x_i\}$ is everywhere dense in X, there is a point x_i in the interval $(\underline{a}, a)$, i.e., a point x_i for which $x_i \prec a$ and $x_i \not\preceq b$. By measuring the event locations with higher and higher accuracy, we will be able to detect this relation. Thus, it is reasonable to require that the following additional condition constructively holds:

$$a \not\preceq b \Rightarrow \exists i \, (x_i \prec a \,\&\, x_i \not\preceq b).$$

Let us show that under this condition, the above implication holds.

Proposition 2. *Let X be a constructive space-time with a constructive causality relation $\prec$ for which*

$$a \not\preceq b \Rightarrow \exists i \, (x_i \prec a \,\&\, x_i \not\preceq b).$$

Then, for every $a \not\preceq b$, there exists a constructive time coordinate t for which $t(a) > t(b)$.

Proof. Indeed, let i_0 be an index for which $x_{i_0} \prec a$ and $x_{i_0} \not\preceq b$. For this i_0, we thus have $f_{x_{i_0}}(a) > 0$ and $f_{x_{i_0}}(b) = 0$. Let us now construct the following time coordinate:

$$t(x) = \frac{2}{f_{x_{i_0}}(a)} \cdot f_{x_{i_0}}(x) + \sum_{i \neq i_0}^{\infty} 2^{-i} \cdot f_{x_i}(x).$$

Similar to the above formula, we can check that the function thus defined is indeed a time coordinate. It is therefore sufficient to show that $t(a) > t(b)$. Indeed:

- For $x = a$, the first term in the sum is equal to $\dfrac{2}{f_{x_{i_0}}(a)} \cdot f_{x_{i_0}}(x) = 2$, so $t(a) \geq 2$.
- For $x = b$, the first term is equal to 0. Since $f_{x_i}(x) \leq 1$ for all i, we thus conclude that

$$t(b) = \sum_{i \neq i_0}^{\infty} 2^{-i} \cdot f_{x_i}(x) \leq \sum_{i=1}^{\infty} 2^{-i} = 1.$$

Here, $t(a) \geq 2$ and $t(b) \leq 1$, and hence indeed $t(a) > t(b)$. □

Comment. Without this additional requirement, we can only prove that

$$a \not\preceq b \Rightarrow \neg\neg\exists t\,(t(a) > t(b)).$$

The existence of a standard metric. Another constructive result is the existence of a standard metric on each space-time model.

Proposition 3. *On every constructive space X with a constructive kinematic causality relation $\prec$, there exists a constructive metric $\rho(a, b)$.*

Proof. The desired constructive metric can be defined as follows:

$$\rho(a, b) \stackrel{\text{def}}{=} \sum_{i=1}^{\infty} 2^{-i} \cdot |f_{x_i}(a) - f_{x_i}(b)|.$$

One can easily check that this function is computable, and that it is indeed a metric – i.e., that it is symmetric and satisfies the triangle inequality. □

7 Remaining Challenges

Need to take symmetries into account. In this paper, given space-time X with the kinematic causality relation $\prec$, we designed a kinematic metric τ that is consistent with this relation.

In physics, however, causality is not everything. One of the most important notions of physics is symmetry. If space-time has symmetries – i.e., is invariant with respect to some transformations – it is therefore desirable to find a kinematic metric τ which is invariant with respect to these symmetries.

In the simplest case of a finite symmetry group G, we can explicitly define such a invariant constructive kinematic metric as

$$\tau_{\text{inv}}(a, b) \stackrel{\text{def}}{=} \sum_{g \in G} \tau(g(a), g(b)).$$

An important case is when X is both an ordered group and a space with a kinematic causality relation $\prec$, and the closures of all intervals are compact sets. It known that in the non-constructive case, there exists a left-invariant metric $\tau(a, b)$: namely, $\tau(a, b) = \mu_H(\{c : a \preceq c \preceq b\})$ where μ_H is the (left-invariant) Haar measure; see, e.g., [14]. It is desirable to constructivize this and similar results.

Need for feasible algorithms. In this paper, we have analyzed the existence of algorithms for computing the kinematic metric. From the practical viewpoint, it is important to make sure not only that such algorithms exist, but that they are *feasible* (i.e., can be computed in polynomial time); see, e.g., [11].

Partial analysis of feasibility of different computational problems related to space-time models is given in [17]. It is desirable to extend this analysis to the problem of computing kinematic metric.

References

1. Aberth, O.: Introduction to Precise Numerical Methods. Academic Press, San Diego, California (2007)
2. Beeson, M. Foundations of Constructive Mathematics: Metamathematical Studies. Springer, Berlin/Heidelberg/New York (1985)
3. Beeson, M.: Some relations between classical and constructive mathematics. Journal of Symbolic Logic **43** (1987) 228–246
4. Bishop, E., Bridges, D. S.: Constructive Analysis. Springer-Verlag, Berlin-Heidelberg-New York (1985)
5. Bridges, D. S.: The construction of a continuous demand function for uniformly rotund preferences. J. Math. Economics **21** (1992) 217–227
6. Bridges, D. S.: The constructive theory of preference orderings on a locally compact space II. Math. Social Sciences **27** (1994) 1–9
7. Bridges, D. S.: Constructive methods in mathematical economics In: Mathematical Utility Theory, J. Econ. (Zeitschrift für Nationalökonomie) Suppl. **8** (1999) 1–21
8. Bridges, D. S., Mehta, G. B.: Representations of Preference Orderings, Lecture Notes in Economics and Mathematical Systems **422**, Springer-Verlag, Heidelberg-Berlin-New York (1996)
9. Bridges, D. S., Vîta, L.: Techniques of Constructive Mathematics. Springer, New York, (2006)
10. Busemann, H.: Timelike Spaces. Warszawa, PWN (1967)
11. Gurevich, Y.: Platonism, constructivism, and computer proofs vs. proofs by hand. Bulletin of EATCS (European Association for Theoretical Computer Science **57** (1995) 145–166
12. Kelley, J. L.: General Topology, Springer, New York (1975)
13. Kronheimer, E. H., Penrose, R.: On the structure of causal spaces. Proc. Cambr. Phil. Soc. **63** (1967) 481–501
14. Kreinovich, V.: On the metrization problem for spaces of kinematic type. Soviet Mathematics Doklady **15** (1974) 1486–1490

15. Kreinovich, V.: Categories of Space-Time Models. Ph.D. dissertation, Soviet Academy of Sciences, Siberian Branch, Institute of Mathematics, Novosibirsk (1979) (in Russian)
16. Kreinovich, V.: Symmetry characterization of Pimenov's spacetime: a reformulation of causality axioms. International J. Theoretical Physics **35** (1996) 341–346.
17. Kreinovich, V., Kosheleva, O.: Computational complexity of determining which statements about causality hold in different space-time models. Theoretical Computer Science **405** (2008) 50–63
18. Kreinovich, V., Lakeyev, A., Rohn, J., Kahl, P.: Computational Complexity and Feasibility of Data Processing and Interval Computations. Kluwer, Dordrecht (1998)
19. Kushner, B. A.: Lectures on Constructive Mathematical Analysis. American Mathematical Society, Providence, Rhode Island (1985)
20. Misner, C. W., Thorne, K. S., Wheeler, J. A.: Gravitation. W. H. Freeman, New York (1973)
21. Nachbin, L.: Topology and Order. Van Nostrand, Princeton, (1965); reprinted by R. E. Kreiger, Huntington (1976)
22. Pimenov, R. I.: Kinematic spaces: Mathematical Theory of Space-Time, Consultants Bureau, N.Y. (1970)