International Journal of Theoretical Physics, Vol. 33, No. 8, 1994

Approximately Measured Causality Implies the Lorentz Group: Alexandrov–Zeeman Result Made More Realistic

Vladik Kreinovich[1]

Received June 21, 1993

If we know the causality relation between approximately known pairs of events, will we still be able to reconstruct affine coordinates? Our answer is: yes, we can reconstruct coordinates uniquely (modulo Lorentz group) from approximately measured coordinates. In this sense, we make the Alexandrov–Zeeman result more realistic.

1. INTRODUCTION

Laws of special relativity are usually described in terms of affine coordinates. Formulas of Minkowski geometry (i.e., geometry of special relativity) are usually described in terms of affine coordinates $x_0, \ldots, x_3$. So, to compare the predictions of special relativity with real life, we need to find a way to measure the affine coordinates of different events.

In the original papers by Einstein and in the textbooks that describe special relativity, usually a simplified situation is considered when we have inertial noninteractive bodies traveling in vacuum. Trajectories of these bodies are described by straight lines in a 4-dimensional space-time. Using these lines, we can measure the affine coordinates of every event, i.e., we can reconstruct the desired affine coordinate system.

In real-life situations, measuring affine coordinates is often a problem. In real life when the objects are interacting with each other it is often difficult to find a way to measure the ideal (affine) coordinates. As an example of such difficulties we can cite celestial mechanics (Brumberg, 1991; Finkelstein and Kreinovich, 1976; Finkelstein *et al.*, 1983; Cannon *et al.*, 1986),

[1]Computer Science Department, University of Texas at El Paso, El Paso, Texas 79968.

0020-7748/94/0800-1733$07.00/0

where very complicated algorithms are used to relate measured data to affine coordinates (of course, part of the complexity in this astronomy problem is caused by the fact that the effects of general relativity also have to be taken into consideration, but the effects of special relativity are also causing much trouble).

It would be nice to have a general way to measure affine coordinates. Methods that celestial mechanics uses to reconstruct affine coordinates and to compare the experimental results with the theoretical predictions of special relativity are based on using specific equations of celestial mechanics.

Since celestial mechanics is not the only possible area of application of special relativity, it is necessary to have a general method of reconstructing affine coordinates from physical measurements.

The Alexandrov–Zeeman theorem shows that such a general method is possible. A mathematical theorem that established the theoretical possibility of reconstructing affine coordinates from measurement results was first proved by Alexandrov (1950) [a detailed proof appeared first in Alexandrov and Ovchinnikova (1953)]. Namely, he showed that if for every pair of events a and b, we know whether a can influence b or not (i.e., whether a causally precedes b or not), then we will be able to reconstruct coordinates almost uniquely (modulo possible linear transformations, namely, modulo Lorentz transform, shifts, and dilations). The precise mathematical formulation of this result is as follows:

Definition 1. By M we denote a 4-dimensional space R^4 (this space is sometimes called an *arithmetic space*) with a special relation *precedes* that is defined as follows. Elements of M will be called *events* [in other words, an event is an element $a = (a_0, a_1, a_2, a_3) \in M$]. The real numbers a_i that form an event a are called its *affine coordinates*. We say that an event a *precedes* event b (or *causally precedes* b), and denote it by $a < b$, if

$$b_0 > a_0 \qquad \text{and} \qquad b_0 - a_0 \geq [(b_1 - a_1)^2 + (b_2 + a_2)^2 + (b_3 - a_3)^2]^{1/2}$$

Alexandrov's Theorem. Assume that $f: M \to M$ is a 1–1 mapping of M onto itself such that $a < b$ iff $f(a) < f(b)$. Then f is linear. Moreover, f is a composition of a Lorentz transformation, a shift in 4-space, and a dilation.

Let us show that this theorem is directly related to the problem of reconstructing affine coordinates from the measurement data. Indeed, suppose that we made many experiments, after which, for every pair of events, we know whether the first event causally precedes the second one or not.

To describe these experiments in mathematical terms, we must somehow mark the events, i.e., describe them by numbers. Since we do not yet know how to measure *affine* coordinates of these events, we will use *some* coordinates. In other words, we will use four arbitrary measurable quantities to serve as four (initial) coordinates. For example, as A_0, we can take the time measured by some noninertial clock, and by A_i, distances measured by noninertial rulers.

We will call the resulting 4-vector $A=(A_0, A_1, A_2, A_3)$ a *physical event*. To distinguish physical events (i.e., initial coordinates) from the events described in terms of their desired affine coordinates, we will use capital letters $A, B, \ldots$ for physical events, and lowercase letters $a, b, \ldots$ for events in terms of affine coordinates x_i.

For every two physical events A and B we know from experiment whether A causally precedes B or not. We want to use this information to reconstruct affine coordinates of all events. In other words, we want to assign to every physical event A four values $a=F(A)$ in such a way that $A<B$ if and only if the events $a=F(A)$ and $b=F(B)$ are causally related according to the formulas of special relativity.

Definition 2. By a *causality relation* $\tilde{<}$, we mean a partial ordering relation on R^4. A 1–1 mapping $F: R^4 \rightarrow M$ of R^4 onto M is called a *coordinate system*. We say that in a coordinate system F, *a causality relation* $\tilde{<}$ *is described by* (the formulas of) *special relativity* if and only if for every $A, B \in R^4$, $A \tilde{<} B$ if and only if for $a=F(A)$ and $b=F(B)$, we have $a<b$, i.e., if

$$b_0 > a_0 \qquad \text{and} \qquad b_0 - a_0 \geq [(b_1-a_1)^2 + (b_2-a_2)^2 + (b_3-a_3)^2]^{1/2}$$

Comment. To avoid misunderstanding, let us repeat the difference between M and R^4: M is a 4-dimensional space with an additional ordering relation $<$ (that is described by the formulas of special relativity); R^4 is a 4-dimensional space on which the causality relation (denoted by $\tilde{<}$) may be defined by different formulas.

We consider the case when such a coordinate system exists. It is well known that this mapping is not uniquely determined by the causality relation. For example, if in a coordinate system F causality is described by the formulas of special relativity, then for every 4-vector $\{s_i\}$ the same is true for the "shifted" coordinates $F_i' = F_i + s_i$. Similarly, if we apply a dilation $F_i \rightarrow \lambda F_i$ ($\lambda = \text{const} > 0$) or a Lorentz transformation to any coordinate system in which causality is described by formulas of special relativity, then the new coordinate system will have the same property.

Alexandrov's theorem shows that these are the only possible sources of nonuniqueness. Namely, the following corollary is true:

Corollary 1. Suppose that a causality relation $\tilde{<}$ is fixed. If F and G are two different coordinate systems in which causality is described by formulas of special relativity, then $G = T(F)$, where T is a composition of shifts, dilations, and Lorentz transformations.

Proof. Indeed, since in both coordinate systems F and G the causality relation is described by the formulas of special relativity, we have $A \precsim B$ iff $F(A) < F(B)$ iff $G(A) < G(B)$. Therefore, for $T = G(F^{-1})$, we have $a < b$ iff $Ta < Tb$. Hence, according to Alexandrov's theorem, T is a composition of a shift, a dilation, and a Lorentz transformation. QED

Historical Comment. This theorem became well known after E. C. Zeeman published a more general result in English (Zeeman, 1964) [for further mathematical results, see, e.g., Kuz'minykh (1975, 1976), Benz (1977), Lester (1977*a*, 1983), and Kosheleva *et al.* (1979*a*); for a modern textbook description, see Naber (1992)]. The relationship between this result and the actual astronomical methods of reconstructing affine coordinates from measurements is shown in Kosheleva *et al.* (1979*b*).

In real life, measurements are never absolutely accurate. Alexandrov's theorem says that if we know exactly which pairs are causally connected and which pairs are not, then we can reconstruct affine coordinates. How can we determine that experimentally? How can we, given events a and b, figure out experimentally whether a causally precedes b or not?

To answer this question, let us recall how causality is determined in real-life situations. For example, how can we prove that a suspect actually murdered a victim? Of course, there can be lots of indirect evidence, but the only direct proof will be if we find an individual trace of the suspect on the victim, a trace that cannot be explained by a random coincidence: e.g., his fingerprints, or his DNA trace, etc.

A similar approach can be used to check causal relation between different events: while in an event a, we emit a signal of a special type (of unique type, so that such a signal cannot appear as a random noise). Then, if a trace of this signal is found in b, we conclude that a causally precedes b, else that a does not precede b.

In a real-life experiment, we can only handle finitely many events, and we can measure the coordinates of events only with some accuracy. As a result, after reconstruction, we will have only approximate values of affine coordinates. So, the more accurate physical interpretation of Alexandrov's theorem is as follows: if we increase the number of events, and make more and more accurate measurements, then we will get more and more accurate values of the affine coordinates of these events. In the limit of infinitely many events covering all space-time and absolutely accurate measurements, we will get the absolutely accurate affine coordinates.

The uncertainty principle leads to a restriction on the accuracy with which we can measure some physical quantities. The above application of

Alexandrov's theorem to real-life physical situations is based on the assumption that in principle we can measure every physical quantity with an arbitrary accuracy. This is true in "classical" (nonquantum) physics. If we apply this idea to real-life events a and b, then for the pairs of events that are separated by microscopic distances, the classical description does not work: at such distances, quantum effects become dominant.

This leads to the following problem. According to the above-described method, the experimental procedure of determining the causality relation is as follows. Suppose that we have an event A for which we have measured the values A_i of its coordinates. To list the events that are causally following A, during an event A, we emit (in all spatial directions) a signal with a certain complicated pattern. At several spatial locations we place sensors that all the time try to pick a signal with this particular pattern. If in some moment of time a sensor picks that signal, this means that this particular event B (picking that signal) causally follows A. To apply the above-described methodology, we must measure (initial, nonaffine) coordinates b_i of this event B. In particular, we must measure the difference $B_0 - A_0$ between the values of noninertial time in A and in B.

According to Heisenberg's uncertainty principle, if we want to measure time with accuracy Δt, then, to implement this measurement, we must use the amount of energy E that is greater than or equal to $h/\Delta t$ (where h is the Planck constant). If we spend less energy, then we will not get the desired accuracy. Therefore, the average power P that needs to be applied is $\approx E/T \geq h/(\Delta t\, T)$, where by T we denote the time interval between the two events. So, the smaller this time interval, the more power we need to apply to measure the coordinates B_0 with a given accuracy. But the more power we apply, the more we disturb the system. If a sufficiently huge amount of power is applied, it will cause the sensors to melt and thus make all the measurements impossible. So, for every measuring technology, there is a maximum amount of power P_0 that can be applied. As a result, if we measure the time interval of size $\approx T$, we can only get an accuracy Δt such that $h/(\Delta t\, T) \leq P_0$, i.e., such accuracy Δt that $\Delta t \geq h/(P_0 T)$. In other words, we cannot measure a time interval of size T with an accuracy better than k/T for some constant $k = h/P_0$.

Similarly, from another Heisenberg inequality that relates measuring spatial coordinates and momentum, we conclude that we cannot measure a spatial distance of size r with an accuracy better than k_1/r for some constant r.

Combining these two inequalities, we can conclude that the accuracy with which we can measure the coordinates B_i of the event B (when a sensor picks a signal) is limited from below (and this accuracy decreases and tends to 0 as B goes further away from A, i.e., as the differences $B_0 - A_0$ and/or $B_i - A_i$ increase).

So, for a given measurement technique, we cannot measure the coordinates of B with arbitrary accuracy. In other words, we can "measure causality" only approximately. A natural question is: *is* this *approximately measured causality sufficient to reconstruct affine coordinates*?

What we plan to do. Our answer to this question is "yes." Yes, we can reconstruct coordinates uniquely (modulo Lorentz group) from an approximately measured causality relation.

In Section 2 we give the necessary definitions, and the precise mathematical formulation of our result. Its proof is given in Section 3.

2. DEFINITIONS AND THE MAIN RESULT

Notation. For $a, b \in M$, let us denote

$$d(a, b) = [(a_0 - b_0)^2 + (a_1 - b_1)^2 + a_2 - b_2)^2 + (a_3 - b_3)^2]^{1/2}$$

Comment. This "Euclidean" metric describes to what extent the points a and b are far away from one another. We are using this metric because as already mentioned, the accuracy with which we can measure the difference between the coordinates of the two events a and b tends to 0 as a and b become more and more separated (either in time or in space). The distance d is introduced to describe this "separation." This description has an evident drawback: namely, this distance is not Lorentz–invariant (i.e., d changes if we apply a Lorentz transformation). However, we will use d in our formulations because the final result will not depend on the exact choice of d.

Definition 3. Assume that a function $h: R^+ \to R^+$ from the set of positive real numbers R^+ to itself is given, and $h(t) \to 0$ as $t \to \infty$. We say that a set $C \subset M \times M$ of pairs (a, b), $a, b \in M$, is a *measured causality* if the following two properties are true:

1. If $(a, b) \in C$, then there exists a $b' \in M$ such that $a < b'$ and $d(b, b') \leq h(d(a, b))$.

2. If $(a, b) \notin C$, then there exists a $b' \in M$ such that $a \not< b'$ and $d(b, b') \leq h(d(a, b))$.

Comments. 1. To avoid confusion, let us mention that measured causality is *not necessarily a causality relation* in the sense of Definition 2. Our only requirement on the measured causality is that it is *close* to the causality relation (close in the precise sense described by this definition).

2. The function $h(t)$ describes the accuracy with which we can measure the difference between the coordinates of the events a and b for which

$d(a, b) = t$. We have already seen that from Heisenberg's principle it follows that $h(t) \leq k/t$ for some constant k and therefore $h(t) \to 0$ as $t \to \infty$.

The set C represents the results of approximately measured causality pairs. If $(a, b) \in C$, this means that for some event b' we actually detected the influence of a. In the ideal case of absolutely precise measurements we would then have placed the pair (a, b') into the list C of the pairs for which we have experimentally confirmed that a causally precedes b'. However, since we are unable to measure the coordinates of this event b' precisely, we placed into this list C a slightly different pair (a, b), where b are approximately measured coordinates of b'. According to the meaning of the function $h(t)$, the accuracy with which we measure these coordinates is $\leq h(d(a, b))$. In other words, $d(b, b') \leq h(d(a, b))$.

The second condition can be interpreted in the same manner: if $(a, b) \notin C$, this means that for some event b' that was approximately measured as b we did not find any influence of a and therefore concluded that $a \not< b'$. Since we can measure the coordinates with accuracy $h(d(a, b))$, we have $d(b, b') \leq h(d(a, b))$.

Theorem. Assume that $C \subset M \times M$ is a measured causality, $f: M \to M$ is a continuous 1–1 mapping of M onto itself such that the inverse mapping is also continuous, and $(a, b) \in C$ iff $(f(a), f(b)) \in C$. Then f is linear. Moreover, f is a composition of a Lorentz transformation, shift in 4-space, and dilation.

Comment. We have explained how Alexandrov's theorem leads to a conclusion that from the causality relation we can reconstruct affine coordinates. Similarly, one can conclude from our theorem that to reconstruct affine coordinates it is sufficient to know approximately measured causality. Let us give the precise definitions.

Definition 4. By a *result of measuring causality* we mean a set of pairs $\tilde{C} \subset R^4 \times R^4$. A continuous 1–1 mapping $F: R^4 \to M$ onto M is called a *coordinate system* if the inverse mapping F^{-1} is also continuous. We say that in a coordinate system F *the result* $\tilde{C}$ *of measuring causality is described by special relativity* iff the set $F(\tilde{C}) = \{(F(A), F(B)) | (A, B) \in \tilde{C}\}$ is a measured causality (in the sense of Definition 3).

Corollary 2. Suppose that a result $\tilde{C}$ of measuring causality is given. If F and G are two different coordinate systems in which $\tilde{C}$ is described by (formulas of) special relativity, then $G = T(F)$, where T is a composition of a shift, a dilation, and a Lorentz transformation.

Comments. 1. This corollary proves that affine structure can be uniquely reconstructed from the approximately measured causality (i.e., whatever coordinates we reconstruct, they will be related by a linear formula and hence the notions of a line, a plane, etc., will be the same in both coordinate systems).

2. This corollary follows from our theorem in the same manner as Corollary 1 follows from Alexandrov's theorem.

3. PROOF OF THE THEOREM

The main idea of this proof is as follows: We will prove that if f satisfies the condition of the theorem (i.e., if f preserves C), then f also preserves the causality relation $<$. Then, we can apply Alexandrov's theorem to get the desired conclusion.

Notation. We will need the following notation (the majority of them are more or less standard in this field; see, e.g., Naber (1992)):

$a \ll b$ denotes

$$b_0 - a_0 > [(b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2]^{1/2}$$

$\mathbf{a}$ denotes a 3D vector part (a_1, a_2, a_3) of an arbitrary event a.

$\rho(\mathbf{a}, \mathbf{b})$ denotes $[(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2]^{1/2}$ (Euclidean distance in a 3D space).

$\tau(a, b)$ denotes $b_0 - a_0 - \rho(\mathbf{a}, \mathbf{b})$.

a^+ denotes $\{b | a < b\}$, called a *future cone* of an event a.

$\tilde{a}^+$ denotes $\{b | (a, b) \in C\}$, called an *approximate future cone* of an event a.

a^- is defined as $\{b | b < a\}$, called a *past cone.*

$\tilde{a}^-$ is defined as $\{b | (b, a) \in C\}$, called an *approximate past cone.*

Lemma. We will prove that for every two events a and b

$$a \ll b \to \exists c(\tilde{b}^+ \subseteq \tilde{a}^+ \cup \tilde{c}^-)$$

and

$$\exists c(\tilde{b}^+ \subseteq \tilde{a}^+ \cup \tilde{c}^-) \to (a < b \vee a = b)$$

Proof of the Lemma. 1. Let us first prove that if $a \ll b$, $b < x$, and $d(x, y) \leq (1/2)\tau(a, b)$, then $a < y$.

By definition of $a \ll b$ we have $\tau(a, b) > 0$. By the definition of τ we have

$$b_0 - a_0 = \rho(\mathbf{a}, \mathbf{b}) + \tau(a, b) \tag{1}$$

From $b < x$ we conclude that

$$x_0 - b_0 \geq \rho(\mathbf{x}, \mathbf{b}) \tag{2}$$

From the inequality

$$d(x, y) = [(x_0 - y_0)^2 + \rho^2(\mathbf{x}, \mathbf{y})]^{1/2} \leq (1/2)\tau(a, b)$$

we conclude that $|y_0 - x_0| \leq (1/2)\tau(a, b)$ [hence

$$y_0 - x_0 \geq -\frac{1}{2}\tau(a, b)] \tag{3}$$

and

$$\rho(\mathbf{x}, \mathbf{y}) \leq \frac{1}{2}\tau(a, b) \tag{4}$$

Adding equation (1) with inequalities (2) and (3), we conclude that

$$\begin{aligned} y_0 - a_0 &= (y_0 - x_0) + (x_0 - b_0) + (b_0 - a_0) \\ &\geq -\frac{1}{2}\tau(a, b) + \rho(\mathbf{x}, \mathbf{b}) + \rho(\mathbf{b}, \mathbf{a}) + \tau(a, b) \\ &= \rho(\mathbf{b}, \mathbf{a}) + \rho(\mathbf{x}, \mathbf{b}) + \frac{1}{2}\tau(a, b) \end{aligned}$$

Because of (4), we can conclude that

$$y_0 - a_0 \geq \rho(\mathbf{b}, \mathbf{a}) + \rho(\mathbf{x}, \mathbf{b}) + \rho(\mathbf{y}, \mathbf{x}) \tag{5}$$

But ρ is a Euclidean metric, so the triangle inequality leads to $\rho(\mathbf{b}, \mathbf{a}) + \rho(\mathbf{x}, \mathbf{b}) + \rho(\mathbf{y}, \mathbf{x}) \geq \rho(\mathbf{a}, \mathbf{y})$. Hence, from (5), we conclude that $y_0 - a_0 \geq \rho(\mathbf{y}, \mathbf{a})$, i.e., that $a < y$.

2. Let us now prove that if $a \ll b$, then there exists a real number $s > 0$ such that if $x \in \tilde{b}^+$ and $d(a, x) \leq s$, then $x \in \tilde{a}^+$.

Indeed, since $h(t) \to 0$ as $t \to \infty$, there exists a t_0 such that if $t \geq t_0$, then $h(t) \leq (1/4)\tau(a, b)$. Let us show that the desired property is true for $s = t_0 + d(a, b)$. Indeed suppose that $x \in \tilde{b}^+$ and $d(a, x) \geq s$. Let us prove that $x \in \tilde{a}^+$ by reduction to a contradiction.

Assume that $x \notin \tilde{a}^+$ [i.e., that $(a, x) \notin C$]. By Definition 3, this means that there exists an x' such that $d(x, x') \leq h(d(a, x))$ and $a \not< x'$. Since $d(a, x) \geq s > t_0$, from the choice of t_0 we conclude that $h(d(a, x)) \leq (1/4)\tau(a, b)$. Therefore,

$$d(x, x') \leq \frac{1}{4}\tau(a, b) \tag{6}$$

From $(b, x) \in C$ we can likewise conclude that there exists an x'' such that $d(x'', x) \leq h(d(b, x))$ and $b < x''$. From triangle inequality we conclude that $d(x, b) \geq d(a, x) - d(a, b)$. Since $d(a, x) \geq s = t_0 + d(a, b)$, we conclude that $d(x, b) \geq t_0 + d(a, b) - d(a, b) = t_0$ and hence $h(d(x, b)) \leq (1/4)\tau(a, b)$. So, $d(x, x'') \leq (1/4)\tau(a, b)$. From this inequality and from (6) we conclude that $d(x', x'') \leq (1/2)\tau(a, b)$. Therefore, applying part 1 of the proof with x'' instead of x and x' instead of y, we can now conclude that $a < x'$. This conclusion contradicts our choice of x', according to which $a \nless x'$. This contradiction shows that the assumption that $x \notin \tilde{a}^+$ was impossible, and hence $x \in \tilde{a}^+$.

3. Similarly, we can prove that if $a \ll b$, $d(a, x) \geq s$ (where s is the same as in part 2) and $b < x$, then $x \in \tilde{a}^+$.

Indeed, if $x \notin \tilde{a}^+$, then, according to Definition 3, there exists an event x' such that $d(x, x') \leq h(d(a, x))$ and $a \nless x'$. Since $d(a, x) \geq s$, we have $h(d(a, x)) \leq (1/4)\tau(a, b)$. Therefore, $d(x, x') \leq (1/4)\tau(a, b) < (1/2)\tau(a, b)$. Since $b < x$, from part 1, we can conclude that $a < x'$, which contradicts $a \nless x'$. This contradiction proves that the assumption $x \notin \tilde{a}^+$ is impossible, and so $x \in \tilde{a}^+$.

4. Let us now prove that if $a \ll b$, then there exists a c such that $\tilde{b}^+ \subset \tilde{a}^+ \cup \tilde{c}^-$.

Indeed, let us take $c = a + (2s + \tau(a, b), 0, 0, 0)$ and $d = a + (2s, 0, 0, 0)$, where s was defined in part 2. Then $d \ll c$ and $\tau(d, c) = \tau(a, b)$. We will prove the desired statement for this c.

4.1. We have already proved that if $x \in \tilde{b}^+$ and $d(a, x) \geq s$, then $x \in \tilde{a}^+$. So, in order to complete the proof of the desired statement, it is sufficient to prove that if $x \in \tilde{b}^+$ and $d(a, x) < s$, then $x \in \tilde{c}^-$. To prove this auxiliary statement, we will prove a slightly stronger statement: that if $d(a, x) < s$, then $x \in \tilde{c}^-$.

4.2. To prove that statement, we will first prove that if $d(a, x) < s$, then $x < d$ and $d(c, x) \geq s$. Then, arguing just as in part 2, we will be able to conclude that $(x, c) \in C$, i.e., that $x \in \tilde{c}^-$.

In order to use the arguments similar to the ones used in part 2, it is necessary to make the following remark. In that proof, from the assumption that $x \notin \tilde{a}^+$ [i.e., that $(a, x) \notin C$], we immediately concluded (using Definition 3) that there exists an x' such that $d(x, x') \leq h(d(a, x))$ and $a \nless x'$. In our case, if we assume that $x \notin \tilde{c}^-$, then we cannot simply refer to Definition 3 and make a similar conclusion that there exists an x' such that $x' \nless c$ and $d(x, x') = d(c, c') \leq h(d(x, c))$. However, this conclusion is still true. Indeed, if $x \notin \tilde{c}^-$, then (by definition of a measured causality) it means that there exists an event c' such that $x \nless c'$ and $d(c, c') \leq h(d(x, c))$. Therefore, for $x' = x + c - c'$, we have $x' \nless c$ and $d(x, x') = d(c, c') \leq h(d(x, c))$.

4.3. Let us prove that $d(c, x) \geq s$. Indeed, from the definition of c, we can conclude that $d(a, c) = 2s + \tau(a, b)$. Therefore, from the triangle inequality, we can conclude that $d(c, x) \geq d(a, c) - d(a, x)$. But $d(a, c) = 2s + \tau(a, b) > 2s$, and $d(a, x) < s$, hence $d(c, x) > s$.

4.4. Let us now prove that $x < d$. Indeed,

$$d(a, x) = [(a_0 - x_0)^2 + \rho^2(\mathbf{a}, \mathbf{x})]^{1/2} < s$$

implies that $|a_0 - x_0| < s$ and $\rho(\mathbf{a}, \mathbf{x}) < s$. We constructed d as a 4-vector $(d_0, \mathbf{d}) = (a_0 + 2s, \mathbf{a})$. Therefore,

$$d_0 - x_0 = (d_0 - a_0) - (a_0 - x_0) = 2s - (a_0 - x_0) > 2s - s = s$$

and $\rho(\mathbf{d}, \mathbf{x}) = \rho(\mathbf{a}, \mathbf{x}) < s$. Hence, $d_0 - x_0 > s > \rho(\mathbf{x}, \mathbf{d})$, and $x < d$.

Both statements are proved, and so is part 4.

5. Let us now prove that if there exists a c such that $\tilde{b}^+ \subset \tilde{a}^+ \cup \tilde{c}^-$, then $a < b$ or $a = b$.

5.1. We will prove this by reduction to a contradiction. Suppose that the condition is true (i.e., $\tilde{b}^+ \subset \tilde{a}^+ \cup \tilde{c}^-$ for some c), but the conclusion is not, i.e., $a \not< b$ and $a \neq b$. This means that $b_0 - a_0 < \rho(\mathbf{a}, \mathbf{b})$, i.e., $\tau(a, b) = b_0 - a_0 - \rho(\mathbf{a}, \mathbf{b}) < 0$.

First, we will consider the case when $\mathbf{b} \neq \mathbf{a}$. In this case, to arrive at a contradiction, we will consider the following sequence of events:

$$x^{(N)} = (b_0 + N\rho(\mathbf{a}, \mathbf{b}) - \frac{1}{2}\tau(a, b), \mathbf{b} + N(\mathbf{b} - \mathbf{a})) \qquad (N = 1, 2, \ldots)$$

We will prove that for sufficiently large N, the following three statements are true:

- $x^{(N)} \in \tilde{b}^+$
- $x^{(N)} \notin \tilde{a}^+$
- $x^{(N)} \notin \tilde{c}^-$

In other words, we will prove that $\tilde{b}^+$ is not a subset of the union $\tilde{a}^+ \cup \tilde{c}^-$, which is contrary to our assumption. So, to get the desired contradiction, let us prove the above three statements.

5.1.1. Let us prove that for sufficiently large N, $x^{(N)} \in \tilde{b}^+$.

Indeed, one can easily check that for all N, $x^{(N)} > b'$, where we denoted

$$b' = b + \left(-\frac{1}{2}\tau(a, b), 0, 0, 0\right)$$

and $b \ll b'$ $[\tau(b, b') = -\frac{1}{2}\tau(a, b)]$. From part 3 we can now conclude that there exists an s such that if $d(b, x^{(N)}) \geq s$, then $x^{(N)} \in \tilde{b}^+$.

To find out when this inequality $d(b, x^{(N)}) \geq s$ is true, let us estimate the distance $d(b, x^{(N)}) \geq s$.

We have $x_0^{(N)} - b_0 = N\rho(\mathbf{a}, \mathbf{b}) - (1/2)\tau(a, b) > N\rho(\mathbf{a}, \mathbf{b})$, and

$$\rho(\mathbf{b}, \mathbf{x}^{(N)}) = \{[N(\mathbf{b} - \mathbf{a})]^2\}^{1/2} = N\rho(\mathbf{a}, \mathbf{b})$$

Therefore,

$$\begin{aligned} d(x^{(N)}, b) &= [(x_0^{(N)} - b_0)^2 + \rho^2(\mathbf{x}^{(N)}, \mathbf{b})]^{1/2} \\ &\geq [2N^2\rho^2(\mathbf{a}, \mathbf{b})]^{1/2} = \sqrt{2}\, N\rho(\mathbf{a}, \mathbf{b}) \end{aligned}$$

We are considering the case when $\mathbf{a} \neq \mathbf{b}$, hence $\rho(\mathbf{a}, \mathbf{b}) > 0$, and $d(x^{(N)}, b) \to \infty$ as $N \to \infty$. Therefore, for sufficiently large N, $d(x^{(N)}, b) \geq s$, and thus $x^{(N)} \in \tilde{b}^+$.

5.1.2. Let us now prove that for sufficiently large N, $x^{(N)} \notin \tilde{a}^+$.

We will prove it by reduction to a contradiction. Assume that there exist arbitrarily large integers N for which $x^{(N)} \in \tilde{a}^+$. This means that for every such N, there exists an event $x^{(N)'}$ such that $d(x^{(N)}, x^{(N)'}) \leq h(d(x^{(N)}, a))$, and $a < x^{(N)'}$. The distance $d(x^{(N)}, a)$ can be estimated from the triangular inequality: $d(x^{(N)}, a) \geq d(x^{(N)}, b) - d(a, b)$. Since we have already proved in part 5.1.1 that $d(x^{(N)}, b) \to \infty$ as $N \to \infty$, we can conclude that $d(x^{(N)}, a) \to \infty$. Since $h(t) \to 0$ as $t \to \infty$, we can conclude that $h(d(x^{(N)}, a)) \to 0$. In particular, starting from some N_0, we will have $h(d(x^{(N)}, a)) < -\frac{1}{4}\tau(a, b)$.

We assumed that there exists an infinite increasing sequence of values of N for which $x^{(N)} \in \tilde{a}^+$. We can delete the values $N < N_0$ and still have an infinite sequence. So, without losing generality, we can assume that for all N from our sequence, $h(d(x^{(N)}, a)) < -\frac{1}{4}\tau(a, b)$ and, therefore $d(x^{(N)}, x^{(N)'}) < -\frac{1}{4}\tau(a, b)$ and $a < x^{(N)'}$.

From $a < x^{(N)'}$ we conclude that $x_0^{(N)'} - a_0 - \rho(\mathbf{x}^{(N)'}, \mathbf{a}) \geq 0$ and hence

$$x_0^{(N)'} - a_0 \geq \rho(\mathbf{x}^{(N)'}, \mathbf{a}) \tag{7}$$

From $d(x^{(N)}, x^{(N)'}) < -\frac{1}{4}\tau(a, b)$ it follows that $|x_0^{(N)} - x_0^{(N)'}| < -\frac{1}{4}\tau(a, b)$ and $\rho(\mathbf{x}^{(N)}, \mathbf{x}^{(N)'}) < -\frac{1}{4}\tau(a, b)$. Therefore, $x_0^{(N)} > x_0^{(N)'} + \frac{1}{4}\tau(a, b)$, and (from the triangle inequality)

$$\rho(\mathbf{x}^{(N)}, \mathbf{a}) \leq \rho(\mathbf{x}^{(N)'}, \mathbf{a}) + \rho(\mathbf{x}^{(N)}, \mathbf{x}^{(N)'}) < \rho(\mathbf{x}^{(N)'}, \mathbf{a}) - \frac{1}{4}\tau(a, b)$$

So, $\rho(\mathbf{x}^{(N)'}, \mathbf{a}) > \rho(\mathbf{x}^{(N)}, \mathbf{a}) + \frac{1}{4}\tau(a, b)$. Therefore, from (7), we conclude that

$$x_0^{(N)} - \tfrac{1}{4}\tau(a, b) - a_0 > \rho(\mathbf{x}^{(N)'}, \mathbf{a}) > \rho(\mathbf{x}^{(N)}, \mathbf{a}) + \tfrac{1}{4}\tau(a, b)$$

So,

$$x_0^{(N)} - a_0 - \rho(\mathbf{x}^{(N)}, \mathbf{a}) \geq \frac{1}{2}\tau(a, b) \tag{8}$$

From the definition of $x^{(N)}$, one can easily find that

$$x_0^{(N)} - a_0 = -\frac{1}{2}\tau(a, b) + (b_0 - a_0) + N\rho(\mathbf{a}, \mathbf{b})$$

and $\rho(\mathbf{x}^{(N)}, \mathbf{a}) = (N+1)\rho(\mathbf{a}, \mathbf{b})$. If we substitute these values into (8), we get the following inequality: $-\frac{1}{2}\tau(a, b) + (b_0 - a_0) - \rho(\mathbf{a}, \mathbf{b}) > \frac{1}{2}\tau(a, b)$. Moving the term with $\tau(a, b)$ into the right-hand side, we conclude that $(b_0 - a_0) - \rho(\mathbf{a}, \mathbf{b}) > \tau(a, b)$, while by definition, $\tau(a, b) = (b_0 - a_0) - \rho(\mathbf{a}, \mathbf{b})$ (a contradiction).

This contradiction shows that at most finitely many elements of the sequence $x^{(N)}$ belong to $\tilde{a}^+$, so, starting from some N, $x^{(N)} \notin \tilde{a}^+$.

5.1.3. Let us now prove that for sufficiently large N, $x^{(N)} \notin \tilde{c}^-$.

We will prove this statement also by reduction to a contradiction. Indeed, assume that there exist arbitrarily large integers N for which $x^{(N)} \in \tilde{c}^-$. By definition, this means that for every such N there exists an event $c^{(N)}$ such that $x^{(N)} < c^{(N)}$ and $d(c^{(N)}, c) \le h(d(x^{(N)}, c))$. Then, as one can easily see, for $x^{(N)\prime} = x^{(N)} + c - c^{(N)}$ we have $x^{(N)\prime} < c$ and $d(x^{(N)}, x^{(N)\prime}) \le h(d(x^{(N)}, c))$. Similarly to part 5.1.2, the distance $d(x^{(N)}, c)$ can be estimated from the triangular inequality: $d(x^{(N)}, c) \ge d(x^{(N)}, b) - d(b, c)$. Since we have already proved in part 5.1.1 that $d(x^{(N)}, b) \to \infty$ as $N \to \infty$, we can conclude that $d(x^{(N)}, c) \to \infty$. Since $h(t) \to 0$ as $t \to \infty$, we can conclude that $h(d(x^{(N)}, c)) \to 0$. In particular, starting from some N_0, we will have $h(d(x^{(N)}, c)) < -\frac{1}{4}\tau(a, b)$.

Similarly to part 5.1.2, without losing generality, we can assume that for all N from our sequence, $h(d(x^{(N)}, c)) < -\frac{1}{4}\tau(a, b)$, and therefore $d(x^{(N)}, x^{(N)\prime}) < -\frac{1}{4}\tau(a, b)$ and $x^{(N)\prime} < c$.

From $x^{(N)\prime} < c$ we conclude that $x_0^{(N)\prime} \le c_0$. From $d(x^{(N)}, x^{(N)\prime}) < -\frac{1}{4}\tau(a, b)$ we conclude that $|x_0^{(N)} - x_0^{(N)\prime}| < -\frac{1}{4}\tau(a, b)$. Therefore,

$$x_0^{(N)} < x_0^{(N)\prime} - \frac{1}{4}\tau(a, b) \le c_0 - \frac{1}{4}\tau(a, b)$$

But we have chosen $x^{(N)}$ in such a way that

$$x_0^{(N)} = b_0 + N\rho(\mathbf{a}, \mathbf{b}) - \frac{1}{2}\tau(a, b) \to \infty$$

as $N \to \infty$. Therefore, the inequality $x_0^{(N)} < c_0 - (1/4)\tau(a, b)$ cannot be true for arbitrarily large N.

This contradiction proves that our assumption was wrong, and, starting from some N, $x^{(N)} \notin \tilde{c}^-$.

5.2. We have proved the lemma for the case when $\mathbf{a} \neq \mathbf{b}$. To complete the proof of the lemma, it is necessary to consider the case when $\mathbf{a} = \mathbf{b}$. In this case $a \not< b$ means that $b_0 < a_0$ (and hence $b < a$). As an example of a sequence $x^{(N)}$ with the property that $x^{(N)} \in \tilde{b}^+$, $x^{(N)} \notin \tilde{a}^+$, and $x^{(N)} \notin \tilde{c}^-$ for sufficiently large N we can take $x^{(N)} = (b_0 - \frac{1}{2}\tau(a, b), \mathbf{b}) + (N, N, 0, 0)$. The proof is similar to the case $\mathbf{a} \neq \mathbf{b}$. The lemma is proved.

Proof of the Theorem. 1. Since f preserves C, f also preserves the approximate future and past cones. This means that $b \in \tilde{a}^+$ iff $f(b) \in f(\tilde{a})^+$ and $b \in \tilde{a}^-$ iff $f(b) \in f(a)^-$. Therefore, if we define $a \tilde{<} b$ as $\exists c(\tilde{b}^+ \subseteq \tilde{a}^+ \cup \tilde{c}^-)$, then we conclude that $a \tilde{<} b$ if $f(a) \tilde{<} f(b)$.

By $C(a)$ let us denote the set $\{b | a \tilde{<} b\}$. Then $b \in C(a)$ iff $f(b) \in C(f(a))$ and, according to the lemma, $a \ll b \to b \in C(a)$ and $b \in C(a) \to (a < b \vee a = b)$.

2. Since f is continuous and its inverse mapping is also continuous, we conclude that $b \in \overline{C(a)}$ (where $\bar{X}$ denotes a closure of the set X) iff $f(b) \in \overline{C(f(a))}$.

3. Because of the lemma, $\{b | a \ll b\} \subseteq C(a) \subseteq \{b | a < b \vee a = b\}$. But it is known that the closures of these two sets $\{b | a \ll b\}$ and $\{b | a < b \vee a = b\}$ coincide with $a^+ \cup \{a\}$. Therefore, $\overline{C(a)} = a^+ \cup \{a\}$. So condition 2 can be written as follows: $b \in a^+ \cup \{a\}$ iff $f(b) \in f(a)^+ \cup \{f(a)\}$. Since f is a 1–1 mapping, from this condition we conclude that $b \in a^+$ iff $f(b) \in f(a)^+$, i.e., $a < b$ iff $f(a) < f(b)$. Therefore, we can apply Alexandrov's theorem to prove that f is linear, and that f is a composition of a shift, a dilation, and a Lorentz transformation. QED

ACKNOWLEDGMENTS

This work was partially sponsored by NSF grant GDA-9015006. The author is thankful to A. D. Alexandrov for a discussion of the first version of this paper, and especially to A. V. Kuz'minykh, who actively participated in the mathematical discussions that led to the proofs, and whose comments helped to make the final text better. From a mathematical viewpoint, he is truly a coauthor of this paper's result.

REFERENCES

Alexandrov, A. D. (1950). *Uspekhi Mathematicheskikh Nauk*, **5**(3, 37), 187 [in Russian].

Alexandrov, A. D., and Ovchinnikova, V. V. (1953). *Leningrad University Vestnik*, **11**, 94–110 [in Russian].

Benz, W. (1977). *Journal of Geometry*, **10**, 45–56.

Brumberg, V. A. (1991). *Essential Relativistic Celestial Mechanics*, Adam Hilger, Bristol, England.

Cannon, W. H., Lisewski, D., Finkelstein, A. M., and Kreinovich, V. (1986). Relativistic effects in Earth based and cosmic long baseline interferometry, in *Proceedings of the IAU Symposium 114 "Relativity in Celestial Mechanics and Astrometry,"* Reidel, Dordrecht, pp. 255–256.

Finkelstein, A. M., and Kreinovich, V. (1976). *Celestial Mechanics*, **13**(2), 151–176.

Finkelstein, A. M., Kreinovich, V., and Pandey, S. N. (1983). *Astrophysics and Space Science*, **94**, 233–247.

Kosheleva, O. M., Kreinovich, V., and Vroegindewey, P. G. (1979*a*). *Proceedings of the Royal Academy of Science of the Netherlands, Series A*, **82**(3), 363–376.

Kosheleva, O. M., Kreinovich V., and Vroegindewey, P. G. (1979*b*). Note on a physical application of the main theorem of chronogeometry, Technological University, Eindhoven, Netherlands.

Kuz'minykh, A. V. (1975). *Soviet Mathematics Doklady*, **16**(6), 1626–1628.

Kuz'minykh, A. V. (1976). *Siberian Mathematical Journal*, **17**(6), 968–972.

Lester, J. A. (1977*a*). *Proceedings of the Cambridge Mathematical Society*, **81**, 455–462.

Lester, J. A. (1977*b*). *Canadian Journal of Mathematics*, **29**, 1247–1253.

Lester, J. A. (1983). *Annals of Discrete Mathematics*, **18**, 567–574.

Naber, G. L. (1992). *The Geometry of Minkowski Space-Time*, Springer-Verlag, New York.

Zeeman, E. C. (1964). *Journal of Mathematical Physics*, **5**, 490–493.