AIP | Journal of Mathematical Physics

Causality Implies the Lorentz Group

E. C. Zeeman

Citation: J. Math. Phys. 5, 490 (1964); doi: 10.1063/1.1704140
View online: http://dx.doi.org/10.1063/1.1704140
View Table of Contents: http://jmp.aip.org/resource/1/JMAPAQ/v5/i4
Published by the American Institute of Physics.

Causality Implies the Lorentz Group

E. C. Zeeman
Institut des Hautes Études Scientifiques, Bures-sur-Yvette, Seine et Oise, France
(Received 30 July 1963)

Causality is represented by a partial ordering on Minkowski space, and the group of all automorphisms that preserve this partial ordering is shown to be generated by the inhomogeneous Lorentz group and dilatations.

Let M denote Minkowski space, the real 4-dimensional space–time continuum of special relativity, and let Q denote the characteristic quadratic form on M,

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

There is a partial ordering on M given by $x < y$ if an event at x can influence an event at y; more precisely, $x<y$ if $y-x$ is a time vector, $Q(y-x)>0$, oriented towards the future, $x_0 < y_0$. Let $f : M \to M$ be a function that is a one-to-one mapping (*we make no assumptions that f is linear or continuous*). We call f a *causal automorphism* if both f and f^{-1} preserve the partial ordering; in other words,

$$x < y \Leftrightarrow fx < fy, \quad \text{all} \quad x, y \in M.$$

The causal automorphisms form a group, which we call the *causality group.*

Let G be the group generated by (i) the orthochronous Lorentz group (linear maps of M that leave Q invariant, and preserve time orientation, but possibly reverse space orientation), (ii) translations of M, and (iii) dilatations of M (multiplication by a scalar).

Theorem. The causality group = G.

Remark 1. The significance of the theorem is that if we interpret the principle of causality mathematically as the set M together with the partial ordering, then the inhomogeneous Lorentz group appears naturally (with dilatations and space reversal) as the symmetry group of M. Therefore the basic invariants of physics, which are the representations of the inhomogeneous Lorentz group, follow natually from the single principle of causality.

Remark 2. It is easy to see that G is contained in the causality group, since the generators of G preserve the partial ordering. The converse is not obvious at first sight, because there seems no reason why a causal automorphism should be linear or even continuous. In fact, the result depends essentially upon space being more than 1-dimensional. If space were 1-dimensional then the causality group would be much larger than G, and the general causal automorphism would map the space and time axes into curved lines, as is shown by the example below. Thus the typical 2-dimensional picture of Minkowski space to be found in most textbooks is misleading.

Remark 3. The condition for f to be a causal automorphism is a global condition, but is equivalent (by an elementary compactness argument using the transitivity of $<$) to the following local condition: given $x \in M$, then there is a neighborhood N of x such that

$$y < z \Leftrightarrow fy < fz, \quad \text{all} \quad y, z \in N.$$

Intuitively this means we need only think of the principle of causality acting in our laboratories for a few seconds, rather than between distant galaxies forever, and still we are able to deduce the Lorentz group.

Remark 4. There is another relation on M given by $x <\cdot\, y$ if light can go from x to y; more precisely $x <\cdot\, y$ if $y - x$ is a light vector, $Q(y - x) = 0$, oriented towards the future, $x_0 < y_0$. The relation $x <\cdot\, y$ is not a partial ordering because it is not transitive,

$$x <\cdot\, y <\cdot\, z \not\Rightarrow x <\cdot\, z.$$

We shall show in Lemma 1 that, in the definition of causal automorphism, it does not matter whether we use $<$ or $<\cdot$ (or both). Intuitively this means that the Lorentz group can be deduced equally well either from causality between heavy particles, or from causality between photons, or from both. Remark 3 also holds for $<\cdot$, although the argument is slightly more complicated due to the lack of transitivity.

Remark 5. In Remarks 2 and 3, when referring to the "continuity" of f or the "neighborhood" of x, we have implicitly assumed a topology on M, although for the proof of the theorem we assume no topology. It is customary to think of M as having the topology of real 4-dimensional Euclidean space, but there are reasons why this is wrong. In particular:

(i) Euclidean topology is locally homogeneous whereas M is not; every point has its associated light cone separating space vectors from time vectors.

(ii) The group of all homeomorphisms of Euclidean space is vast, and of no physical significance.

In a subsequent paper[1] we suggest alternative topologies for M, which are not homogeneous, and have the property that any homeomorphism maps light cones to light cones. Therefore any homeomorphism preserves or reverses the relation $<\cdot$, and so the group of all homeomorphisms of M with such a topology will be the double cover of G. Consequently, the topology is physically significant because it implies the Lorentz group.

The existence of such topologies on Minkowski space suggest the possibility of similar topologies on the inhomogeneous Lorentz group, finer than the Lie-group topology. Any representation with the Lie-group topology would *a fortiori* be a representation with a finer topology, but not necessarily conversely. This raises the question: are there some new representations of the inhomogeneous Lorentz group?

Example.

Let K denote 2-dimensional Minkowski space with characteristic quadratic form

$$Q(x) = x_0^2 - x_1^2, \qquad x = (x_0, x_1) \in K.$$

Choose new coordinates

$$y_0 = x_0 - x_1, \qquad y_1 = x_0 + x_1.$$

Let $f_0, f_1 : R \to R$ be two arbitrary nonlinear orientation-preserving homeomorphisms of the real line onto itself. Define $f : K \to K$ by

$$f(y_0, y_1) = (f_0 y_0, f_1 y_1).$$

Then f is a causal automorphism, but $f \notin G$ because f is nonlinear. In general, the images of the space and time axes will not be straight lines.

Lemma 1. *Let $f : M \to M$ be a function that is a one-to-one mapping. Then f, f^{-1} preserve the partial ordering $<$ if and only if they preserve the relation $<\cdot$.*

[1] E. C. Zeeman, "The topology of Minkowski space" (to be published).

Proof. If $x < y$ implies $f^{-1}x < f^{-1}y$, then $x \nless y$ implies $fx \nless fy$. Therefore if f, f^{-1} preserve $<$, then f preserves $<$ and $\nless$. Now

$$x <\cdot\, y \Leftrightarrow \begin{cases} x \nless y \\ y < z \Rightarrow x < z. \end{cases}$$

Therefore if f preserves $<$ and $\nless$, then f preserves $<\cdot$. Therefore if f, f^{-1} preserve $<$, then f, f^{-1} preserve $<\cdot$. Conversely,

$$x < y \Leftrightarrow \begin{cases} x \nless\cdot\, y \\ x <\cdot\, z <\cdot\, y, \quad \text{for some} \quad z. \end{cases}$$

Therefore if f preserves $<\cdot$ and $\nless\cdot$, then f preserves $<$, and so if f, f^{-1} preserve $<\cdot$, then f, f^{-1} preserve $<$.

Notation.

If $x \in M$, let C_x denote the *light cone* through x,

$$C_x = \{y; x <\cdot\, y \quad \text{or} \quad x = y \quad \text{or} \quad y <\cdot\, x\}.$$

If $x <\cdot\, y$, we call the line through x and y a *light ray* and denote it by $R_{x,y}$. We deduce

$$R_{x,y} = C_x \cap C_y.$$

Lemma 2. A causal automorphism maps light rays to light rays.

Proof: Let f be a causal automorphism. By Lemma 1, f and f^{-1} preserve $<\cdot$, and so $fC_x = C_{fx}$. Therefore if $x <\cdot\, y$,

$$fR_{x,y} = f(C_x \cap C_y) = C_{fx} \cap C_{fy} = R_{fx,fy}.$$

Lemma 3. A causal automorphism maps parallel light rays to parallel light rays.

Proof: Let a_1, a_2 be parallel light rays, and let P be the plane through them. There are two cases according to whether or not P is a tangent to all the light cones with vertex in P.

Case (1). Suppose P is not a tangent (this is the usual case). Then P contains two families $\{a\}$, $\{b\}$ of light rays, where $\{a\}$ consists of all lines parallel to a_1 (and a_2), and $\{b\}$ consists of all lines parallel to another direction. If $x \in P$, then the light cone with vertex x meets P in the two light rays through x, one from each family.

Let f be a causal automorphism. The images $\{fa\}$, $\{fb\}$ are families of lines with the property that each fa meets each fb, but no two of any one family meet. There are two possibilities; (i) fa_1 and fa_2 are coplanar, or (ii) they are not. We shall show that (ii) leads to a contradiction. For if fa_1, fa_2 are not coplanar they lie in a 3-dimensional subspace S, say, of M. Then each $fb \subset S$, because it meets fa_1 and fa_2, and therefore each $fa \subset S$. Therefore the two families are generators of a nondegenerate quadric surface in S. If this quadric surface is a hyperboloid (meeting S_∞ in a conic, where S_∞ denotes the plane at infinity), then each fa is parallel to some fb (the unique fb through $fa \cap S_\infty$), contradicting that fa meets fb (in a finite point). Alternatively, if the quadric is a paraboloid (meeting S_∞ in two lines), then the directions of $\{fa\}$ are parallel to all the lines in a plane (which meets S_∞ in one of the lines). But all light rays are parallel to the rays in a single light cone (which meets M_∞ in a sphere), and so any plane contains at most two lines parallel to light rays (through the points in M_∞ where the sphere meets the line in $S_\infty \subset M_\infty$), and so again we have a contradiction.

Therefore fa_1, fa_2 must be coplanar, and, since they do not meet, must be parallel.

Case (2). Suppose P is a tangent to all light cones with vertex in P (this is the exceptional case). The argument of Case (1) breaks down because P has the property that it contains only one family of light rays, namely all the lines parallel to a_1 and a_2. The planes through a_1 with this property span a 3-dimensional subspace A_1, say, of M (the tangent prime to the light cones through a_1). Similarly the planes through a_2 with the property span A_2. Choose a_3 parallel to a_1 and a_2, and not in $A_1 \cup A_2$. Then by Case (1), a_1 and a_2 are both parallel to a_3, and hence parallel to each other. The proof of Lemma 3 is complete.

Remark. So far, everything we have done applies to the 2-dimensional example above. As yet we have not proved that f maps each light ray linearly, nor have we proved that f maps straight lines other than light rays into straight lines. We prove this in the next lemma, using the fact that the dimension of space is greater than 1.

Lemma 4. A causal automorphism maps each light ray linearly.

Proof: Suppose a, a_1 are parallel light rays, as in Case (1) of Lemma 3. The family $\{b\}$ of parallel light rays meeting a and a_1 determine a linear map $g_1 : a \to a_1$, and if f is a causal automorphism the image family $\{fb\}$ determine a linear map $e_1 : fa \to fa_1$ such that the diagram

$$\begin{array}{ccc} a & \xrightarrow{f} & fa \\ \downarrow{\scriptstyle g_1} & f & \downarrow{\scriptstyle e_1} \\ a_1 & \longrightarrow & fa_1 \end{array}$$

is commutative. If a_2 is also parallel to a_1, we can define similar maps g_2, e_2 for the pair a_1, a_2, and maps g_3, e_3 for the pair a_2, a, provided neither of the pairs is exceptional as in Case (2) of Lemma 3. Composing the three diagrams gives a commutative diagram

$$\begin{array}{ccc} a & \xrightarrow{f} & fa \\ \downarrow{\scriptstyle g} & f & \downarrow{\scriptstyle e} \\ a & \longrightarrow & fa \end{array}$$

where $g = g_3g_2g_1$ is a translation of a, and $e = e_3e_2e_1$ is a translation of fa (g, e are translations because they are compositions of parallel displacements).

If Minkowski space were 2-dimensional, then any such translations would have to be the identity. But in higher dimensions—and this is where the difference is essential—we claim that any given translation g of a can be obtained in this manner. It suffices to construct an arbitrary translation on one particular light ray, for then the result will be true for all light rays since G is transitive on the set of all light rays.

Let $x = (0, 0, 0, 0)$, $y = (0, -t, 0, t)$, $z = (0, 0, 0, 2t)$ and $x^* = (0, 0, t, t)$. Let a, a_1, a_2 be the light rays through x, y, z, respectively, parallel to the direction $[0, 0, 1, 1]$.
Then

$$x \xrightarrow[g_1]{} y \xrightarrow[g_2]{} z \xrightarrow[g_3]{} x^*.$$

Therefore $gx = x^*$, and the given translation g can be obtained by suitable choice of the parameter t.

Let r, s be coordinates chosen on a, fa such that $f(0) = 0$. Suppose that when g is the translation $r \to r + t$, then e (which is uniquely determined by g) is the translation $s \to s + u$, where $u = u(t)$. Then

$$f(r + t) = fg(r) = ef(r) = f(r) + u(t),$$

for all r, t. Putting $r = 0$, we have

$$f(t) = u(t),$$

and so

$$f(r + t) = f(r) + f(t).$$

Therefore by induction $f(nt) = nf(t)$, for positive and negative integers n. If m is also an integer, then $nf[(m/n)t] = f(mt) = mf(t)$, and so

$$f(rt) = rf(t)$$

for r rational. But the last equation is also true for r real, because f preserves $< \cdot$, and so is orderpreserving on each light ray. Hence f is linear on the light ray, and Lemma 4 is proved.

Lemma 5. A causal automorphism maps parallel equal intervals on light rays to parallel equal intervals.

Proof: Parallel light rays must be mapped with the same linear expansion because the family of parallel light rays meeting them both also remains parallel. (In the exceptional case use a third ray, as in the proof of Lemma 3.) Therefore equal intervals are mapped with the same linear expansion onto equal intervals.

Proof of the Theorem: We are given a causal automorphism $f : M \to M$. We can assume the f keeps the origin fixed, by first composing f with a translation if necessary. Choose four linearly independent vectors v_1, v_2, v_3, v_4 directed along four light rays through the origin: these form a base for the vector-space structure of M, and so an arbitrary vector $x \in M$ can be written

$$x = \sum x_i v_i, \qquad x_i \text{ scalar.}$$

Let $g : M \to M$ be the linear map given by

$$gx = x_i(fv_i).$$

We shall show that f is linear by proving that $f = g$. For each i, $1 \leq i \leq 4$, let M_i denote the i-dimensional vector subspace spanned by v_j, $1 \leq j \leq i$. We shall show that $f = g$ on M_i by induction on i. The induction starts with $i = 1$ by Lemma 4, and finishes with $i = 4$. Assume the induction for $i - 1$. Given $x \in M_i$, write $x = y + x_i v_i$, where $y \in M_{i-1}$. Then the interval from y to x is parallel and equal in length to the vector $x_i v_i$. By Lemma 5 the interval from fy to fx is parallel and equal in length to $f(x_i v_i)$. Therefore

$$\begin{aligned} fx &= fy + f(x_i v_i) \\ &= gy + g(x_i v_i), \text{ by induction and by Lemma 4,} \\ &= gx, \text{ because } g \text{ is linear.} \end{aligned}$$

This completes the inductive step, and the proof that f is linear.

Since f preserves $<\cdot$, the light cone, $Q(x) = 0$, through the origin is kept fixed. Therefore, multiplying f by a scalar if necessary, we deduce that f leaves Q invariant. In other words, in modulo multiplication by a translation and a dilatation, f is a time-orientation-preserving element of the Lorentz group. Therefore $f \in G$, and the proof of the theorem is complete.

ACKNOWLEDGMENT

The author is indebted to the physicists of the Institut des Hautes Études Scientifiques, and in particular to F. Lurçat, for their patience in explaining to him which mathematics in modern physics is significant.