Special Relativity without Physics

WASHEK F. PFEFFER [*]

SUMMARY. - *Using only causality and the constant speed of light, I derive the Poincaré transformation group. In this derivation I make no a priori assumptions about the linearity or continuity of the transformations.*

By "understanding special relativity" I mean understanding how the coordinate systems associated with different observers transform into each other. In my opinion, this rather limited concept of "understanding special relativity" is the most fundamental. Indeed, I show by examples that once the coordinate transformations are properly understood, the standard special relativistic phenomena such as length contraction, slowing of clocks, and Einstein's law for addition of velocities follow with a minimal effort.

Although, the coordinate transformations are derived in virtually any textbook on special relativity, these derivations have much to be desired. Most of them are based on highly restrictive and unintuitive assumptions, such as linearity or preservation of the space-time intervals, which are justified by physical reasoning of questionable rigor. In contrast, my objective is to formulate precisely a few simple physical principles from which the coordinate transformations are derived in a purely deductive manner.

I am obliged to Karel Kuchař for valuable discussions concerning the physical aspects of special relativity.

---

[*] Author's address: Department of Mathematics, University of California, Davis, CA 95616, USA, e-mail: wfpfeffer@ucdavis.edu

1. The Poincaré group

We shall use two primitive (i.e., undefined) notions: *event* and *coincidence of events*. Intuitively, event is when something happens, such as the emission or absorption of a photon. The intuitive meaning of coinciding events is that these events occur simultaneously at the same place, for instance, a simultaneous emission of photons from the same place in different directions. We assume that all events form a set U, called the *universe*, and that the coincidence of events is an equivalence relation on U, denoted by $\sim$. Our main object of study is the family

$$E = U/\sim$$

of all equivalence classes of coinciding events, called the *space-time*.

We assume in E there is a binary relation $\prec$, called the *light connection*. Intuitively, a point $v \in E$ is light connected to a point $u \in E$, in writing $u \prec v$ if the equivalence class u is determined by the emission of a photon $\hbar$ (*light signal*), and the equivalence class v is determined by the absorption of the same photon $\hbar$. For each $u \in E$, the set

$$C(u) = \{v \in E : u \prec v \text{ or } v \prec u\}$$

is called the *light cone* at u, and the sets

$$C_+(u) = \{v \in E : u \prec v\} \quad \text{and} \quad C_-(u) = \{v \in E : v \prec u\}$$

are called the *forward* and *backward light cones* at u. In essence, our approach to special relativity is based on a single assumption: **the light cones are the only physical reality detectable by an observer**. Naturally, an observer may detect other physical phenomena too, but those we shall ignore in our exposition.

By $\mathbb{R}$ we denote the set of all real numbers. For an integer $m \geq 2$, we denote by $\mathbb{R}^m$ the m-fold Cartesian product of $\mathbb{R}$. We shall be mainly concerned with $\mathbb{R}^4$ in which the first three coordinates are viewed as *spatial coordinates* and the forth coordinate is viewed as a *time coordinate*. For $x = [x_1, x_2, x_3, x_4]$ in $\mathbb{R}^4$, we let

$$x^2 = (x_1)^2 + (x_2)^2 + (x_3)^2 - (x_4)^2 .$$

Intuitively speaking, an *observer* is a person, named α, equipped with a measuring tape and a clock (both broadly interpreted), who tries to associate with each class of equivalent events a place and time represented by a point $x \in \mathbb{R}^4$. This association is not arbitrary, since it must reflect the light connection. Thus α associates points of $\mathbb{R}^4$ with points of the space-time E so that the following condition is met: if

$$x = [x_1, x_2, x_3, x_4] \quad \text{and} \quad y = [y_1, y_2, y_3, y_4]$$

are associated, respectively, with the emission of a photon $\hbar$ and the absorption of the same photon $\hbar$, then

$$\sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2} = c(y_4 - x_4)$$

where c is the average speed with which the photon $\hbar$ travels from the time of its emission to the time of its absorption. The next definition and axiom give a precise mathematical formulation to our intuition.

DEFINITION 1.1. *An observer is a bijection* $\alpha = [\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ *from* E *to* $\mathbb{R}^4$ *such that*

$$u \prec v \iff \left\{ \big[\alpha(v) - \alpha(u)\big]^2 = 0 \text{ and } \alpha_4(u) \leq \alpha_4(v) \right\}$$

for each u *and* v *in* E.

We note that what we call an observer is often referred to as the *coordinate system* or *reference frame* of an observer.

AXIOM 1.2. *The set of all observers is not empty.*

REMARK 1.3. *Postulating the existence of an observer encompasses important physical claims.*

(i) *The* speed of light *is* constant *and equal to* 1. *The constancy of the speed of light is crucial; the fact we made the speed of light* 1 *is merely a matter of convenience achieved by a suitable choice of units.*

(ii) *The* causality, *i.e., the direction of time flow, holds between any points of space-time which are light connected. In short, it says a photon has to be emitted before it is absorbed.*

*(iii) The space-time E has the structure of a four-dimensional linear
     space over* $\mathbb{R}$*.*

Claims (i)–(iii) are the *physical assumptions* underlying special
relativity. They are *physical* in the sense of being independent of
the choice of an observer. In *vacuum*, claim (i) has been verified
locally to a high degree of accuracy. Claim (ii) is extrapolated from
time-wise irreversible processes, such as those encountered in ther-
modynamics. On the other hand, claim (iii) is a deliberate simplifi-
cation which, strictly speaking, holds only in the universe completely
void of any matter (including gravitational fields). However, from a
practical point of view, claim (iii) is still a good approximation of
physical reality in the absence of strong gravitational fields. For in-
stance, claim (iii) is a useful assumption in the study of microscopic
phenomena, in particular, in quantum mechanics. Investigating the
actual structure of the space-time is the subject of *general relativity*,
which is beyond the scope of my lectures.

Note an observer α transforms the light cone $C(u)$ at $u \in E$ into
a genuine cone in $\mathbb{R}^4$ with the vertex $x = \alpha(u)$. Indeed,

$$\alpha\big[C(u)\big] = \big\{y \in \mathbb{R}^4 : (y - x)^2 = 0\big\}$$

is a quadric in $\mathbb{R}^4$ whose only singular point is x. The set $\mathbb{R}^4 -
\alpha\big[C(u)\big]$ has two connected components

$$\big\{y \in \mathbb{R}^4 : (y - x)^2 < 0\big\} \quad \text{and} \quad \big\{y \in \mathbb{R}^4 : (y - x)^2 > 0\big\}.$$

If $(y - x)^2 < 0$ and $y_4 > x_4$, then

$$\sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2} = s(y_4 - x_4)$$

where $0 \le s < 1$. The physical interpretation of this is a particle
emitted from the space-time point $\alpha^{-1}(x)$ that travels with a speed
s, lesser than the speed of light, is absorbed at the space-time point
$\alpha^{-1}(y)$. Since massive particles can travel with any speed that is
lesser than the speed of light, we say the space-time points $\alpha^{-1}(x)$
and $\alpha^{-1}(y)$ are *mass connected*.

The following observation is an immediate consequence of Defi-
nition 1.1 and Axiom 1.2. It shows the relation $\prec$ of light connection
is antisymmetric.

PROPOSITION 1.4. *For each pair $u, v \in E$, we have*

$$\{u \prec v \text{ and } v \prec u\} \iff u = v\,.$$

Proof. If $\alpha : E \to \mathbb{R}^4$ is an observer, then $\alpha_4(u) = \alpha_4(v)$, and hence

$$\sum_{i=1}^{3} \big[\alpha_i(u) - \alpha_i(v)\big]^2 = 0\,.$$

It follows $\alpha(u) = \alpha(v)$, and since α is a bijection, $u = v$. $\qquad\square$

Given two observers α and β, we have the following commutative diagram

$$
\begin{array}{ccc}
E & =\!=\!=\!= & E \\
{\scriptstyle\alpha}\big\downarrow & & \big\downarrow{\scriptstyle\beta} \\
\mathbb{R}^4 & \xrightarrow{\ \ f\ \ } & \mathbb{R}^4
\end{array}
$$

where $f = \beta \circ \alpha^{-1}$ is a bijection of $\mathbb{R}^4$. The map f is the coordinate transformation whose understanding is critical for any communication between observers α and β. Assuming $f : x \mapsto x'$, it is easy to see

$$
\begin{aligned}
(x - y)^2 &= 0 \iff (x' - y')^2 = 0\,, \\
(x - y)^2 &= 0 \implies \operatorname{sign}(x_4 - y_4) = \operatorname{sign}(x_4' - y_4')\,.
\end{aligned}
\tag{1}
$$

THEOREM 1.5. *Let $m \geq 3$ be an integer, and let Q be an indefinite non-singular quadratic form in $\mathbb{R}^m$. Assume $f : x \mapsto x'$ is a bijection of $\mathbb{R}^m$ such that*

$$Q(x' - y') = 0 \iff Q(x - y) = 0$$

for all $x, y \in \mathbb{R}^m$. Then f is an affine map, and there is a real number $k \neq 0$ such that

$$Q(y' - x') = kQ(y - x)$$

for all $x, y \in \mathbb{R}^m$. If the signature of Q differs from zero, then $k > 0$.

Theorem 1.5 was proved at various levels of generality by several authors. The present version was established in 1962 by V. KNICHAL, who never published his result. For $m = 4$ and $Q(x) = x^2$, Theorem 1.5 is proved in [9, 2]. For any $m \geq 3$ and

$$Q(x) = (x_1)^2 - \sum_{i=2}^{m} (x_i)^2 \,,$$

a proof can be found in [4]. No proof is easy, and I shall not attempt it at this point. However, in Section 5 below, I outline the proof of a stronger result, from which Theorem 1.5 follows.

Theorem 1.5 is **false** if Q is positively definite or if $m = 2$. The first claim is obvious, since $Q(x) = 0$ implies $x = 0$ whenever Q is positively definite. To see the second claim, consider a nonsingular indefinite quadratic form

$$Q(x) = (x_1)^2 - (x_2)^2$$

in $\mathbb{R}^2$, and a nonlinear bijection

$$[x_1, x_2] \mapsto \left[(x_1)^3, (x_2)^3 \right] : \mathbb{R}^2 \to \mathbb{R}^2 \,.$$

This is interesting, since in many textbooks on special relativity, the Lorentz transformation is derived for the case of one spatial coordinate only, i.e., in $\mathbb{R}^2$. Such derivations are either incorrect, or they require additional assumptions.

DEFINITION 1.6. *The* Poincaré *group* $\mathcal{P}$ *is the group of all bijective affine maps* $f : x \mapsto x'$ *of* $\mathbb{R}^4$ *such that for a constant* $k(f) > 0$ *and all* $x, y \in \mathbb{R}^4$, *the following conditions are satisfied:*

$$(x' - y')^2 = k(f)(x - y)^2 \,,$$
$$(x - y)^2 = 0 \Longrightarrow \mathrm{sign}(x_4' - y_4') = \mathrm{sign}(x_4 - y_4) \,.$$

Elements of the Poincaré group are called Poincaré *transformations.*

The following proposition is an immediate consequence of equations (1) and Theorem 1.5.

PROPOSITION 1.7. *If* α *and* β *are observers if and only if* $\beta \circ \alpha^{-1}$ *belongs to* $\mathcal{P}$.

2. The Lorentz group

DEFINITION 2.1. *The* Lorentz group $\mathcal{L}$ *is the group of all linear Poincaré transformations*

$$x \mapsto x' : \mathbb{R}^4 \to \mathbb{R}^4$$

such that $(x')^2 = x^2$ *for each* $x \in \mathbb{R}^4$. *Elements of the Lorentz group are called* Lorentz transformations.

A *normalization* is a map

$$x \mapsto cx + z : \mathbb{R}^4 \to \mathbb{R}^4$$

where $c > 0$ and $z \in \mathbb{R}^4$. It is easy to verify the family $\mathcal{N}$ of all normalizations is a normal subgroup of $\mathcal{P}$. We denote by $\mathbf{0}$ and $\mathbf{1}$ the zero vector in $\mathbb{R}^4$ and the identity map of $\mathbb{R}^4$, respectively.

PROPOSITION 2.2. *There is a split short exact sequence*

$$1 \longrightarrow \mathcal{N} \overset{\subset}{\longrightarrow} \mathcal{P} \overset{\Phi}{\longrightarrow} \mathcal{L} \longrightarrow 1 \qquad (2)$$

where

$$\Phi(f) : x \mapsto \frac{1}{\sqrt{k(f)}} \big[f(x) - f(\mathbf{0}) \big] : \mathbb{R}^4 \to \mathbb{R}^4 .$$

Proof. A straight forward calculation shows that $\Phi(f) \in \mathcal{L}$ for each $f \in \mathcal{P}$, and that $\Phi(g) = g$ for every $g \in \mathcal{L}$. Moreover, $\Phi(f) = \mathbf{1}$ if and only if

$$x = \frac{1}{\sqrt{k(f)}} \big[f(x) - f(\mathbf{0}) \big]$$

for all $x \in \mathbb{R}^4$, or equivalently, if and only if $f(x) = \sqrt{k(f)}x + f(\mathbf{0})$ is a normalization. Thus (2) is a short exact sequence, which splits because the inclusion map $\mathcal{L} \hookrightarrow \mathcal{P}$ is the right inverse of Φ. $\qquad \square$

From the physical point of view, two observers are related by a Lorentz transformation whenever both of them

- map the same point of the space-time to the zero vector $\mathbf{0}$ of $\mathbb{R}^4$,

- choose the same units of length and time.

Since the Lorentz transformations are linear maps of $\mathbb{R}^4$, they can be represented by matrices. We say a 4×4 matrix A is a *Lorentz matrix* if the map $x \mapsto Ax$ is a Lorentz transformation. To facilitate the calculation with matrices, we view the points of $\mathbb{R}^4$ as the column vectors, and if A is a matrix we denote by A^T its transpose. The identity 4×4 matrix is denoted by I, and we let

$$J = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

LEMMA 2.3. *If A is a 4×4 matrix, then $(Ax)^2 = x^2$ for each $x \in \mathbb{R}^4$ if and only if $A^T J A = J$, or equivalently $A J A^T = J$.*

Proof. Recall there is a one-to-one correspondence between the quadratic forms $x \mapsto x^T A x$ and symmetric matrices A. Since J and $A^T J A$ are the symmetric matrices of the quadratic forms $x \mapsto x^2$ and $x \mapsto (Ax)^2$, respectively, we have $(Ax)^2 = x^2$ if and only if $A^T J A = J$. Moreover, in the following string of equalities each is equivalent to the next:

$$\begin{aligned} A^T J A &= J \\ J A^T J A &= I \\ J A^T J &= A^{-1} \\ A J A^T J &= I \\ A J A^T &= J \end{aligned}$$

$\square$

LEMMA 2.4. *Let $A = (a_{ij})$ be a matrix such that $A^T J A = J$. Then $|a_{44}| \geq 1$ and*

$$\mathrm{sign}(Ax)_4 = \mathrm{sign}(a_{44} x_4)$$

for each $x \in \mathbb{R}^4$ with $x^2 \leq 0$.

Proof. Since Lemma 2.3 implies $A J A^T = J$, we obtain

$$(a_{41})^2 + (a_{42})^2 + (a_{43})^2 - (a_{44})^2 = -1\,,$$

and consequently $(a_{44})^2 \geq 1$. Now select an $x \in \mathbb{R}^4$ with $x^2 \leq 0$. As

$$x'_4 = \sum_{i=1}^{4} a_{4i} x_i \,,$$

Schwartz's inequality yields

$$
\begin{aligned}
(x'_4 - a_{44} x_4)^2 &= (a_{41} x_1 + a_{42} x_2 + a_{43} x_3)^2 \\
&\leq \left[(a_{41})^2 + (a_{42})^2 + (a_{43})^2 \right] \cdot \left[(x_1)^2 + (x_2)^2 + (x_3)^2 \right] \\
&\leq \left[(a_{44})^2 - 1 \right] (x_4)^2
\end{aligned}
$$

Thus $x'_4 = 0$ whenever $x_4 = 0$, and

$$|x'_4 - a_{44} x_4| \leq \left[(a_{44})^2 - 1 \right]^{\frac{1}{2}} |x_4| < |a_{44} x_4|$$

whenever $x_4 \neq 0$. The lemma follows. $\square$

From Lemmas 2.3 and 2.4 we obtain immediately the following corollary.

COROLLARY 2.5. *A 4×4 matrix $A = (a_{ij})$ is a Lorentz matrix if and only if $A^T J A = J$ and $a_{44} \geq 1$, or equivalently, if and only if $A^T J A = J$ and*

$$\operatorname{sign} x_4 = \operatorname{sign}(Ax)_4$$

for a single $x \in \mathbb{R}^4$ with $x \neq 0$ and $x^2 \leq 0$.

To appreciate the physical meaning of Lemma 2.4 and Corollary 2.5, call a bijection $\alpha : E \to \mathbb{R}^4$ a *semiobserver* if

$$\{ u \prec v \text{ or } v \prec u \} \iff \left[\alpha(v) - \alpha(u) \right]^2 = 0 \,.$$

Thus a semiobserver recognizes the speed of light is 1, but his flow of time may be mixed up. Nonetheless, since normalizations do not affect causality, Lemma 2.4 shows the flow of time of a semiobserver α is either the correct one, in which case α is an observer, or the reversed one. In view of Corollary 2.5, the causality of a semiobserver can be decided by a single experiment which verifies the flow of time between two distinct points of the space-time that are either light connected or mass connected.

3. Boosts

If u and v are numbers in the open interval $(-1, 1)$, let

$$u \dotplus v = \frac{u + v}{1 + uv} \tag{3}$$

and observe $\{(-1, 1), \dotplus\}$ is an abelian group, denoted by $\mathcal{B}$. For $v \in \mathcal{B}$, a direct calculation reveals

$$A_v = \begin{pmatrix} \frac{1}{\sqrt{1-v^2}} & 0 & 0 & \frac{-v}{\sqrt{1-v^2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{-v}{\sqrt{1-v^2}} & 0 & 0 & \frac{1}{\sqrt{1-v^2}} \end{pmatrix}$$

is a Lorentz matrix, called a *boost*, and

$$A_u A_v = A_{u \dotplus v} \tag{4}$$

for each pair $u, v \in \mathcal{B}$. Thus the map $v \mapsto A_v$ is a homomorphism from $\mathcal{B}$ into the multiplicative group of Lorentz matrices.

A *spatial adjustment*, or simply an *adjustment*, is a linear map of $\mathbb{R}^4$ given by a matrix

$$R = \begin{pmatrix} C & \mathbf{n} \\ \mathbf{n}^T & 1 \end{pmatrix}$$

where C is an orthogonal 3×3 matrix and $\mathbf{n}$ is a 3×1 zero matrix. Note spatial adjustments live up to their name: they adjust the spatial coordinates x_1, x_2, x_3 by rotations and reflections, while leaving the time coordinate x_4 intact. A direct verification shows the family $\mathcal{A}$ of all adjustments is a subgroup of the Lorentz group $\mathcal{L}$.

PROPOSITION 3.1. *Let A be a Lorentz matrix. There are matrices R and S corresponding to adjustments, and a real number v with $|v| < 1$ such that*

$$A = RA_v S \,.$$

Proof. By $a, b, c, \ldots$ and $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$ we denote real numbers and 3×1 matrices, respectively. Keep in mind $\mathbf{n}$ denotes the 3×1 zero matrix. Using this notation, we can write

$$A = \begin{pmatrix} C & \mathbf{a} \\ \mathbf{b}^T & c \end{pmatrix}$$

where C is a 3×3 matrix.

(i) Choose an orthogonal 3×3 matrix $(\mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3)$ so that $\mathbf{b} = b\mathbf{b}_1$ for some $b \in \mathbb{R}$. Let

$$B_1 = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{n} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and observe

$$A_1 = AB_1 = \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \mathbf{a} \\ b & 0 & 0 & c \end{pmatrix}$$

where $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$ are some 3×1 matrices, not necessarily the columns of C.

(ii) Choose an orthogonal 3×3 matrix $(\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3)$ so that $\mathbf{a} = a\mathbf{a}_1$ for some $a \in \mathbb{R}$. Let

$$B_2 = \begin{pmatrix} \mathbf{a}_1^T & 0 \\ \mathbf{a}_2^T & 0 \\ \mathbf{a}_3^T & 0 \\ \mathbf{n}^T & 1 \end{pmatrix}$$

and observe

$$A_2 = B_2 A_1 = \begin{pmatrix} \mathbf{d}_1^T & a \\ \mathbf{d}_2^T & 0 \\ \mathbf{d}_3^T & 0 \\ \mathbf{d}_4^T & c \end{pmatrix}$$

where $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ are some 3×1 matrices and $\mathbf{d}_4^T = (b \quad 0 \quad 0)$.

(iii) Since A, B_1, and B_2 are Lorentz matrices, so is $A_2 = B_2 A B_1$. As Lemma 2.3 implies $A_2 J A_2^T = J$, we infer $\mathbf{d}_1$, $\mathbf{d}_2$, and $\mathbf{d}_3$ are mutually perpendicular vectors, and $\mathbf{d}_2$ and $\mathbf{d}_3$ are unit vectors. Choose a unit vector $\mathbf{e}$ so that $\mathbf{d}_1 = d\mathbf{e}$ for some $d \geq 0$. Let

$$B_3 = \begin{pmatrix} \mathbf{e} & \mathbf{d}_2 & \mathbf{d}_3 & \mathbf{n} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and observe

$$A_3 = A_2 B_3 = \begin{pmatrix} d & 0 & 0 & a \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ p & q & r & c \end{pmatrix}.$$

(iv)   As A_2 and B_3 are Lorentz matrices, so is A_3. Lemma 2.3 yields

$$A_3^T J A_3 = A_3 J A_3^T = J \,,$$

from which we infer

$$r = q = 0 \,, \tag{5}$$
$$d^2 - p^2 = d^2 - a^2 = 1 \,, \tag{6}$$
$$p^2 - c^2 = a^2 - c^2 = -1 \,, \tag{7}$$
$$ad - cp = pd - ac = 0 \,. \tag{8}$$

From equations (6) and (7), we obtain

$$d^2 = 1 + a^2 = c^2 \,,$$

and as $d \geq 0$ by our choice, and $c \geq 1$ by Corollary 2.5, we conclude $d = c$. Hence $a = p$, since equation (8) implies

$$0 = ad - cp = ac - cp = c(a - p)$$

and $c \geq 1$. If $v = -a/c$, then

$$v^2 = \frac{a^2}{c^2} = \frac{c^2 - 1}{c^2} < 1$$

and

$$1 = c^2 - a^2 = c^2 - c^2 v^2 = c^2 \left(1 - v^2\right) \,.$$

Therefore,

$$c = d = \frac{1}{\sqrt{1 - v^2}} \quad \text{and} \quad a = p = -vc = \frac{-v}{\sqrt{1 - v^2}} \,,$$

and it follows from equation (5) that A_3 is the boost A_v. Since $A_v = B_2 A B_1 B_3$ where B_1, B_2, and B_3 are matrices corresponding to adjustments, it suffices to let $R = B_1^{-1}$ and $S = (B_1 B_3)^{-1}$.   □

If $\alpha : E \to \mathbb{R}^4$ is an observer and $g \in \mathcal{N}$ or $g \in \mathcal{A}$, we say the observer $g \circ \alpha : E \to \mathbb{R}^4$ is a *normalization* or an *adjustment* of the observer α, respectively. Using this terminology, we can give Proposition 3.1 a more intuitive formulation.

THEOREM 3.2. *Given observers α and β, there is an adjustment α_1 of α and an adjustment and normalization β_1 of β such that the diagram*

$$
\begin{array}{ccc}
E & =\!=\!=\!= & E \\
\alpha_1 \downarrow & & \downarrow \beta_1 \\
\mathbb{R}^4 & \xrightarrow[A_v]{} & \mathbb{R}^4
\end{array}
$$

commutes for some $v \in \mathbb{R}$ with $|v| < 1$.

Since $\mathcal{N}$ is a normal subgroup of $\mathcal{P}$, in Theorem 3.2 it does not matter whether observer β is first adjusted and then normalized or vice versa. By symmetry, it is also irrelevant which observer is normalized. On the other hand, we shall see that, in general, both observers α and β must be adjusted. To this end, we need to understand the physical meaning of boosts.

Let α and β be suitably adjusted and normalized observers so that the diagram

$$
\begin{array}{ccc}
E & =\!=\!=\!= & E \\
\alpha \downarrow & & \downarrow \beta \\
\mathbb{R}^4 & \xrightarrow[A_v]{} & \mathbb{R}^4
\end{array}
$$

commutes for a $v \in \mathbb{R}$ with $|v| < 1$. To emphasize the forth coordinate is time, we write $x = [x_1, x_2, x_3, t]$ for each $x \in \mathbb{R}^4$. If $x' = A_v x$, then

$$
x_1' = \frac{x_1 - vt}{\sqrt{1 - v^2}}, \quad x_2' = x_2, \quad x_3' = x_3, \quad t' = \frac{t - vx_1}{\sqrt{1 - v^2}}. \quad (9)
$$

View observer α as a person, named α, holding a clock $c(\alpha)$. We may assume α associates the coordinates $[0, 0, 0, t]$ with the event when the clock $c(\alpha)$ shows time t. In view of equations (9), observer β associates with the same event the coordinates

$$
\left[-vt, 0, 0, \frac{t}{\sqrt{1 - v^2}} \right].
$$

It follows α sees β as moving along the x_1-axis with the constant velocity v. By symmetry, β sees α as moving along the x_1-axis with the constant velocity $-v$.

From the above considerations we deduce that any pair of observers move with respect to each other on a line in $\mathbb{R}^3$ with a constant speed less than 1 — the speed of light. In accordance with the *first Newton law*, such observers are usually referred to as *inertial observers*. In order to relate a pair of inertial observers by a boost, it is necessary to make two steps.

(i) **Normalize one** of the observers so that both observers map the same point of the space-time to $\mathbf{0} \in \mathbb{R}^4$, and use the same units of time and length. In other words, by a suitable normalization of one of the observes, we achieve that the observers are related by a Lorentz transformation.

(ii) **Adjust both** observers so that their first coordinate axes lie in the line of their relative motion and point in the same direction, and that their second and third coordinate axes coincide.

Once the parameter v in the boost A_v is interpreted as velocity, it follows from identity (4) that the addition $\dotplus$ defined by equation (3) is the Einstein law for the **addition of velocities**.

For observers α and β related by a boost A_v, we explain two additional relativistic phenomena. Since A_v leaves the coordinates x_2 and x_3 intact, we can reduced our considerations to $\mathbb{R}^2$ where A_v is given by equations

$$x' = \frac{x - vt}{\sqrt{1 - v^2}} \qquad \text{and} \qquad t' = \frac{t - vx}{\sqrt{1 - v^2}} \,. \tag{10}$$

To facilitate ones intuition, it is best to give observers α and β a rather concrete interpretation.

We think of observer α as a person, named α, who holds a clock $c(\alpha)$ and stands by a railroad track lying on the x-axis. We assume the event when $c(\alpha)$ shows time $t = 0$ is mapped by α to $[0,0]$. We further assume that at each spatial point x along the railroad track there is a clock c_x *synchronized* with $c(\alpha)$ in the following sense: if $c(\alpha)$ shows time t (event A), then c_x shows time $t + x$ (event B); note $|x|$ is the time needed for a photon to travel between the spatial points 0 and x. Since $A \prec B$ or $B \prec A$ (according to whether $x \geq 0$ or $x < 0$, respectively), observer α maps event A to $[0,t]$ and event B to $[x, t+x]$.

We think of observer β as a person, named β, standing at the end of a train $\mathbf{T}$ which moves on the railroad track with a constant positive velocity $v < 1$. We assume that β holds a clock $c(\beta)$, and that the event when $c(\beta)$ shows time $t = 0$ is mapped by β to $[0, 0]$. If $\ell' > 0$ is the *rest length* of $\mathbf{T}$ (i.e., the length of $\mathbf{T}$ measured by β), then ℓ' is the spatial coordinate of the front of $\mathbf{T}$ in the reference frame of β.

Since α and β are related by a Lorentz transformation, both α and β map the same space-time point to $[0, 0]$, which plays the role of $\mathbf{0} \in \mathbb{R}^4$ due to the dimension reduction. Consequently the events when the clocks $c(\alpha)$ and $c(\beta)$ show time $t = 0$ coincide. In other words, when α and β are at the same place, the clocks $c(\alpha)$ and $c(\beta)$ show the same time $t = 0$.

Slowing of clocks Let A be the event when the clock $c(\beta)$ shows time $t' > 0$. Then $\beta(A) = [0, t']$, and solving equations (10) reveals

$$\alpha(A) = \left[\frac{vt'}{\sqrt{1 - v^2}}, \frac{t'}{\sqrt{1 - v^2}} \right] \, .$$

Thus at the place $x = vt'/\sqrt{1 - v^2}$ where the clock $c(\beta)$ shows time t', the clock c_x shows time

$$t = \frac{t'}{\sqrt{1 - v^2}} > t' \, .$$

Consequently, the clock $c(\beta)$ is *late* with respect to the clock c_x.

Length contraction Suppose observer α wants to measure the length of $\mathbf{T}$. To this end, α finds two spatial points $x < y$ such that the clocks c_x and c_y show the same time t when the beginning of $\mathbf{T}$ is at y and the end of $\mathbf{T}$ is at x. Then α interprets the number $\ell = y - x$ as the length of $\mathbf{T}$. Assuming $x = t = 0$, the event when the front of $\mathbf{T}$ is at y is mapped by α to $[\ell, 0]$. The same event is mapped by β to $[\ell', t']$ where

$$\ell' = \frac{\ell}{\sqrt{1 - v^2}} \qquad \text{and} \qquad t' = \frac{-v\ell}{\sqrt{1 - v^2}}$$

according to equations (10). Consequently,

$$\ell = \ell' \sqrt{1 - v^2} < \ell' \,,$$

which shows the length ℓ of **T** measured by α is *shorter* than the *rest length* ℓ' of **T** measured by β.

REMARK 3.3. *Theorem 3.2 of its own does not preclude the possibility of two particles moving with a relative speed larger than or equal to* 1 *— it merely says observers cannot be attached simultaneously to these particles. There are dynamical reasons why the relative speed of two massive particles is always less than* 1. *On the other hand, the relative speed of two photons or of a photon and a massive particle equals* 1.

4. The Hilbert space

PROPOSITION 4.1. *Let* $\mathcal{U}$ *and* $\mathcal{V}$ *be maximal orthonormal families in Hilbert spaces* X *and* Y, *respectively. If* X *and* Y *are linearly homeomorphic, then* $\mathcal{U}$ *and* $\mathcal{V}$ *have the same cardinality.*

Proof. As the proposition is a standard result of linear algebra when $\mathcal{U}$ or $\mathcal{V}$ is finite, assume $\mathcal{U}$ and $\mathcal{V}$ are infinite, and select a surjective linear homeomorphism $T : X \to Y$. If (x, y) denote the inner product in Y, the family

$$\mathcal{V}_u = \big\{ v \in \mathcal{V} : (v, Tu) \neq 0 \big\}$$

is countable for each $u \in \mathcal{U}$ by Bessel's inequality. Since the linear hull of $T(\mathcal{U})$ is a dense subset of Y, we see $\mathcal{V} = \bigcup_{u \in \mathcal{U}} \mathcal{V}_u$. Consequently, the cardinality of $\mathcal{V}$ is smaller than or equal to the cardinality of $\mathcal{U}$, and the proposition follows by symmetry. □

An immediate corollary of Proposition 4.1 is that all maximal orthonormal families in a Hilbert space X have the same cardinality, called the *dimension* of X and denoted by $\dim X$. Hilbert spaces X and Y are called *isomorphic* if there is a linear isometry from X onto Y. It is well known and easy to prove two Hilbert spaces are isomorphic if and only if they have the same dimension [8, Section 4.19]. In

view of Proposition 4.1, two Hilbert spaces are isomorphic whenever they are linearly homeomorphic.

Throughout the reminder of my lectures, X will be an arbitrary but fixed **real** Hilbert space of any dimension (possibly finite). If x and y are elements of X, we denote by (x, y) their inner product, and let $|x| = \sqrt{(x, y)}$. The zero vector of X is denoted by $\mathbf{0}$. For a set $A \subset X$, we denote by A^- and $[A]$ the closure and linear hull of A, respectively.

A *functional* on a linear space Z over $\mathbb{R}$ is a real-valued function defined on Z. A functional S on $X \times X$ is called:

- *bilinear* if the functionals

$$S_x : y \mapsto S(x, y) : X \to \mathbb{R} \quad \text{and} \quad S_y : x \mapsto S(x, y) : X \to \mathbb{R}$$

  are linear for all $x, y \in X$;

- *symmetric* if $S(x, y) = S(y, x)$ for all $x, y \in X$;

With each functional S on $X \times X$ we associate a *quadratic functional*

$$x \mapsto S(x, x)$$

on X. If S is symmetric and bilinear, then

$$S(x, y) = \frac{1}{2}\big[S(x + y, x + y) - S(x, x) - S(y, y)\big] \qquad (11)$$

for all $x, y \in X$. Thus a symmetric bilinear functional S on $X \times X$ and the associated quadratic functional $x \mapsto S(x, x)$ on X determine each other.

Given an $x \in X$ and a symmetric bilinear functional S on $X \times X$, we call the set

$$C(x, S) = \big\{y \in X : S(y - x, y - x) = 0\big\}$$

the *light cone* of S at x.

A symmetric bilinear functional Q on $X \times X$ is called a *form* if there is a bijection $T : X \to X$ such that

$$Q(x, y) = (Tx, y)$$

for all $x, y \in X$. Since Q is symmetric,

$$(Tx, y) = Q(x, y) = Q(y, x) = (Ty, x),$$

and we infer T is a linear map whose graph is closed. Thus T is a homeomorphism by the closed graph theorem, and the Schwartz's inequality yields

$$\big|Q(x, y)\big| \leq \|T\| \cdot |x| \cdot |y| \qquad (12)$$

for all $x, y \in X$; here $\|T\|$ denotes the usual norm of T.

We say a form Q is, respectively, *positively* or *negatively definite* in a subspace Y of X if $Q(x, x) > 0$ or $Q(x, x) < 0$ for each $x \in Y$ with $x \neq \mathbf{0}$. A form that is neither positively nor negatively definite in Y, is called *indefinite* in Y.

DEFINITION 4.2. *Let Q be a form. An ordered pair (X_+, X_-) of linear subspaces of X is called a Q-decomposition if*

(i) $X = X_+ \oplus X_-$,

(ii) Q is positively definite in X_+ and negatively definite in X_-,

(iii) $Q(x, y) = 0$ whenever $x \in X_+$ and $y \in X_-$.

The ordered pair $(\dim X_+, \dim X_-)$ is called the signature *of Q.*

While a Q-decomposition is by no means unique, the next proposition shows the signature of a form Q does not depend on the choice of a particular Q-decomposition.

PROPOSITION 4.3. *Let Q be a form, and let (X_+, X_-) be a Q-decomposition. The spaces X_+ and X_- are closed, and if (Y_+, Y_-) is another Q-decomposition, then*

$$\dim Y_+ = \dim X_+ \quad and \quad \dim Y_- = \dim X_- .$$

Proof. Let $\{x_i\}$ be a sequence in X_+ that converges to an $x \in X$, and let $x = x_+ + x_-$ where $x_+ \in X_+$ and $x_- \in X_-$. Since

$$0 = \lim Q(x_i, x_-) = Q(x, x_-) = Q(x_+, x_-) + Q(x_-, x_-) = Q(x_-, x_-),$$

condition (ii) of Definition 4.2 implies $x_- = \mathbf{0}$, and we see $x \in X_+$. A similar argument shows X_- is also closed.

According to condition (i) of Definition 4.2 there are linear maps $P_\pm$ from X onto $X_\pm$ such that

$$x = P_+ x + P_- x$$

for every $x \in X$. The first part of the proof and the closed graph theorem show the maps $P_\pm$ are bounded. Moreover, Definition 4.2, (ii) implies

$$X_+ \cap Y_- = X_- \cap Y_+ = \{\mathbf{0}\} \,,$$

and hence the maps $P_\pm : Y_\pm \to X_\pm$ are injective. Applying the closed graph theorem again, we see these maps are linear homeomorphisms. In particular, $P_\pm(Y_\pm)$ are closed subspaces of $X_\pm$, and Proposition 4.1 implies

$$\dim Y_\pm = \dim P_\pm(Y_\pm) \leq \dim X_\pm \,.$$

The proposition follows by symmetry. □

Let $U : X \to X$ be a linear map such that

$$(Ux, y) = (x, Uy) \quad \text{and} \quad (Ux, x) \geq 0$$

for all $x, y \in X$. By [7, Section 104], there is a unique linear map $V : X \to X$ with

$$(Vx, y) = (x, Vy) \,, \quad (Vx, x) \geq 0 \,, \quad \text{and} \quad V^2 = U \,.$$

We call V the *square root* of U, denoted by $U^{\frac{1}{2}}$. Observe

$$UV = V^2 V = VV^2 = VU \,,$$

and note V is bijective if and only if U is bijective. If U is bijective, then

$$(Ux, x) = (V^2 x, x) = (Vx, Vx) = |Vx|^2 > 0$$

for each $x \in X$ with $x \neq \mathbf{0}$.

PROPOSITION 4.4. *Let Q be a form, and let $T : X \to X$ be the associated bijection. There is a Q-decomposition (X_+, X_-) such that $X_+ = T(X_+)$ is orthogonal to $X_- = T(X_-)$. Letting*

$$(x|y) = \begin{cases} Q(x,y) & \text{if } x, y \in X_+, \\ -Q(x,y) & \text{if } x, y \in X_-, \\ 0 & \text{if } x \in X_+ \text{ and } y \in X_-, \end{cases}$$

defines an inner product $(x|y)$ in X such that the norms $\|x\| = \sqrt{(x|x)}$ and $|x|$ are equivalent.

Proof. Let $U = (T^2)^{\frac{1}{2}}$. As both T and U are bounded, the sets

$$X_\pm = \{x \in X : Tx = \pm Ux\}$$

are closed T-invariant subspaces of X. Given $x \in X_+$, find a $z \in X$ with $x = Tz$, and observe

$$Tx = Ux = UTz = TUz.$$

Since T is bijective, $Uz = x = Tz$. Hence $z \in X_+$, which implies $T(X_+) = X_+$. Similarly we prove $T(X_-) = X_-$. Let $x \in X_+$ and $y \in X_-$. Then

$$Q(x,y) = (Tx, y) = (Ux, y) = (x, Uy) = (x, -Ty) = -Q(x,y),$$

and so $Q(x,y) = 0$. In particular, finding $z \in X_+$ with $x = Tz$, we obtain

$$(x, y) = (Tz, y) = Q(z, y) = 0.$$

Therefore $X_+ \subset X_-^\perp$. Now $(U + T)(X) \subset X_+$, since

$$(U - T)(U + T) = U^2 - T^2 = 0.$$

Thus if $x \in X_+^\perp$, then,

$$\big([U + T]x, y\big) = \big(x, [U + T]y\big) = 0$$

for each $y \in X$. Consequently $(U + T)x = 0$, which means $x \in X_-$. It follows $X_+^\perp = X_-$, or equivalently $X = X_+ \oplus X_-$. Since U is bijective, we have

$$Q(x,x) = (Ux, x) > 0 \quad \text{and} \quad Q(y,y) = -(Uy, y) < 0$$

for all $x \in X_+$ and $y \in X_-$ different from $\mathbf{0}$. This establishes (X_+, X_-) is a Q-decomposition. Denote by $P_\pm$ the orthogonal projections from X onto $X_\pm$, and let

$$(x|y) = Q(P_+x, P_+y) - Q(P_-x, P_-y)$$

for all $x, y \in X$. If $V = U^{\frac{1}{2}}$, then

$$\|x\|^2 = (UP_+x, P_+x) + (UP_-x, P_-x) = |VP_+x|^2 + |VP_-x|^2, \quad (*)$$

and we infer $(x|y)$ is an inner product in X. If $\lim \|x_i\| = 0$, then $\lim |x_i| = 0$, because $\lim P_+x_i = \lim P_-x_i = \mathbf{0}$ by $(*)$, and

$$|x_i| = |P_+x_i + P_-x_i| \leq |P_+x_i| + |P_-x_i|.$$

On the other hand, $\|x\| \leq \sqrt{2}\|V\| \cdot |x|$ according to $(*)$. $\qquad \square$

PROPOSITION 4.5. *If Q is an indefinite form, then $\big[C(x, Q)\big] = X$ for each $x \in X$.*

Proof. Since $C(x, Q) = C(\mathbf{0}, Q) - x$, it suffices to show $\big[C(\mathbf{0}, Q)\big] = X$. To this end choose Q-decomposition (X_+, X_-), and select a $y \in X_+$ with $y \neq \mathbf{0}$. If $z \in X_-$, let $u = z + ty$ and $v = z - ty$ where

$$t = \sqrt{-\frac{Q(z, z)}{Q(y, y)}}\,.$$

A direct calculation shows $u, v \in C(\mathbf{0}, Q)$, and so $z = \frac{1}{2}(u + v)$ belongs to $\big[C(\mathbf{0}, Q)\big]$. Thus $X_- \subset \big[C(\mathbf{0}, Q)\big]$, and similarly we prove $X_+ \subset \big[C(\mathbf{0}, Q)\big]$. The proposition follows. $\qquad \square$

In accordance with one's intuition, we actually proved X is the *convex hull* of $C(x, Q)$ for each $x \in X$, but we shall not need this stronger result.

PROPOSITION 4.6. *Let Q be an indefinite form, and let S be a symmetric bilinear functional on $X \times X$. If $C(\mathbf{0}, S) = C(\mathbf{0}, Q)$, then $S = cQ$ for a real number $c \neq 0$; in particular S is a form.*

Proof. Let (X_+, X_-) be a Q-decomposition, and select $x \in X_+$ and $y \in X_-$ so that $Q(x, x) = -Q(y, y)$. Then $u = x + y$ and $v = x - y$ belong to $C(\mathbf{0}, Q)$, and so

$$0 = S(u, u) = S(x, x) + 2S(x, y) + S(y, y),$$
$$0 = S(v, v) = S(x, x) - 2S(x, y) + S(y, y).$$

It follows $S(x, y) = 0$ and $S(x, x) = -S(y, y)$. Now select a point $z \in X_+$ so that $Q(z, z) = 1$, and let $c = S(z, z)$. Since Q is indefinite, there are nonzero vectors $x \in X_+$ and $y \in X_-$. The numbers $a = \sqrt{Q(x, x)}$ and $b = \sqrt{-Q(y, y)}$ are positive, and letting $x_1 = x/a$ and $y_1 = y/b$, we obtain

$$Q(x_1, x_1) = -Q(y_1, y_1) = 1.$$

According to our previous result, $S(x_1, y_1) = 0$ and

$$S(x_1, x_1) = -S(y_1, y_1) = S(z, z) = c.$$

It follows

$$S(x, y) = abS(x_1, y_1) = 0,$$
$$S(x, x) = a^2 S(x_1, x_1) = c\,Q(x, x),$$
$$S(y, y) = b^2 S(y_1, y_2) = c\,Q(y, y),$$

where the last two equalities hold also for $x = y = \mathbf{0}$. If $u \in X$, then $u = x + y$ where $x \in X_+$ and $y \in X_-$. Consequently

$$S(u, u) = S(x, x) + S(y, y) = c\big[Q(x, x) + Q(y, y)\big] = c\,Q(u, u),$$

and equality (11) implies $S = c\,Q$. As $C(\mathbf{0}, Q) \neq X$, we have $c \neq 0$. $\square$

5. The main result

Throughout the rest of my lecture, $\dim X \geq 3$ and Q is an arbitrary but fixed indefinite form whose signature is (σ_+, σ_-). For $x, y \in X$, we let

$$xy = Q(x, y) \quad \text{and} \quad x^2 = Q(x, x),$$

and write C_x instead of $C(x, Q)$. By $\mathcal{P}$ we denote the group of all bijective transformations $f : X \to X$ such that

$$f(C_x) = C_{f(x)}$$

for each $x \in X$. Thus a bijective map $f : X \to X$ belongs to $\mathcal{P}$ if and only if

$$[f(y) - f(x)]^2 = 0 \iff (y - x)^2 = 0$$

for all $x, y \in X$. It follows immediately $\mathcal{P}$ is a transformation group of X. The elements of $\mathcal{P}$ are called *Poincaré transformations*.

Instead of dealing with the whole group $\mathcal{P}$, it will be more convenient to study its homogeneous subgroup

$$\mathcal{P}_0 = \left\{ f \in \mathcal{P} : f(\mathbf{0}) = \mathbf{0} \right\}.$$

Throughout, $x \mapsto x'$ will stand for a map in $\mathcal{P}_0$, and we let $A' = \{x' : x \in A\}$ for every set $A \subset X$. The essential relationship between the groups $\mathcal{P}$ and $\mathcal{P}_0$ is provided the following simple lemma.

LEMMA 5.1. *Let $x \mapsto x^*$ be a map from $\mathcal{P}$, and let $z \in X$. If*

$$x' = (x + z)^* - z^*$$

then the map $x \mapsto x'$ belongs to $\mathcal{P}_0$.

Proof. Since the map $x \mapsto x^*$ is bijective, so is the map $x \mapsto x'$, and

$$0 = (y' - x')^2 = \left[(y + z)^* - (x + z)^* \right]^2$$

is equivalent to

$$0 = \left[(y - z) - (x - z) \right]^2 = (y - x)^2.$$

As $\mathbf{0}' = \mathbf{0}$, the lemma is proved. $\qquad\qquad\square$

MAIN THEOREM 5.2. *Each map $x \mapsto x'$ is linear.*

Before we outline the proof of Theorem 5.2, we derive some of its major consequences.

THEOREM 5.3. *Each map $x \mapsto x'$ is continuous, and there is a constant $c \neq 0$, depending on the map $x \mapsto x'$, such that $(x')^2 = cx^2$ for each $x \in X$. Moreover, $c > 0$ whenever $\sigma_+ \neq \sigma_-$.*

Proof. By Theorem 5.2, the map

$$S : (x, y) \mapsto x'y' : X \times X \to \mathbb{R}$$

is a bilinear symmetric functional on $X \times X$. Since $C(\mathbf{0}, S) = C_{\mathbf{0}}$, by Proposition 4.6, there is a $c \neq 0$ such that $x'y' = cxy$ for all $x, y \in X$. Select a Q-decomposition (X_+, X_-) as in Proposition 4.4, and let $(x|y)$ be the inner product associated with (X_+, X_-) according to Proposition 4.4. Since the norms $|x|$ and $\|x\|$ induced, respectively, by the inner products (x, y) and $(x|y)$ are equivalent, it suffices to establish the continuity of $x \mapsto x'$ with respect to the norm $\|x\|$. To this end, denote by $P_{\pm}$ the orthogonal projections, with respect to the inner product $(x|y)$, of X onto $X_{\pm}$.

Find linear spaces $Y_{\pm} \subset X$ so that $X_{\pm} = Y'_{\pm}$, and observe (Y_+, Y_-) or (Y_-, Y_+) is a Q-decomposition depending on whether $c > 0$ or $c < 0$, respectively. If x belongs to Y_+ or Y_-, then

$$
\begin{aligned}
\|x'\|^2 = \left|(x')^2\right| &= |c| \cdot |x^2| = |c| \cdot \left|(P_+x + P_-x)^2\right| \\
&= |c| \cdot \left|(P_+x)^2 + (P_-x)^2\right| = |c| \cdot \left|\|P_+x\|^2 - \|P_-x\|^2\right| \qquad (*) \\
&\leq |c|\left(\|P_+x\|^2 + \|P_-x\|^2\right) = |c| \cdot \|x\|^2 \, .
\end{aligned}
$$

According to condition (i) of Definition 4.2 there are linear maps $E_{\pm}$ from X onto $Y_{\pm}$ such that

$$x = E_+x + E_-x$$

for every $x \in X$. Proposition 4.3 and the closed graph theorem show the maps $E_{\pm}$ are bounded. Thus by $(*)$, for each $x \in X$,

$$
\begin{aligned}
\|x'\| = \left\|(E_+x)' + (E_-x)'\right\| &\leq \left\|(E_+x)'\right\| + \left\|(E_-x)'\right\| \\
&\leq \sqrt{|c|}\left(\|E_+x\| + \|E_-x\|\right) \leq \sqrt{|c|}\left(\|E_+\| + \|E_-\|\right)\|x\| \, ,
\end{aligned}
$$

and the map $x \mapsto x'$ is continuous. Finally, if $c < 0$ then

$$\sigma_+ = \dim X_+ = \dim Y_- = \dim X_- = \sigma_- \, .$$

$\square$

COROLLARY 5.4. *Each $f \in \mathcal{P}$ is a continuous affine map, and there is a constant $c(f) \neq 0$ such that*

$$\left[f(y) - f(x)\right]^2 = c(f)(y - x)^2$$

for all $x, y \in X$. Moreover, $c(f) > 0$ whenever $\sigma_+ \neq \sigma_-$.

In the remainder of this section, we outline the proof of Theorem 5.2. We begin with a definition.

DEFINITION 5.5. *A set $N \subset X$ is called a* null set *whenever $\mathbf{0} \in N$ and $(y - x)^2 = 0$ for all $x, y \in N$. If a null set is also a closed linear subspace of X, we call it a* null space.

Since a bijection of X belongs to $\mathcal{P}_0$ if and only if it maps null sets to null sets, understanding null sets is essential for the proof of Theorem 5.2.

LEMMA 5.6. *A set $N \subset X$ is a null set if and only if $\mathbf{0} \in N$ and $xy = 0$ for all $x, y \in N$. If N is a null set, then $[N]^-$ is a null space.*

Proof. The first part of the lemma follows from the equality

$$(y - x)^2 = y^2 - 2xy - x^2 = (y - \mathbf{0})^2 - 2xy + (x - \mathbf{0})^2 .$$

Now it is easy to verify $[N]$ is a null set whenever N is a null set, and the lemma follows from inequality (12). □

It follows from Zorn's lemma each null set is contained in a maximal (with respect to inclusion) null set, which is a null space by Lemma 5.6. Denoting by $\mathfrak{M}$ the family of all maximal null sets, it is easy to see

$$N \in \mathfrak{M} \iff N' \in \mathfrak{M} .$$

LEMMA 5.7. *Let $N \in \mathfrak{M}$, and let $L \subset N$ be a null space. If x is in $N - L$, then there is an $M \in \mathfrak{M}$ such that $L \subset M$ and $x \notin M$.*

Proof. Given a point z in $X - C_{\mathbf{0}}$, let

$$y^* = y - 2 \left(\frac{yz}{z^2}\right) z$$

for each $y \in X$. Intuitively, y^* is the reflection of y across the hyperplane

$$\{u : uz = 0\}\,.$$

A straightforward calculation reveals the map $y \mapsto y^*$ is a linear involution of X, and $(y^*)^2 = y^2$ for every $y \in X$. By inequality (12), the involution $y \mapsto y^*$ is also continuous, and hence

$$M = \{y^* : y \in N\}$$

belongs to $\mathfrak{M}$. Now assume $yz = 0$ for each $y \in L$, and $xz \neq 0$. Then $y^* = y$ for every $y \in L$, and

$$xx^* = x^2 - 2\left(\frac{xz}{z^2}\right)(xz) = -2\frac{(xz)^2}{z^2} \neq 0\,.$$

Thus $L \subset M$, and $x \notin M$ as Lemma 5.6 implies $x^* \notin N$.

We complete the proof by finding a point z in $X - C_0$ such that $xz \neq 0$, and $yz = 0$ for every $y \in L$. To this end, let T be the bijection associated with Q. According to Section 4, the map T is a linear homeomorphism, and hence $T(L)$ is a closed subspace of X. Denote by P the orthogonal projection onto $T(L)^\perp$, and let $u = PTx$. Then $u \neq \mathbf{0}$ and

$$yu = (Ty, u) = (Ty, PTx) = (PTy, Tx) = 0 \qquad (*)$$

for each $y \in L$. On the other hand,

$$
\begin{aligned}
xu &= (Tx, u) = (Tx, PTx) \\
&= (Tx, P^2Tx) = (PTx, PTx) \qquad (**) \\
&= |u|^2 > 0.
\end{aligned}
$$

Let

$$z = \begin{cases} u & \text{if } u^2 \neq 0, \\ u + x & \text{if } u^2 = 0 \end{cases}$$

and observe

$$z^2 = \begin{cases} u^2 & \text{if } u^2 \neq 0, \\ (u + x)^2 = 2xu & \text{if } u^2 = 0. \end{cases}$$

Inequality $(**)$ implies $z \notin C_0$ and $xz = xu > 0$. If $y \in L$, then $yu = 0$ by $(*)$, and $yx = 0$ by Lemma 5.6. $\qquad \square$

COROLLARY 5.8. *If $L \subset X$ is a null space, then*

$$L = \bigcap \{M \in \mathfrak{M} : L \subset M\}.$$

PROPOSITION 5.9. *If $N \subset X$ is a null set, then $\left([N]^-\right)' = [N']^-$.*

Proof. If $L = [N]^-$, then

$$L = \bigcap \{M \in \mathfrak{M} : L \subset M\}$$

by Corollary 5.8. Since $x \mapsto x'$ maps maximal null spaces to maximal null spaces,

$$L' = \bigcap \{M' : M \in \mathfrak{M} \text{ and } L \subset M\}$$

is a closed subspace of X. Now $N' \subset L'$ implies $[N']^- \subset L'$, and we obtain

$$[N']^- \subset \left([N]^-\right)'.$$

Applying this result to the null set N' and the map $x \mapsto x^*$ inverse to $x \mapsto x'$ provides the reverse inclusion. Indeed,

$$[N]^- = \left[(N')^*\right]^- \subset \left([N']^-\right)^*$$

and consequently,

$$\left([N]^-\right)' \subset [N']^-.$$

$\square$

COROLLARY 5.10. *Let $x \in X$ be such that $x^2 = 0$, and let $t \in \mathbb{R}$. Then*

$$(tx)' = t'x$$

for a $t' \in \mathbb{R}$.

Proof. Since the one-dimensional linear spaces $[x]$ and $[x']$ are closed, Proposition 5.9 implies $[x]' = [x']$. $\square$

PROPOSITION 5.11. *Let $x \in X$ and $t \in \mathbb{R}$. then $(tx)' = t'x'$ for some $t' \in \mathbb{R}$.*

This refinement of Corollary 5.10 is obtained by a series of rather technical lemmas, first for the case when

$$\min\{\sigma_+, \sigma_-\} \geq 2,$$

and then for the case when $\sigma_- = 1$; if $\sigma_+ = 1$, it suffices to replace Q by $-Q$. The argument relies on the assumption $\dim X \geq 3$. The interested reader is referred to [6, Section 4] for details.

PROPOSITION 5.12. *If $L \subset X$ is a line, then so is L'.*

Proof. Select distinct points x and y in L, and let K be the line passing through the points x' and y'. If $z \in L$ and $z \neq x$, find a $t \in \mathbb{R}$ with $y - z = t(x - z)$. According to Lemma 5.1, the map

$$x \mapsto x^\bullet = (x + z)' - z' : X \to X$$

belongs to $\mathcal{P}_0$. In view of Proposition 5.11, there is a $t^\bullet \in \mathbb{R}$ such that

$$y' - z' = (y - z)^\bullet = t^\bullet (x - z)^\bullet = t^\bullet (x' - z').$$

Thus $z' \in K$, and we have $L' \subset K$. Applying this result to the map $x \mapsto x^*$ inverse to the map $x \mapsto x'$, we obtain $K^* \subset L$, and consequently $K \subset L'$.   □

Since the field $\mathbb{R}$ has no nontrivial inner automorphisms [3, Theorem 1.19], the Main Theorem (Theorem 5.2) follows from the fundamental theorem of projective geometry [3, Theorem 2.26]: *If U is a linear space over $\mathbb{R}$ of dimension larger than one, then each bijection of U that maps lines onto lines is affine.*

6. An application

A map $x \mapsto x^* : X \to X$ is called an *isometry* whenever

$$|y^* - x^*| = |y - x|$$

for all $x, y \in X$. It is well known [1, Section 35] the family of all bijective isometries is a subgroup of the affine group of X.

DEFINITION 6.1. *A bijective map* $x \mapsto x^* : X \to X$ *is called a* pseudoisometry *whenever*

$$|y^* - x^*| = |v^* - u^*| \iff |y - x| = |v - u|$$

for all x, y, u, v *in* X.

Pseudoisometries may not preserve distances, but they preserve pairs of equidistant points. The family of all pseudoisometries is a transformation group of X which contains the group of all bijective isometries as a proper subgroup (e.g., the map $x \mapsto 2x$ is a pseudoisometry but not an isometry).

THEOREM 6.2. *If* $\dim X \geq 2$, *then each pseudoisometry is a continuous affine map. Moreover, given a pseudoisometry* $x \mapsto x^*$, *there is a constant* $a > 0$ *such that*

$$|y^* - x^*| = a|y - x|$$

for all $x, y \in X$.

Proof. In the Hilbert space $Y = X \oplus X$, the bijection

$$f : x_1 \oplus x_2 \mapsto x_1^* \oplus x_2^*$$

is a Poincaré transformation of Y with respect to the indefinite form

$$Q(x_1 \oplus x_2, y_1 \oplus y_2) = (x_1, y_1) - (x_2, y_2).$$

As before, for each $u \in Y$, let $u^2 = Q(u, u)$. Since $\dim Y \geq 4$, the results of Section 5 imply $f : Y \to Y$ is a continuous affine map, and there is a constant $c \neq 0$ such that

$$
\begin{aligned}
|y_1^* - x_1^*|^2 - |y_2^* - x_2^*|^2 &= \left[(y_1^* - x_1^*) \oplus (y_2^* - x_2^*) \right]^2 \\
&= (y_1^* \oplus y_2^* - x_1^* \oplus x_2^*)^2 \\
&= \left[f(y_1 \oplus y_2 - x_1 \oplus x_2) \right]^2 \\
&= c(y_1 \oplus y_2 - x_1 \oplus x_2)^2 \\
&= c\left[(y_1 - x_1) \oplus (y_2 - x_2) \right]^2 \\
&= c\left(|y_1 - x_1|^2 - |y_2 - x_2|^2 \right).
\end{aligned}
$$

Letting $x = x_1$, $y = y_1$, and $x_2 = y_2 = \mathbf{0}$, we obtain

$$|y^* - x^*|^2 = c|y - x|^2.$$

For $y \neq x$, the last inequality yields $c > 0$, and it suffices to let $a = \sqrt{c}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

COROLLARY 6.3. *If* $\dim X \geq 2$, *then each pseudoisometry is the composition of an isometry and a dilation.*

The assumption $\dim X \geq 2$ in Theorem 6.2 and Corollary 6.3 is essential. Indeed, viewing $\mathbb{R}$ as a linear space over the rationals, it is easy to construct a nonlinear bijection $f : \mathbb{R} \to \mathbb{R}$ such that

$$f(rx + sy) = rf(x) + sf(y)$$

for all $x, y \in \mathbb{R}$ and all rational numbers r and s. Such an f is a pseudoisometry of $\mathbb{R}$ that is not continuous, and hence not affine.

REMARK 6.4. *An interesting generalization of Corollary 6.3 was obtained in [5, Theorem 18]:* If $\dim X \geq 2$, than every bijection of X that maps circles onto circles is the composition of an isometry and a dilation.

REFERENCES

[1] N.I. AKHIEZER AND I.M. GLAZMAN, *Theory of Linear Operators in Hilbert Space*, vol. 1, Ungar, New York, 1966.

[2] A.D. ALEXANDROV, *A contribution to chronogeometry*, Cand. J. Math **19** (1967), 1119–1228.

[3] E. ARTIN, *Geometric Algebra*, Interscience, London, 1957.

[4] H.J. BORCHERS AND G.C. HEGERFELDT, *The structure of space-time transformation*, Comm. Math. Phys. **28** (1972), 259–266.

[5] R.J. GARDNER AND R.D. MAULDIN, *Bijections of* $\mathbb{R}^n$ *onto itself*, Geometrica Dedicata (1988), no. 26, 323–332.

[6] W.F. PFEFFER, *Lorentz tranformations of a Hilbert space*, Amer. J. Math **103** (1981), 691–709.

[7] F. RIESZ AND B. SZ.-NAGY, *Functional Analysis*, Dover, New York, 1990.

[8] W. RUDIN, *Real and Complex Analysis*, McGraw-Hill, New York, 1987.

[9] E.C. ZEEMAN, *Causality implies the Lorentz group,*, J. Math. Phys. **5** (1964), 490–493.