Mathematical Structures and Modeling
2013. N. 2(28). PP. 35–42

UDC 514.8

HOW TO DEFINE DISTANCE IN TERMS OF KINEMATIC METRIC AND/OR OTHER PHYSICAL QUANTITIES

F. Zapata

It is desirable to have a metric $d(a, b)$ which describes the closeness between space-time events. In Newtonian space, we can use Euclidean metric to describe the closeness of different spatial points. In relativistic physics, spatial distance is not invariant, it changes when we go from the original reference frame to another one which moves in relation to the original one. The only invariant characteristic is the proper time — known as kinematic metric. Can we use kinematic metric to define the desired distance $d(a, b)$ between space-time events? In this paper, we show that for the full space-time of special relativity, it is not possible to define a metric $d(a, b)$ in terms of the kinematic metric; however, if we limit ourselves to a compact (bounded) part of space-time, such a definition becomes possible. We also show that, more generally, for compact parts of space-time, we can define metric in terms of any number of physical fields.

1. Introduction

Distance is relative. Different measuring instruments ranging from rulers to laser-based super-accurate devices measure distance between two spatial points. By using these measuring instruments, for every two spatial points x and y, we can determine the distance $d(x, y)$ between these two points. This measured distance satisfies the usual properties $d(x, x) = 0$, $d(x, y) > 0$ when $x \neq y$, $d(x, y) = d(y, x)$, and the triangle inequality $d(x, z) \leq d(x, y) + d(y, z)$. In mathematics, a function $d(x, y)$ which satisfies these properties is called a *distance function*.

In Newtonian physics, distance between the two points had an absolute physical meaning. However, relativity theory showed that distance is not absolute: a measured distance between the two points depends, e.g., on whether we measure this distance by the original instrument or by another instrument which is moving with respect to the original one; see, e.g., [2, 4].

Research Institute for Manufacturing and Engineering Systems (RIMES), University of Texas at El Paso, El Paso, TX 79968, USA
E-mail: fazg74@gmail.com

Is there an absolute distance function? The usual physical distance is not absolute, it depends on the coordinate system. A natural question is: can we define an alternative continuous distance function which will be absolute? In other words, can we use coordinate-independent physical quantities to define a continuous coordinate-independent distance function?

What we do in this paper. We show that in some cases — e.g., for the full empty space-time — such a continuous coordinate-independent distance function is not possible. On the other hand, if we limit ourselves to a part of space-time which is compact, then such a continuous distance function is possible. We also show that, more generally, for compact parts of space-time, we can define metric in terms of any number of physical fields.

2. Case of Full Empty Space-Time

Empty space-time: what is observable. Let us first consider the case of an empty space-time. In Newtonian space, the distance $d(a, b)$ can be defined as a shortest path between the two spatial points a and b. In an empty space-time, we do not have a direct way of measuring distance, we can also measure time. For every two events a and b, if a causally precedes b (we will denote this by $a \le b$), we can measure proper time of different particles which start at a and end at b. In contrast to the Newtonian space, where the positive shortest distance corresponds to the straight line connecting a and b, in relativistic physics, we can always get from a to b in zero proper time: by sending a photon which is reflected so that it reaches a point b. In relativistic physics, the proper time is the *longest* along the straight line connecting a and b. If another particle travels with a certain speed in relation to the straight-line particle, then, according to the known feature of relativity theory, the time on that other particle slows down, so the overall travel time will be shorter than on the original straight-line particle.

We can therefore define a natural analogue of metric $\tau(a, b)$ as the longest proper time of a particle that travels from a to b. This function $\tau(a, b)$ is known as *kinematic* metric; see, e.g., [1, 3–5].

For the usual metric, the shortest path from x to z cannot be longer than the path from x to y, and then from y to z; thus, we have the usual triangle inequality $d(x, z) \le d(x, y) + d(y, z)$. For the kinematic metric, the longest time between a and c cannot be shorter than time of going from a to b and then from b to c. Thus, when $a \le b \le c$, we have $\tau(a, c) \ge \tau(a, b) + \tau(b, c)$; this inequality is known as the *anti-triangle inequality*.

For the case when a cannot causally influence b, the kinematic metric is usually defined as $\tau(a, b) = 0$.

The question. In this case, the question is: if we only know the kinematic metric $\tau(a, b)$, can we define a new function $d(a, b)$ in terms of the kinematic metric?

It is impossible to define a distance function in terms of kinematic metric: a proof. Let us show that in the empty Minkowski space-time of Special Relativity, where the kinematic metric takes the form

$$\tau((t,x),(s,y)) = \sqrt{(s-t)^2 - \rho^2(x,y)} \text{ if } s-t \geq d(x,y), \tag{1}$$

where $\rho(x,y) \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{3}(x_i - y_i)^2}$ is the usual metric in a 3-D Euclidean space, it is not possible to define a distance function in terms of $\tau(a,b)$.

Indeed, the kinematic metric $\tau(a,b)$ of special relativity is known to be invariant with respect to Lorentz transformations, transformations that describe a transition to a coordinate system associated with a moving body. Thus, if we could have a metric function $d(a,b)$ which is described in terms of the kinematic metric $\tau(a,b)$, then this distance function $d(a,b)$ would also be Lorentz-invariant. Let us prove that this is impossible, i.e., that there is no Lorentz-invariant distance function.

We will prove this by contradiction. Let us assume that such a distance function $d(a,b)$ exists. Let us pick some point $a=(t,x)$ in Minkowski space-time and let us pick a real number $\varepsilon > 0$. Since $d(a,a) = 0$ and the function $d(a,b)$ is continuous, there exists a neighborhood U of the point a for which $d(a,b) \leq \varepsilon$ for all $b \in U$. This neighborhood included points from the future light cone

$$L^+_a = \{(s,y) : s-t = \rho(x,y)\}.$$

Thus, for some points $b \in L^+_a$, we have $d(a,b) \leq \varepsilon$.

It is known that every two points from $L^+_a - \{a\}$ can be transformed into each other by an appropriate Lorentz transformation. Since $d(a,b) \leq \varepsilon$ for some

$$b \in L^+_a - \{a\}$$

and the distance function $d(a,b)$ is Lorentz-invariant, we can therefore conclude that $0 \leq d(a,b) \leq \varepsilon$ for all the points $b \in L^+_a$. This inequality $0 \leq d(a,b) \leq \varepsilon$ holds for all $\varepsilon > 0$, thus $d(a,b) = 0$ — which contradicts to the fact that for distance functions, we should have $d(a,b) > 0$ when $a \neq b$.

3. Case of a Compact Part of Space-Time

Formulation of the problem. Instead of the whole space-time, let us now consider a separable compact part of space-time, with a continuous kinematic metric $\tau(a,b)$.

Need to consider physically meaningful space-time models. Of course, we cannot define a distance function in terms of $\tau(a,b)$ if there exist two points a and b which cannot be distinguished by the kinematic metric, i.e., for which $\tau(a,c) = \tau(b,c)$ and $\tau(c,a) = \tau(c,b)$ for all c. In this case, such points a and b, while mathematically different, cannot be distinguished by any physical properties and are, therefore, identical from the physical viewpoint. It is therefore reasonable to consider space-time models which are, in this sense, *physically meaningful*, i.e., for which, for every $a \neq b$, there exists a point c for which either $\tau(a,c) \neq \tau(b,c)$ or $\tau(c,a) \neq \tau(c,b)$.

We also assume that the kinematic metric $\tau(a,b)$ is continuous with respect to some original (non-physical) metric $d_0(a,b)$.

Comment. For example, in a 4-D space time with points $a = (a_0, a_1, a_2, a_3)$, we can use a coordinate-dependent distance function $d_0(a,b) = \sqrt{\sum\limits_{i=0}^{3}(a_i - b_i)^2}$.

Result. It turns out that for physically meaningful compact space-time models, we can define a distance function in terms of kinematic metric $\tau(a,b)$: namely, we can take

$$d(a,b) \stackrel{\text{def}}{=} \sup_c \max(|\tau(a,c) - \tau(b,c)|, |\tau(c,a) - \tau(c,b)|). \tag{2}$$

Let us prove that this formula indeed defines a continuous metric, and that the topology generated by this metric $d(a,b)$ is equivalent to the topology generated by the original metric $d_0(a,b)$.

Proof. Let us first check that the function defined by the formula (2) is indeed a distance function.

Proving that $d(a,a) = 0$. For $a = b$, we have

$$|\tau(a,c) - \tau(b,c)| = |\tau(a,c) - \tau(a,c)| = 0$$

and

$$|\tau(c,a) - \tau(c,b)| = |\tau(c,a) - \tau(c,a)| = 0,$$

so we have

$$\max(|\tau(a,c) - \tau(b,c)|, |\tau(c,a) - \tau(c,b)|) = 0$$

and thus, $d(a,a) = 0$.

Proving that $d(a,b) = 0$ implies $a = b$. Since the kinematic metric is assumed to be physically meaningful, for each pair $a \neq b$, there exists a c for which either $|\tau(a,c) - \tau(b,c)| > 0$ or $|\tau(c,a) - \tau(c,b)| > 0$. In both cases, $\max(|\tau(a,c) - \tau(b,c)|, |\tau(c,a) - \tau(c,b)|) > 0$, and that, the supremum $d(a,b)$ of all such values is also positive. Also, from the definition (2), it follows that $d(a,b) = d(b,a)$ for all a and b.

Proving the triangle inequality. Let us prove that the expression (2) satisfies the triangle inequality, i.e., that $d(a,a') \le d(a,b) + d(b,a')$ for all a, b, and a'. The value $d(a,a')$ is the largest of all possible values $|\tau(a,c) - \tau(a',c)|$ and $|\tau(c,a) - \tau(c,a')|$. Thus, to prove that $d(a,a')$ does not exceed the sum $d(a,b) + d(b,a')$, it is sufficient to prove that each of these values $|\tau(a,c) - \tau(a',c)|$ and $|\tau(c,a) - \tau(c,a')|$ does not exceed the sum $d(a,b)+d(b,a')$. Indeed, for all real numbers A, B, and A', we have $|A - A'| \le |A - B| + |B - A'|$; this is a 1-D case of the triangle inequality. In particular, for $A = \tau(a,c)$, $B = \tau(b,c)$, and $A' = \tau(a',c)$, we get

$$|\tau(a,c) - \tau(a',c)| \le |\tau(a,c) - \tau(b,c)| + |\tau(b,c) - \tau(a',c)|.$$

Here, by definition of $d(a,b)$ and $d(a',b)$, we have $|\tau(a,c) - \tau(b,c)| \le d(a,b)$ and $|\tau(a',c) - \tau(b,c)| \le d(a',b)$, thus we conclude that

$$|\tau(a,c) - \tau(a',c)| \le d(a,b) + d(b,a').$$

Similarly, we can prove that

$$|\tau(c,a) - \tau(c,a')| \le d(a,b) + d(b,a')$$

for all c. This proves the triangular inequality. So, we have proved that the expression (2) defines a distance function.

Proving continuity. Let us show that the distance function $d(a,b)$ is continuous with respect to the original metric $d_0(a,b)$.

We know that the kinematic metric $\tau(a,b)$ is continuous. It is known that a continuous function on a compact set is always uniformly continuous with the respect to the original (non-physical coordinate-dependent) distance function d_0, i.e., for every $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$ such that if $d_0(a,a') \le \delta(\varepsilon)$ and $d_0(b,b') \le \delta(\varepsilon)$ then $|\tau(a,b) - \tau(a',b')| \le \varepsilon$. Let us show that if $d_0(a,a') \le \delta_0 \stackrel{\text{def}}{=} \delta\left(\dfrac{\varepsilon}{2}\right)$ and $d_0(b,b') \le \delta_0$ then $|d(a,b) - d(a',b')| \le \varepsilon$; this will prove that the function $d(a,b)$ is indeed continuous.

It is sufficient to prove that $d(a,b) \le d(a',b') + \varepsilon$; in this case, we can similarly prove that $d(a',b') \le d(a,b) + \varepsilon$ and thus, that $|d(a,b) - d(a',b')| \le \varepsilon$. By definition (2), $d(a,b)$ is the supremum of the values

$$\max(|\tau(a,c) - \tau(b,c)|, |\tau(c,a) - \tau(c,b)|)$$

corresponding to different points c. Thus, to prove that $d(a,b) \le d(a',b') + \varepsilon$, it is sufficient to prove that all these values are smaller than or equal to $d(a',b') + \varepsilon$, i.e., that

$$\max(|\tau(a,c) - \tau(b,c)|, |\tau(c,a) - \tau(c,b)|) \le d(a',b') + \varepsilon$$

for all c.

The largest of the two numbers $|\tau(a,c) - \tau(b,c)|$ and $|\tau(c,a) - \tau(c,b)|$ is smaller than or equal to a certain bound if and only if both numbers do not exceed this bound. Thus, it is sufficient to prove that for every c, we have $|\tau(a,c) - \tau(b,c)| \le$

$d(a', b') + \varepsilon$ and $|\tau(c, a) - \tau(c, b)| \le d(a', b') + \varepsilon$. Without losing generality, we will prove the first inequality; the second one is proven similarly.

From $d_0(a, a') \le \delta_0 = \delta\left(\dfrac{\varepsilon}{2}\right)$ and $d(c, c') = 0$ for $c' = c$, we conclude that $|\tau(a, c) - \tau(a', c)| \le \dfrac{\varepsilon}{2}$. Similarly, from $d_0(b, b') \le \delta_0$, we conclude that

$$|\tau(b, c) - \tau(b', c)| \le \frac{\varepsilon}{2}.$$

Thus, we have

$$|(\tau(a, c) - \tau(b, c)) - (\tau(a', c - \tau(b', c))| = |(\tau(a, c) - \tau(a', c)) - (\tau(b, c) - \tau(b', c))| \le$$

$$|\tau(a, c) - \tau(a', c)| + |\tau(b, c) - \tau(b', c))| \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence,

$$|\tau(a, c) - \tau(b, c)| \le |\tau(a', c) - \tau(b', c)| + \varepsilon.$$

By definition,

$$|\tau(a', c) - \tau(b', c)| \le \sup_c \max(|\tau(a', c) - \tau(b', c)|, |\tau(c, a') - \tau(c, b')|) = d(a', b').$$

Thus, we indeed conclude that $|\tau(a', c) - \tau(b', c)| \le d(a', b') + \varepsilon$. The inequality is proven and so, the metric $d(a, b)$ is indeed continuous with respect to the original metric $d_0(a, b)$.

Proving the equivalence of topologies generated by metrics $d_0(a, b)$ and $d(a, b)$. On a compact set, continuity of the metric $d(a, b)$ implies its uniform continuity. Thus, for every $\varepsilon > 0$, there exists a $\delta > 0$ for which $d_0(a, b) \le \delta$ implies $d(a, b) \le \varepsilon$. This implies that when $d_0(a_n, b_n) \to 0$, we also have $d(a_n, b_n) \to 0$.

To prove that the metrics $d_0(a, b)$ and $d(a, b)$ generate the same topology, it is sufficient to prove that, vice versa, for every $\varepsilon > 0$, there exists a $\delta > 0$ for which $d(a, b) \le \delta$ implies that $d_0(a, b) \le \varepsilon$. We will prove this by contradiction. Suppose the above statement is not true. This means that there exists an $\varepsilon > 0$ for which, for every $\delta > 0$, there exist points a and b for which $d(a, b) \le \delta$ and $d_0(a, b) > \varepsilon$. In particular, for each natural number n, we can take $\delta = 2^{-n}$ and conclude that there exist points a_n and b_n for which $d(a_n, b_n) \le 2^{-n}$ and $d_0(a_n, b_n) > \varepsilon$.

On a separable compact set, every sequence has a convergent subsequence. Without losing generality, let us assume that the sequences a_n and b_n themselves converge (in the sense of the original metric $d_0(a, b)$), i.e., that $a_n \to A$ and $b_n \to B$ for some limit points A and B. Since both the original metric $d_0(a, b)$ and the new metric $d(a, b)$ are continuous in the original topology, we conclude that $d(A, B) = \lim_n 2^{-n} = 0$ and that $d_0(A, B) \ge \varepsilon > 0$. Since $d_0(A, B) > 0$, this means that $A \ne B$. Since we assumed that the space-time is physically meaningful, this means that there exists a point C for which either $\tau(A, C) \ne \tau(B, C)$ or $\tau(C, A) \ne \tau(C, B)$. In both cases, we have $\max |\tau(A, C) - \tau(B, C)|, |\tau(C, A) - \tau(C, B)| > 0$ and thus,

$$d(A, B) = \sup_c \max |\tau(A, c) - \tau(B, c)|, |\tau(c, A) - \tau(c, B)| \ge$$

$$\max |\tau(A,C) - \tau(B,C)|, |\tau(C,A) - \tau(C,B)| > 0,$$

which contradicts to $d(A,B) = 0$.

This contradiction proves that the above statement is true, i.e., $d_0(a,b)$ is continuous with respect to $d(a,b)$ and thus, the metrics $d_0(a,b)$ and $d(a,b)$ indeed generate the same topology.

4. General Case of Physical Fields Over a Compact Part of Space-Time

Description of the general case. Let us assume that we have a separable compact space-time A. On this space-time, we can have physical fields $f_1(a), \ldots, f_n(a)$, i.e., functions which assign a real number to each space-time point a (a vector-valued or a tensor-valued field can be viewed as a tuple consisting of the component fields). We may also have functions $g_1(a_1, \ldots, a_{n_1})$, $\ldots$, $g_m(a_1, \ldots, a_{n_m})$ that assign numbers to *tuples* of space-time points — just like kinematic metric assigns a number $\tau(a,b)$ to each pair of space-time points a and b.

Similarly to the case of kinematic metric, we assume that the fields are physically meaningful, in the sense that for every two space-time points a and b, we have either $f_i(a) \neq f_i(b)$ for some i, or we have $g_j(a_1, \ldots, a_{k-1}, a, a_{k+1}, \ldots, a_{n_j}) - g_j(a_1, \ldots, a_{k-1}, b, a_{k+1}, \ldots, a_{n_j})$ for some j and k and for some points $a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_{n_j}$. We also assume that all these functions $f_i(a)$ and $g_j(a_1, \ldots, a_{n_j})$ are continuous with respect to some original metric $d_0(a,b)$.

How to define metric in this case. Similarly to the case of kinematic metric, we can define the metric $d(a,b)$ as follows:

$$d(a,b) = \max(d_{f_1}(a,b), \ldots, d_{f_n}(a,b), d_{g_1}(a,b), \ldots, d_{g_m}(a,b)), \tag{3}$$

where

$$d_{f_i}(a,b) \stackrel{\text{def}}{=} |f_i(a) - f_i(b)| \text{ and}$$

$$d_{g_j}(a,b) = \sup_{k, a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_{n_j}} \Delta_{jk}(a, b, a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_{n_j}),$$

where

$$\Delta_{jk}(a, b, a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_{n_j}) \stackrel{\text{def}}{=}$$

$$|g_j(a_1, \ldots, a_{k-1}, a, a_{k+1}, \ldots, a_{n_j}) - g_j(a_1, \ldots, a_{k-1}, b, a_{k+1}, \ldots, a_{n_j})|.$$

Similarly to the previous section, we can prove that this formula defines a continuous metric, and that the topology generated by this metric $d(a,b)$ is equivalent to the topology generated by the original metric $d_0(a,b)$.

References

1. Busemann H. Timelike Spaces, Warszawa:PWN, 1967.
2. Feynman R., Leighton R., and Sands M. The Feynman lectures on physics, Boston, Massachusetts : Addison Wesley, 2005.
3. Kronheimer E.H., and Penrose R. On the structure of causal spaces // Proc. Cambr. Phil. Soc. 1967, V. 63, No. 2, P. 481-501.
4. Misner C.W., Thorne K.S., and Wheeler J.A. Gravitation, New York : W.H. Freeman, 1973.
5. Pimenov R.I. Kinematic Spaces: Mathematical Theory of Space-Time, New York : Consultants Bureau, 1970.