2-1-2022

# A Natural Causality-Motivated Description of Learning

Olga Kosheleva
*The University of Texas at El Paso*, olgak@utep.edu

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



Comments:

Technical Report: UTEP-CS-22-24

*Mathematical*
*Structures and Modeling*
*2022. N. X(XX). PP. 1–??*

UDC XXX
DOI: 10.25513/2222-8772.2022.X.1-??

# A NATURAL CAUSALITY-MOTIVATED DESCRIPTION OF LEARNING

**Olga Kosheleva**
Ph.D. (Phys.-Math.), Associate Professor, e-mail: olgak@utep.edu
**Vladik Kreinovich**
Ph.D. (Phys.-Math.), Professor, e-mail: vladik@utep.edu

University of Texas at El Paso, El Paso, Texas 79968, USA

**Abstract.** Teaching is not easy. One of the main reasons why it is not easy is that the existing descriptions of the teaching process are not very precise – and thus, we cannot use the usual optimization techniques, techniques which require a precise model of the corresponding phenomenon. It is therefore desirable to come up with a precise description of the learning process. To come up with such a description, we notice that on the set of all possible states of learning, there is a natural order $s \leq s'$ meaning that we can bring the student from the state $s$ to the state $s'$. This relation is similar to the causality relation of relativity theory, where $a \leq b$ means that we can move from point $a$ to point $b$. In this paper, we use this analogy with relativity theory to come up with the basics of such an order-based description of learning. We hope that future studies of these basics will help to improve the teaching process.

**Keywords:** teaching, learning, relativity theory, causality, metric, kinematic metric.

## 1. Formulation of the Problem

**How can we train instructors faster?** Effective teaching is not easy. For most instructors, it takes several years to master teaching – and even after these years, no matter how experienced the instructor is, there is always room for significant improvement.

How can we speed up this process? How can we make sure that instructors learn the teaching skills as soon as possible – and not, as now, spend several years learning these skills?

**This problem is not easy.** One of the reasons why training instructors takes so long is that teaching is not a precise science.

Naive thinking is that since we want to achieve optimal teaching, why not use optimization techniques – that have been so successful in many other application areas? Unfortunately, this is not that easy: optimization techniques require that the problem is formulated in precise terms, and the teaching problem is far from such a description.

**We need to describe teaching in precise terms.** To be able to utilize the effectiveness of the existing optimization tools, it is therefore desirable to be able to come up with a formal description of the teaching process.

**Our approach to this description.** One of our areas of interest is foundations of relativity, where an ordering relation – namely, the causality relation – turned out to play a fundamental role; see, e.g., [1–3]. In view of this fact, a natural idea is to look for a description of teaching in terms of an ordering relation.

## 2. Towards a Natural Description of the Learning Process

**What are states of student knowledge.** Before we start analyzing what is the natural relation on the set $S$ of all possible states of knowledge, we first need to find out what is a natural way to describe these states.

To fully characterize the student's knowledge of the class material, we need to describe this student's degree of knowledge in each topic. Usually, the knowledge of each topic is described by a grade, and grades are somewhat subjective. To avoid this subjectivity, we can use some objective (or at least inter-subjective) way to gauging this knowledge: e.g., by the number of hours that an average student would take to get to this level of knowledge.

Thus, at any given moment of time, the student's knowledge can be characterized by the values $s_1, \ldots, s_n$ describing this student's knowledge of all $n$ topics, or, in other words, by an $n$-dimensional point $s = (s_1, \ldots, s_n)$. In this case, $S$ is simply a subset in the $n$-dimensional affine space.

To be more precise, since the number of hours is always non-negative, $S$ is a quadrant of the $n$-dimensional affine space in which all the coordinates $s_i$ are non-negative.

**What is a natural ordering relation between states of knowledge.** Now that we have an idea of what is the set $S$ of states of knowledge, we can start analyzing what is a natural relation between these states. To come up with such a description, let us use an analogy with causality. Causality relation $a \leq b$ between two points (events) in space-time means that, in principle, we can go from the point $a$ to the point $b$, i.e., an observer can first observe $a$ and then observe $b$.

Similarly, we can define an order $s \leq s'$ between two different states of knowledge as the possibility to go from the state $s$ to the state $s'$, i.e., the possibility that a student was first in the state $s$ and at some future moment of time, the knowledge of this student is characterized by the state $s'$.

We assume that the skills the students learn are not forgotten during for the duration of the course – or, to be more precise, that, in the first approximation, we can ignore the effects of possible forgetting. Under this assumption, the student's level of knowledge in each topic cannot decrease, so we cannot have $s_i > s'_i$. On the other hand, if $s_i \leq s'_i$ for all $i$, then we can move from the state $s$ to the state $s'$ by teaching the student additional material in each topic. Thus, a natura ordering relation on the set $S$ of all states of the student knowledge is the coordinate-wise order:

$$(s = (s_1, \ldots, s_n) \le s' = (s'_1, \ldots, s'_n)) \Leftrightarrow (s_i \le s'_i \text{ for all } i). \quad (1)$$

**How much effort do we need to move the students from one state to another.** The ultimate goal of teaching is to bring the student from the original state $(0, \ldots, 0)$ (in which the student does not yet have any knowledge of any of the class topics) to the desired state $\ell = (\ell_1, \ldots, \ell_n)$, where $\ell_i$ is the student's desired level of knowledge on the $i$-th topic.

Our goal is to bring the student there the fastest way. To find out which way is the fastest, we need to know, for every two states $s \le s'$, how much student effort (e.g., measured by hours) we need to get from the state $s$ to the state $s'$. Let us denote this amount by $d(s, s')$. To be more precise, $d(s, s')$ denote the smallest possible effort needed to get from the state $s$ to the state $s'$.

*Comment.* In the idealized case when all topics are independent, the only way to go from state $s$ to state $s'$ is to teach the student additional material for each topic. Because of our selection of the way we measure the student's knowledge in each topic – by the number of hours needed to go from 0 to $s_i$ – the additional time needed for the student to improve his/her knowledge from level $s_i$ to the level $s'_i$ is to spend time $s'_i - s_i$. In this case, the overall time needed to go from $s$ to $s'$ is equal to the sum of these times:

$$d(s, s') = \sum_{i=1}^{n} (s'_i - s_i). \quad (2)$$

In reality, topics are interdependent, so the knowledge of one topic helps to study another topic. For example, knowing basic physics helps students to better understand calculus – for example, the derivative can be naturally understood when a student realizes the velocity is the derivative of the coordinate. Similarly, knowing the derivative can help the student better understand the ideas of velocity and acceleration. Because of such mutual help, the overall time needed to go from $s$ to $s'$ can be smaller that the sum of the corresponding times:

$$d(s, s') \le \sum_{i=1}^{n} (s'_i - s_i). \quad (3)$$

**What are natural properties of the function $d(s, s')$.** Any transition requires some efforts: the only time when the effort is 0 is when $s = s'$. So, we have the following property:

$$d(s, s) = 0 \text{ and } (d(s, s') = 0 \Rightarrow s = s'). \quad (4)$$

If we can go from state $s$ to $s'$ by using the amount $d(s, s')$ of resources, and then we can go from $s'$ to $s''$ by using the amount $d(s', s'')$ of resources, then one of the possible ways to go from the state $s$ to the state $s''$ is to go through $s'$. For this two-stage transition, we spend the amount $d(s, s') + d(s', s'')$. Thus, the smallest possible amount $d(s,s'')$ of resources needed to go from $s$ to $s''$ cannot exceed this sum. Thus, we have the usual triangle inequality:

$$d(s,s'') \le d(s,s') + d(s',s''). \tag{5}$$

**How is our function $d(s,s')$ related to metric and to its space-time analogue – kinematic metric (as measured by proper time)?** Properties (4) and (5) resemble the usual properties of metric. The main difference is that in our case, the value $d(s,s')$ is only defined when $s \le s'$. From this viewpoint, this notion resembles *kinematic metric* $\tau(a,b)$ – the proper time that an inertial particle would measure when it goes from event $a$ to event $b$: this values is also only defined when $a \le b$. However, kinematic metric is known to be the *largest* value of the proper time – not the smallest as in our case – and thus, instead of the triangle inequality, it satisfies the opposite ("anti-triangle") inequality

$$\tau(a,a'') \ge \tau(a,a') + \tau(a',a'').$$

From this viewpoint, the proposed model is intermediate between the regular metrics and the kinematic metrics.

**Towards a formal general definition.** Let us combine the above-described natural properties of this "learning" metric (we will call it *$\ell$-metric*, $\ell$ for "learning") into the following definition:

**Definition 1.** *Let $(S,\le)$ be a partially ordered set. By an $\ell$-*metric *we mean a function $d(s,s')$ that is defined for all pairs $s,s' \in S$ for which $s \le s'$ and that satisfies the properties* (4) *and* (5).

**Challenge.** Since $\ell$-metrics provide a natural description of learning, we believe that to enhance the learning process, it will be beneficial to study the properties of such $\ell$-metrics.

One such property is described in the next section.

## 3. Shift-Invariant Scale-Invariant $\ell$-Metrics on the Finite-Dimensional Affine Space with Component-Wise Order

**Why shift-invariant and scale-invariant.** In the study of casuality, a good approximation to real-life causality is provided by the Special Relativity Theory, in which:

- the causality relation is invariant with respect to shift $a \le b \Leftrightarrow a+c \le b+c$ and with respect to scalings: $a \le b \Leftrightarrow \lambda \cdot a \le \lambda \cdot b$ for all $\lambda > 0$, and
- the kinematic metric is both shift- and scale-invariant: $\tau(a,b) = \tau(a+c,b+c)$ and $\tau(\lambda \cdot a, \lambda \cdot b) = \lambda \cdot \tau(a,b)$ for all $\lambda > 0$.

In our case, component-wise order is also clearly shift- and scale-invariant. It is therefor reasonable to consider shift- and scale-invariant $\ell$-metrics.

**Definition 2.** *Let $(S, \le)$ be an $n$-dimensional affine space with coordinate-wise order* (1)*. We say that an $\ell$-metric is:*

- shift-invariant *if $d(s, s') = d(s + s'', s' + s'')$ for all $s$, $s'$, and $s''$, and*
- scale-invariant *if $d(\lambda \cdot s, \lambda \cdot s') = \lambda \cdot d(s, s')$ for all $s$, $s'$, and $\lambda > 0$.*

**Proposition 1.** *Let $(S, \le)$ be an $n$-dimensional affine space with coordinate-wise order* (1)*. The following two conditions are equivalent to each other for any function $d(s, s')$ defined for all pairs $s, s' \in S$ for which $s \le s'$:*

- *$d(s, s')$ is a shift- and scale-invariant $\ell$-metric;*
- *$d(s, s')$ has the form*

$$d(s, s') = (s'_1 - s_1) \cdot F\left(\frac{s'_2 - s_2}{s'_1 - s_1}, \ldots, \frac{s'_2 - s_2}{s'_1 - s_1}\right)$$

*for some positive-valued convex function $F(r_2, \ldots, r_n)$.*

**Proof.** Shift-invariance clearly implies that $d(s, s') = d(0, s - s')$, and scale-invariance, with $\lambda = s'_1 - s_1$ (when $s'_1 > s_1$) implies that

$$d(s, s') = d(0, s' - s) = (s'_1 - s_1) \cdot d(0, (1, r_2, \ldots, r_n)),$$

where we denoted

$$r_i \stackrel{\text{def}}{=} \frac{s'_i - s_i}{s'_1 - s_1}.$$

So, if we denote

$$F(r_2, \ldots, r_n) \stackrel{\text{def}}{=} d(0, (1, r_2, \ldots, r_n)),$$

we almost get the desired result – the only thing remaining to prove is that the triangle inequality for the original function $d(s, s')$ is equivalent to convexity of the function $F(r_2, \ldots, r_n)$.

Indeed, due to shift-invariance, the triangle inequality gets the following equivalent form

$$d(0, s'' - s) \le d(0, s' - s) + d(0, s'' - s').$$

So, if we denote $a \stackrel{\text{def}}{=} s' - s$ and $b \stackrel{\text{def}}{=} s'' - s'$, we get the following equivalent form:

$$d(0, a + b) \le d(0, a) + d(0, b). \tag{6}$$

Due to scale-invariance, we have

$$d(0, a) = a_1 \cdot d(0, (1, A_2, \ldots, A_n)) = a_1 \cdot F(A_2, \ldots, A_n),$$

where we denoted
$$A_i \stackrel{\rm def}{=} \frac{a_i}{a_1}.$$
Similarly, we have
$$d(0,b)=b_1\cdot F(B_1,\ldots,B_n),$$
where we denoted
$$B_i \stackrel{\rm def}{=} \frac{b_i}{b_1},$$
and
$$d(0,a+b)=(a_1+b_1)\cdot F(C_1,\ldots,C_n),$$
where we denoted
$$C_i \stackrel{\rm def}{=} \frac{a_i+b_i}{a_1+b_1}.$$
Thus, the triangle inequality (6) takes the form
$$(a_1+b_1)\cdot F(C_2,\ldots,C_n)\le a_1\cdot F(A_2,\ldots,A_n)+b_1\cdot F(B_2,\ldots,B_n).$$
Dividing both sides of this inequality by $a_1+b_1$, we get an equivalent inequality
$$F(C_2,\ldots,C_n)\le \frac{a_1}{a_1+b_1}\cdot F(A_2,\ldots,A_n)+\frac{b_1}{a_1+b_1}\cdot F(B_2,\ldots,B_n),$$
i.e., if we denote
$$\alpha \stackrel{\rm def}{=} \frac{a_1}{a_1+b_1},$$
the form
$$F(C_2,\ldots,C_n)\le \alpha\cdot F(A_2,\ldots,A_n)+(1-\alpha)\cdot F(B_2,\ldots,B_n). \qquad (7)$$
By definition of $A_i$ and $B_i$, we have $a_i=a_1\cdot A_i$ and $b_i=b_1\cdot B_i$. Substituting these expressions for $a_i$ and $b_i$ into the formula that defined $C_i$, we get
$$C_i=\frac{a_i+b_i}{a_1+b_1}=\frac{a_1\cdot A_i+b_1\cdot B_i}{a_1+b_1}=\frac{a_1}{a_1+b_1}\cdot A_i+\frac{b_1}{a_1+b_1}\cdot B_i=\alpha\cdot A_i+(1-\alpha)\cdot B_i.$$
Thus, the inequality (7) takes the equivalent form
$$F(\alpha\cdot A_2+(1-\alpha)\cdot B_2,\ldots,\alpha\cdot A_n+(1-\alpha)\cdot B_n)\le$$
$$\alpha\cdot F(A_2,\ldots,A_n)+(1-\alpha)\cdot F(B_2,\ldots,B_n),$$
which is exactly the definition of convexity.

The equivalence between triangle inequality for the function $d(s,s')$ and the convexity of the corresponding function $F(r_2,\ldots,r_n)$ is thus proven, and so is the proposition.

*Comment.* The case when all topics are independent and the $\ell$-metric is described by the formula (2) corresponds to the convex function $F(r_2,\ldots,r_n)=r_2+\ldots+r_n$.

## Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

## References

1. A. D. Alexandrov, "On Lorentz transformations", *Uspekhi Math. Nauk*, 1950, Vol. 5, No. 1, p. 187 (in Russian).
2. A. D. Alexandrov and V. V. Ovchinnikova, "Remarks on the foundations of Special Relativity", *Leningrad University Vestnik*, 1953, No. 11, pp. 94–110 (in Russian).
3. E. C. Zeeman, "Causality implies the Lorentz group", *Journal of Mathematical Physics*, 1964, Vol. 5, No. 4, pp. 490–493.