From Aristotle to Newton, from Sets to Fuzzy Sets, and from Sigmoid to ReLU: What Do All These Transitions Have in Common?

Christian Servin[1], Olga Kosheleva[2], and Vladik Kreinovich[2]

[1]Information Technology Systems Department
El Paso Community College (EPCC)
919 Hunter Dr., El Paso, TX 79915-1908, USA
cservin1@epcc.edu

[2]Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

1. Aristotelian physics: a brief reminder

- In ancient times, most researchers believed that a natural state of an un-disturbed object is when this object is not moving at all.
- In order to make the object move, we need to apply some force.
- If we stop applying this force, the object stops moving.
- Now we know that this description is wrong.
- However, from the commonsense viewpoint, this description makes perfect sense.
- For example, if we push a cart, it will start rolling, but once we stop pushing, it will pretty fast come to a stop.
- Because of this description, people believed that there is a force constantly pushing the planets and the Sun across the sky.
- For example, many religious people believed that angels are constantly pushing the Sun and the planets with their wings.

2. How can we describe this worldview in mathematical terms

- In mathematical terms, this description means that in order to change the location of an object, we need to apply some force.
- In other words, any change in the object's coordinates x has to be explained by an application of some force.
- From the mathematical viewpoint:
  - a change from the object's coordinate $x(t)$ at some moment of time t and its coordinate $x(t+\Delta t)$ at the "next" moment of time
  - can be described by the derivative:

$$\dot{x}(t) \approx \frac{x(t+\Delta t) - x(t)}{\Delta t}.$$

- In these terms:
  - if no force is applied, the object does not move, so the derivative $\dot{x}(t)$ is equal to 0;
  - thus, the only way the derivative becomes different from 0 is when some external force is applied.

3. How can we describe this worldview in mathematical terms (cont-d)

- In other words, the derivative $\dot{x}$ is determined by the external forces.
- So we have an equation $\dot{x}(t) = F$, where F is the corresponding force.
- In mathematics, such equations – involving the first derivative of the function – are called *differential equations of first order*.

4. Galileo's discovery and the resulting Newtonian physics

- The Aristotelian approach explains the behavior of physical objects on the *qualitative* level.
- However:
  - when researchers started performing experiments,
  - they realized that this approach does not lead to a good *quantitative* explanation.
- One of the first researchers who experimentally tested effects of different forces on motion was Galileo.
- Based on his experiments, Galileo came up with a completely different approach to motion.

5. Galileo's discovery and the resulting Newtonian physics (cont-d)

- Namely, Galileo has discovered what he called *the Law of Inertia*:
  - if an object starts moving with a constant speed in some direction,
  - then, in the absence of external forces, it will continue this movement indefinitely, and
  - if the object stops moving, this means that some forces were applied.
- For example, when we stop pushing a cart, it does stop – because the force of friction is applied.
- And if we place a cart of a very smooth surface, where friction is small – e.g., on a smooth ice – it will continue rolling for a very long time.
- Similarly, if we throw a ball into the air, it will stop – because its motion is affected by the force of gravity.

6. Galileo's discovery and the resulting Newtonian physics (cont-d)

- Galileo provided quantitative description of some of the phenomena.
- However, the full description of motion came from Newton.
- Newton used the Law of Inertia as one of the main three laws of motion.
- However, he supplemented it with the other two laws.
- He was able to do it, since he invented calculus, the mathematical technique appropriate for describing motion.

7. How can we describe Newton's worldview in mathematical terms

- In mathematical terms, the above description means that in order to change the velocity of an object, we need to apply some force.
- In other words, any change in the object's velocity v has to be explained by an application of some force.
- From the mathematical viewpoint:
  - a change from the object's velocity $v(t)$ at some moment of time t and its velocity $v(t+\Delta t)$ at the "next" moment of time
  - can be described by the derivative:

$$\dot{v}(t) \approx \frac{v(t+\Delta t) - v(t)}{\Delta t}.$$

8. How can we describe Newton's worldview in mathematical terms (cont-d)

- In these terms:
  - if no force is applied, the object does not change its velocity, so the derivative $\dot{v}(t)$ is equal to 0;
  - thus, the only way the derivative becomes different from 0 is when some external force is applied.
- In other words, the derivative $\dot{v}$ is determined by the external forces.
- So we have an equation $\dot{v}(t) = F$, where F is the corresponding force.
- The velocity v itself is the derivative of the coordinates $v = \dot{x}$.
- So, its derivative $\dot{v}$ is the second derivative of the coordinate $\dot{v} = \ddot{x}$.
- Thus, the dynamic equation takes the form $\ddot{x} = F$.
- In mathematics, such equations – involving the second derivative of the function – are called *differential equations of second order*.

9. Why second order: a kind-of explanation

- So far:
  - the only argument that we provided in favor of second-other equations was
  - that they lead to a more accurate description of the physical world.
- But *why* do they provide a more accurate description of the physical world?
- A possible – although probably not very convincing – answer to this question comes from the fact that:
  - physical equations are usually described
  - in terms of an optimization principle.
- For example, light passing through several media follows the path for which the transition time is the smallest.
- Equations of motion can be equivalently described as saying that some functional $\int L(x, \dot{x})\, dt$ attains its smallest possible value.

10. Why second order: a kind-of explanation (cont-d)

- This functional is called *action*.
- It is known that the minimum of a function is attained when its derivative is equal to 0.
- Similarly, it can be shown that the minimum of a functional is attained when its so-called *functional derivative* is equal to 0.
- This derivative takes the form

$$\frac{\delta L}{\delta x} = \frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = \frac{\partial L}{\partial x} - \frac{\partial^2 L}{\partial x \partial \dot{x}} \cdot \dot{x} - \frac{\partial^2 L}{\partial^2 \dot{x}} \cdot \ddot{x}.$$

- Equating this to 0 is always a second-order equation.
- In principle, we can consider functionals in which L depends only on x.
- But in this case, we simply have a static equation, no dynamics.

11. Why second order: a kind-of explanation (cont-d)

- We can also consider the case when L also depends on the second derivative.
- In this case, we get a 4-th order differential equation.
- Optimization never leads to the first- (or any odd-) order differential equations.
- The simplest even-order differential equation is the second-order one.

12. Case of ecology

- Interestingly, a similar transition, from first- to second-order differential equations, leads to more accurate models in ecology.
- In ecology, the above optimization-based explanation makes more sense than in physics.
- In physics, the fact that equations come from optimization principles is largely an empirical fact with no clear physical meaning.
- In biology, the need to have an optimal fit with the environment is one of the main principles.
- Without such fit, the species would not survive.

13. Other cases when optimality naturally appears

- There are other cases when optimality naturally appears: namely, the cases when we design objects or when we design algorithms.
- In all such cases, we want to select the design which is the best for our purpose.
- Let us consider two such examples.

14. Two examples: a general description

- In the first example, we are interested in how our degree of confidence $x(t)$ that a value t satisfies some property changes with t.
- This property may be "t is small", it may be "t is larger than 0", etc.
- In the second example, we are interested in the activation function $x(t)$ of a neuron, i.e., in a function that transforms:
  - a linear combination t of neuron's input signals
  - into the output signal.
- From the application viewpoint, these two examples deal with completely different situations.
- However, from the mathematical viewpoint, they are similar.
- In both cases, we are interested in a real-valued function $x(t)$ of a real-valued variable t.
- Let us analyze, for these two examples, what would descriptions in terms of first- and second-order differential equations lead to.

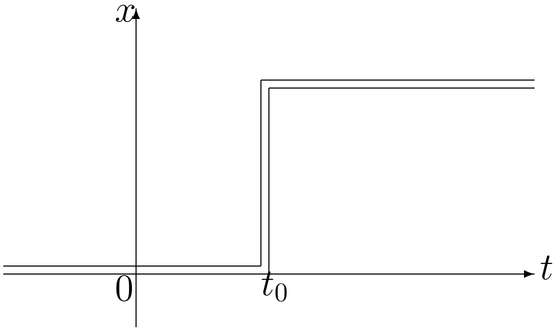
15. What will happen if we use first-order differential equations: mathematical analysis

- In general, as we have mentioned, a first-order differential equation has the form $\dot{x} = F$, for some function F describing force.
- In principle, at different moments of time, we can apply different forces.
- From this viewpoint, the basic, elementary effect is the application of a force which is active only at one specific moment of time.
- All other forces can be represented as combinations of such instantaneous forces applied at different moments of time.
- In mathematical terms, such an elementary force is described by a so-called delta-function $\delta(t - t_0)$.
- This function is equal to 0 for all other moments of time $t \neq t_0$.
- Let us therefore consider the first-order differential equation describing this force, i.e., the equation $\dot{x} = \delta(t - t_0)$.

16. What will happen if we use first-order differential equations: mathematical analysis (cont-d)

- For $t \neq t_0$, the derivative $\dot{x}$ of the function $x(t)$ is equal to 0.
- So both for $t < t_0$ and for $t > t_0$, the function $x(t)$ does not change.
- In other words, it is equal to a constant for $t < t_0$, and it is equal to another constant when $t > t_0$.
- It is reasonable to start with $x(t) = 0$, so for $t < t_0$, we have $x(t) = 0$.
- At the point t_0, the value of $x(t)$ jumps to some other value.
- So, we have the following step function $x(t)$:
  - for $t < t_0$, we have $x(t) = 0$, and
  - for $t > t_0$, we have $x(t) = c$ for some constant c.

17. What will happen if we use first-order differential equations: mathematical analysis (cont-d)



- Let us analyze what this means for both our examples.

18. Case of degrees of confidence

- In this case, we go from no confidence to the next level – which is natural to be associated with full confidence.
- This corresponds to the usual Boolean 2-valued logic, and this corresponds to sets – or, to be more precise, to infinite or final intervals.
- Specifically, when we have only one delta-function, we get an infinite interval.
- However, if we add another (negative) delta-function at the endpoint of the desired interval, we will get a finite interval as well.

19. Case of an activation function

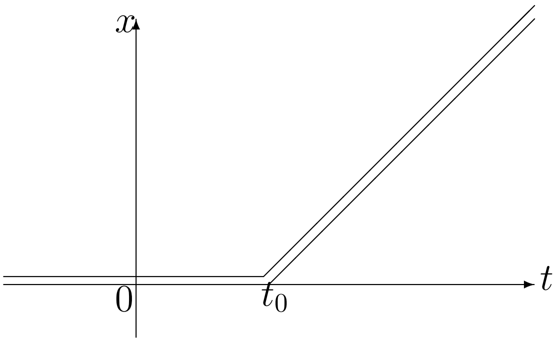
- For activation functions, a step-function is a good approximation to the *sigmoid* function.
- Sigmoid is the activation function that was, until recently, most widely used in artificial neural networks.

20. What will happen if we use second-order differential equations: mathematical analysis

- Second-order differential equations have the form $\ddot{x} = F$.
- As in the first-order case, the elementary force is a delta-function, so we get the equation $\ddot{x} = \delta(t - t_0)$.
- As we have mentioned earlier, this equation can be equivalently re-written as $\dot{v} = \delta(t - t_0)$, where we denoted $v \stackrel{\text{def}}{=} \dot{x}$.
- In terms of v, this is the same equation that we considered when we analyzed first-order equations.
- So, we can use the solution that we derived during that analysis.
- Namely, we conclude that:
  - for $t < t_0$, we have $\dot{x}(t) = v(t) = 0$, and
  - for $t > t_0$, we have $\dot{x}(t) = v(t) = c$ for some constant c.
- We know the derivative $\dot{x}(t)$ of the desired function $x(t)$.

21. What will happen if we use second-order differential equations: mathematical analysis (cont-d)

- So, to reconstruct this function, we can simply integrate the above expression.
- If we start with 0 as before, we get the following result:
  - for $t < t_0$, we have $x(t) = 0$, and
  - for $t > t_0$, we have $x(t) = c \cdot (t - t_0)$.



- Let us analyze what this means for both our examples.

22. What it means

- For degrees of confidence:
  - instead of only two possible values – as in the case of first-order equations,
  - we have degrees of confidence that continuously change from 0 to larger values.
- This is one of the main ideas behind *fuzzy logic*.
- Moreover, what we get is a linear shape of this increase.
- This is exactly what we get when we use *triangular membership functions* – the most widely used shape of membership functions.
- For activation functions, what we get is exactly *Rectified Linear Unit* (ReLU, for short).
- This is currently the most widely used activation function.

23. Conclusion

We showed that a natural optimization-motivated transition from first-order to second-order differential equations explains transitions:

- in physics,
- in describing degrees of confidence, and
- in describing activation functions in neural networks:

application area	first-order differential equations	second-order differential equations
physics	Aristotelian	Newtonian
degrees of confidence	sets, intervals	fuzzy sets, triangular membership functions
neural networks	sigmoid activation function	ReLU activation function

24. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).