Formalizing the Informal, Precisiating the Imprecise: How Fuzzy Logic Can Help Mathematicians and Physicists by Formalizing Their Intuitive Ideas

Olga Kosheleva[1], Renata Reiser[2], and Vladik Kreinovich[1]

[1]University of Texas at El Paso, El Paso, TX 79968, USA
{olgak,vladik}@utep.edu
[2]Universidade Federal de Pelotas, 96010-610, Pelotas, RS, Brazil
renata.reiser@gmail.com

Abstract. Fuzzy methodology transforms expert ideas – formulated in terms of words from natural language – into precise rules and formulas. In this paper, we show that by applying this methodology to intuitive physical and mathematical ideas, we can get known fundamental physical equations and known mathematical techniques for solving these equations. This fact makes us confident that in the future, fuzzy techniques will help physicists and mathematicians to transform their imprecise ideas into into new physical equations and new techniques for solving these equations.

1 Fuzzy and Physics: Past and Present

Fuzzy methodology: main objective. Fuzzy methodology has been invented to transform expert ideas – formulated in terms of words from natural language – into precise rules and formulas, rules and formulas understandable by a computer (and implementable on a computer); see, e.g., [6, 10, 12].

Fuzzy methodology: numerous successes. Fuzzy methodology has led to many successful applications, especially in intelligent control [6, 10].

What is fuzzy methodology good for? Traditional viewpoint. When are soft fuzzy techniques mostly used now? Let us take, as an example, *control*, which is one of the major success stories of fuzzy methods.

In control, if we know the exact equations that describe the controlled system, and if we know the exact objective function of the control, then we can often apply the optimal control techniques developed in traditional (crisp) control theory and compute the *optimal* control.

Even in these situations, we can, in principle, use soft computing methods instead: e.g., we can use simpler fuzzy control rules instead of (more complicated) traditional techniques. As a result, we may get a control that is much easier to compute but that it somewhat worse in quality.

However, the major application of fuzzy methodology in control is to the situations when we only have *partial* knowledge about the controlled system and about the objective functions and in which, therefore, traditional optimal control theory is not directly applicable. Here is where all known success stories come from: utilities like washing machines or camcoders, car parking automation, and other applications all share one thing: they all have to operate in a partially known environment.

From this viewpoint, as we gain more and more knowledge about a system, a moment comes when we do not need to use fuzzy techniques any longer: when we have accumulated enough knowledge, we will then be able to use traditional (crisp) techniques.

From this viewpoint, fuzzy techniques look like a (successful but still) *intermediate* step, "poor man's" data processing techniques, that need to be used only if we cannot apply "more optimal" traditional methods.

Physics and applied mathematics as application areas. When we study the physical world, our first objective is to find the *physical laws*, the *equations* that describe how the values of physical quantities change with time.

Once we have found these equations, the next task is *mathematical*: we need to solve these equations to predict the future values of physical quantities.

Both tasks are not easy. In both tasks, we start with informal ideas, and gradually move to exact equations and exact algorithms for solving these equations.

Current use of fuzzy techniques in physics and applied mathematics. The current use of fuzzy techniques in physics and applied mathematics following the similar lines as other applications to fuzzy methodology: fuzzy techniques are useful when we do not know the exact equations describing the corresponding system; see, e.g., [3, 6, 10, 11] and references therein.

Once we know these equations, more traditional (crisp) techniques can be used to solve these equations.

2 Fuzzy and Physics: Towards the Future

A natural new idea of using fuzzy techniques. As we have mentioned, physics research has two main objectives:

- to find the *physical laws*, the *equations* that describe how the values of physical quantities change with time, and
- to solve these equations, so that we will be able to predict the future values of physical quantities.

In both tasks, we start with *informal* ideas, and gradually move to *exact* equations and *exact* algorithms for solving these equations.

But transforming informal ideas into exact ones, their *precisiation* is exactly what fuzzy techniques have been invented for. It is therefore reasonable to use fuzzy techniques for this precisiation.

In this section, we will show that fuzzy logic techniques can indeed help in transforming informal ideas into exact equations and algorithms.

Comment. Preliminary versions of some of the results from this section first appeared in [4, 7, 8]; similar results are also presented in [9].

2.1 Case Study: Newton's Physics

Let us start our analysis with the simplest example of physical equations: namely, with the equations of Newton's physics.

Newton's physics: informal description. Let us consider a simple case when we have a single body in a potential field $V(x)$. It is a commonsense knowledge that a body usually tries to go to the points x where its potential energy $V(x)$ is the smallest. For example, a rock left at the top of the mountain, when it starts moving, it may sometimes move up (due to the original push), but mostly it tries to go down.

If we take friction into account, then a body also tries to stop. In the idealized case when there is no friction, there is a conservation of energy: the sum of the potential energy $V(x)$ and the kinetic energy $K = \dfrac{1}{2} \cdot m \cdot \sum_{i=1}^{3} \left(\dfrac{dx_i}{dt} \right)^2$ is constant. Thus, when the body minimizes its potential energy, it thus tries to maximize its kinetic energy.

What we plan to do. The above text does not sound like a very accurate description of a physical system. However, we will show that when we apply the usual fuzzy methodology to this description, we get a very precise formulation – all the way to Newton's equations

$$m \cdot \frac{d^2 x_i}{dt^2} = -\frac{\partial V}{\partial x_i}. \tag{1}$$

We will perform this derivation step-by-step.

First step: selecting a physically meaningful membership function corresponding to "small $V(x)$". The body tries to get to the areas where the potential energy $V(x)$ is small. "Small" is an imprecise word from natural language. For such words, the fuzzy methodology recommends to select a membership function $\mu(V)$ describing, for each possible value V, to what extent this value V is small.

How membership functions are determined: one of the possible ways. For each individual value V, the value $\mu(V)$ can be obtained, e.g., by polling several (n) experts and assigning, as $\mu(V)$, the ratio $\mu(V) = \dfrac{n(V)}{n}$, where $n(V)$ is the number of experts who answered that this value V is small.

Specifics of a physical system. In assigning the appropriate membership function, we must take into account the specifics of the physical system. One of

the features of this physical system (see, e.g., [5]) is that the potential energy has no absolute numerical value. All we know is the *relative* potential energy relative to some level. If we change that level by some value V_0, then all the numerical values of the potential energy get shifted, from V to $V + V_0$.

Crudely speaking, this means that the numerical values V and $V + V_0$ may represent the exact same value of the potential energy – but measured in comparison to different levels.

How to describe these physical specifics: first try. A seemingly natural formalization of this idea is to simply require that the degrees to which the values V and $V + V_0$ are small should be the same: $\mu(V) = \mu(V + V_0)$.

The first try does not work. However, this formalization does not work: if we require that this equality holds for all V and V_0, then for every two real numbers V and V', by taking $V_0 = V' - V$, we would be able to conclude that $\mu(V) = \mu(V + V_0) = \mu(V')$ – and thus, that the resulting membership function is simply constant. This does not make intuitive sense, since we know that the smaller the value V, the larger should be our confidence that V is small.

A better idea. We therefore cannot simply require that the functions $\mu(V)$ and $\mu(V + V_0)$ corresponding to two different levels are identical. However, we should require that these two membership functions be, to some extent, equivalent to each other.

How to formalize this idea: re-analyzing the polling method. How can we formalize this idea? To do that, let us go back to the polling method of determining a membership function.

Our objective is to find the value $\mu(V)$ as accurately as possible. It is known that in the poll, the more people we ask, the more accurate is the resulting opinion. Thus, a natural way to improve the accuracy of the poll is to ask more experts. However, there is a catch. When at first, we could only afford to poll n people, we thus selected top experts in the field. Now that we add m extra folks, these folks may be too intimidated by the original experts to voice their opinions – especially in case the original experts disagree. With the new experts mute, we still have the same number $n(V)$ of experts who believe that the value V is small – but now we have to divide it not by the original number n, but by the new number $n + m$. As a result, instead of the original value $\mu(V) = \dfrac{n(N)}{n}$, we get a new value $\mu'(V) = \dfrac{n(N)}{n + m}$. It is easy to see that $\mu'(V) = c \cdot \mu(V)$, where $c = \dfrac{n}{n + m}$.

Thus, for the exact same opinion, by selecting two different numbers of experts n and $n+m$, we get two numerically different membership functions: $\mu(V)$ and $c \cdot \mu(V)$. These two membership functions represent the same expert opinion and are, thus, equivalent in some reasonable sense.

Resulting formalization of the physical intuition. Now, we have a meaningful interpretation of the requirement that the membership functions $\mu(V)$

and $\mu(V + V_0)$ – corresponding to two different starting levels for measuring potential energy – are equivalent: that for every V_0, there should be a value $c(V_0)$ for which $\mu(V + V_0) = c(V_0) \cdot \mu(V)$.

Resulting selection of the membership function. This functional equation is known (see, e.g., [1]). Its only monotonic solution is a function

$$\mu(V) = a \cdot \exp(-k \cdot V).$$

So we will use this exponential function to describe the fact that potential energy should be small.

Comment. Since, as we have mentioned, the membership function is determined modulo a factor c, we can, for simplicity, set a to 1 and get an even simpler formula $\mu(V) = \exp(-k \cdot V)$.

Second step: selecting a membership function corresponding to "large value of kinetic energy K". As we have mentioned, kinetic energy tends to increase, i.e., should be large.

Instead of starting a derivation from scratch, let us use the fact that we already have a physically meaningful membership function for "small". Intuitively, a value K is large if $-K$ is small.

So, the statement "kinetic energy K is large" is equivalent to saying "the value $-K$ is small". By using the above membership function for small, we thus conclude that the membership function describing our intuition about the kinetic energy is $\mu(K) = \exp(-k \cdot (-K)) = \exp(k \cdot K)$.

Third step: selecting a physically meaningful t-norm ("and"-operation). We want to describe the intuition that the potential energy is small *and* that the kinetic energy is large *and* that the same is true at different moments of time. According to fuzzy methodology, we must therefore select an appropriate "and"-operation (t-norm) to combine our degrees of certainty in individual statements into a single degree describing the degree to which we believe in the composite statement.

Let us use physical intuition to select such a t-norm $f_\&(a, b)$.

Specific of the physical system. In principle, if we have two completely independent systems, we can consider them as a single system. Since these systems do not interact with each other, the total energy E of the combined system is simply equal to the sum $E_1 + E_2$ of the energies of the components.

Using the physical specifics. Intuitively, if both component energies are small, then the resulting total energy should also be small. We can therefore estimate the smallness of the total energy in two different ways:

– first, we can simply apply the above membership function "small" to the total energy $E = E_1 + E_2$, and get the value $\mu(E_1 + E_2)$;
– second, we can first estimate the degrees $\mu(E_1)$ and $\mu(E_2)$ to which each of the components is small, and then use a t-norm $f_\&(a, b)$ to combine these degrees into a degree that E_1 is small *and* E_2 is small: $f_\&(\mu(E_1), \mu(E_2))$.

In view of the above motivation, it is reasonable to require that these two estimates should coincide, i.e., that we should have $\mu(E_1 + E_2) = f_\&((\mu(E_1), \mu(E_2)))$. We know that $\mu(E) = \exp(-k \cdot E)$, thus, we conclude that the following equality should hold for all E_1 and E_2:

$$\exp(-k \cdot (E_1 + E_2)) = f_\&(\exp(-k \cdot E_1), \exp(-k \cdot E_2)).$$

This requirement enables us to uniquely determine the corresponding t-norm. Namely, to find the value $f_\&(a_1, a_2)$, we must first find the values E_i for which $\exp(-k \cdot E_i) = a_i$. For these values, we then have

$$f_\&(a_1, a_2) = \exp(-k \cdot (E_1 + E_2)) = \exp(-k \cdot E_1) \cdot \exp(-k \cdot E_2) = a_1 \cdot a_2.$$

Resulting selection of a t-norm. Thus, the physically meaningful t-norm is the algebraic product $f_\&(a_1, a_2) = a_1 \cdot a_2$.

Resulting model. Now, we are ready to estimate to what extent a given trajectory $x(t)$ satisfies the intuitive ideas that the potential energy be small and the kinetic energy be large at all moments of time $t_1, \ldots, t_N$. We know the degrees to which each of these requirements is satisfied at each moment of time, so to get the overall degree, we can simply multiply all these degrees. As a result, we get the following product: $\prod_{i=1}^{N} \exp(-k \cdot V(t_i)) \cdot \prod_{i=1}^{N} \exp(k \cdot K(t_i))$. Since, as we have already mentioned,

$$\exp(-k \cdot a) \cdot \exp(-k \cdot b) = \exp(-k \cdot (a + b)),$$

this expression can be reformulated as $\exp(-k \cdot S)$, where $S \stackrel{\text{def}}{=} \sum_{i=1}^{N} (V(t_i) - K(t_i))$.

It is reasonable to select, as the most reasonable, a trajectory for which our degree of confidence that this trajectory is reasonable is the highest. To find such a trajectory, we must maximize the value $\exp(-k \cdot S)$. Since the function $\exp(-k \cdot S)$ is strictly decreasing, this is equivalent to minimizing S.

So, we arrive at the requirement that we should minimize the sum S. In reality, the number of moments of time is infinite, so instead of a sum, we get an integral $S \sim \int L \, dt$, where we denoted

$$L = V(t) - K(t) = V(t) - \frac{1}{2} \cdot m \cdot \sum_{i=1}^{3} \left(\frac{dx_i}{dt}\right)^2.$$

This model leads to Newton's equations. In modern physics, most physical laws are formulated in terms of the Principle of Least Action, according to which the trajectory is selected in such a way that the *action* $S = \int L \, dt$ is the smallest possible. In particular, for Newtonian physics, the exact same expression S – as we came up with based on fuzzy methodology – leads *exactly* to Newton's laws; see, Appendix A.

Comment. With the fuzzy approach, we not only get the most reasonable New-ton's trajectory, we also get the degree $\exp(-k \cdot S)$ with which all other trajectories are reasonable. In Newton's physics, only one trajectory is possible, but in quantum physics, non-Newtonian trajectories are also possible, and the "amplitude" of each trajectory is determined by exactly this formula $\exp(-k \cdot S)$, albeit with a complex value k. This fact makes the above derivation even more interesting.

2.2 Beyond the Simplest Netwon's Equations

Need to go beyond Newton's equations. In our analysis of the Newton's equations, we assume that the expression for the potential energy $V(x)$ is given. However, in reality, this expression also needs to be determined. The potential energy represents a *field* – e.g., electrostatic, gravitational, etc. – so, in addition to mechanics, we must also find the equations that describe the corresponding field.

Gravitational field: main idea. Let us consider the simplest case of a gravitational field. We will consider it in the Newtonian approximation, where it is described by a scalar function $V(x)$.

The main physical property of the gravitational field is that it changes very slowly: gravitational pull of the Earth, for example, is caused by the Earth as a whole, so if we move a little bit, we still feel approximately the same gravitation. It is a known empirical fact that the differences in the gravitational field at different earth locations are very small (but, by the way, very important for geophysics, because they provide a good overall understanding of what is located below the Earth surface).

Thus, all the components $\partial V \, \partial x_i$ of the gradient of the gravitational field must be small. This situation is similar to kinetic energy and different from potential energy in the sense that we want these values to be close to 0. Similarly to the case of kinetic energy, this is equivalent to requiring that the squares of the derivatives be small.

Derivation of the resulting model. Thus, we arrive at the condition that for all locations x, all squares of partial derivatives must be small. For each location and for each i, the corresponding requirement that the square of the derivative is small can be described by the degree $\exp\left(-k \cdot \left(\dfrac{\partial V}{\partial x_i}\right)^2\right)$. By using the product t-norm to combine these values, we get the expression $\prod\limits_x \prod\limits_{i=1}^3 \exp\left(-k \cdot \left(\dfrac{\partial V}{\partial x_i}\right)^2\right)$. As in the Newton's case, this expression can be represented as $\exp(-k \cdot S)$, where $S = \sum\limits_x \sum\limits_{i=1}^3 \left(\dfrac{\partial V}{\partial x_i}\right)^2$. Taking into account that we have infinitely many spatial locations x, we get an integral instead of the sum: $S = \int L \, dx$, where $L(x) = \sum\limits_{i=1}^3 \left(\dfrac{\partial V}{\partial x_i}\right)^2$.

This model leads to Newton's formulas for the gravitation force. It is known that minimizing this expression leads to the equation $\sum\limits_{i=1}^{3} \dfrac{\partial^2 V}{\partial x_i^2} = 0$, that leads to Newton's gravitational potential $V(x) \sim \dfrac{1}{r}$ that, in turns, leads to the known expression for the gravitational force $F \sim r^{-2}$.

Comment. Similar arguments can lead to other known action principles and thus, to other fundamental physical equations.

2.3 From Equations to Solutions: Fuzzy Techniques Help to Deal with Divergent Series

Small-parameter method. Once we know the equations that describe the dynamics of the corresponding particles and/or fields, a natural next step is to solve these equations under the given information – and thus, predict the future values of the corresponding physical quantities. The equations are often complex, and in many situations, no analytical solution is known, so we have to consider approximate methods.

The complexity of solving a system of complex equations is often eased by the fact that our knowledge is usually incremental. At any given moment of time, we have a model which is a reasonably good approximation to reality, and then we find a new model which is even more accurate. The ideas behind the new model may be revolutionary – as they were for quantum physics or relativity theory – but in terms of predictions, the new theories usually provide a small adjustment to the previous known one. For example, General Relativity was confirmed when it turned out that it better describes the bending of light near the Sun – better by 1.75 arc-seconds.

Usually, by the time new complex equations appear, we already know how to solve previous equations. Thus, we can use the solution x_0 to the previous equations as a first approximation to the solution x of the new equations.

The difference $x - x_0$ between these solutions can be characterized by some small parameter q. The original solution x_0 (to the previous theory) corresponds to taking into account only the 0-th order term in the Taylor expansion of x into a series in terms of q. To get a better approximation, we can take into account terms which are linear in q, terms which are quadratic in q, etc. Ideally, we thus get an expression for x as an infinite power series; see, e.g., [5]:

$$x = \sum_{i=0}^{\infty} q^i \cdot x_i = x_0 + q \cdot x_1 + q^2 \cdot x_2 + \dots \tag{2}$$

In practice, as an estimate for x, we compute the first few terms in this sum

$$s_k \stackrel{\text{def}}{=} \sum_{i=0}^{k} q^i \cdot x_i. \tag{3}$$

In general, the more terms we take – i.e., the larger cutoff value k we use – the more accurate is the resulting estimate.

The Taylor series method also provides us with a reasonable estimate of the accuracy of the next approximation, i.e., of the approximation error

$$\Delta x_k \stackrel{\text{def}}{=} x - x_k = \sum_{i=k+1}^{\infty} q^i \cdot x_i;$$

namely, since we assume that the terms decrease with k, the first ignored term $q^{k+1} \cdot x_{k+1}$ provides a reasonably accurate description of the approximation error.

This method often works well. In many cases, this idea works very well [5]. It works, e.g., in celestial mechanics, when the two-body problem – which describes e.g., how the Earth goes around the Sun – has an explicit analytical solution, and we would like to analyze how the presence of the Moon affects this solution. In this problem, the Moon is much lighter than the Earth, so the ratio of these mass $m_{\text{Moon}}/m_{\text{Earth}}$ is the desired small dimensionless parameter q.

Sometimes, this method leads to divergent series. In some other cases, the small-parameter method only works for small k: we get a good approximation s_0, a more accurate approximation s_1, an even more accurate approximation s_2, etc. – until we reach a certain threshold k_0. Once this threshold is reached, the approximation accuracy decreases. In other words, the series (2) diverge. This is, e.g., of Taylor series corresponding to quantum electrodynamics (see, e.g., [2, 5]): the first few terms of the expansion in the weak interaction constant $\alpha \approx 1/137$ lead to very accurate predictions, but the whole series diverge.

Divergence is one of the main problems of quantum field theory. The above divergence is one of the main challenges preventing physicists from coming up with exact mathematical formulations of quantum field theory.

Fuzzy techniques can potentially help in solving this problem. Divergence is largely a theoretical problem; in practice, physicists use semi-heuristic methods to come up with meaningful predictions. Formalizing imprecise semi-heuristic ideas is one of the main reasons why fuzzy techniques were invented in the first place. Let us therefore try to use fuzzy techniques to formalize the physicists' reasoning.

How physicists use divergent series. Let us describe how physicists come up with answers when the series converge.

When the series representing the answer are divergent, physicists usually consider only the approximations until the remaining term $s_{k+1} - s_k$ starts increasing. In other words, the last approximation they consider is the one for which the difference $s_{k+1} - s_k$ is smaller than both the previous difference $s_k - s_{k-1}$ and the next difference $s_{k+2} - s_{k+1}$.

Usually, the difference is in the orders of magnitude – just like for Taylor series in general, so we have $s_{k+1} - s_k \ll s_k - s_{k-1}$ and $s_{k+1} - s_k \ll s_{k+2} - s_{k+1}$.

Challenge. The divergent character of the corresponding series presents a mathematical challenge: how do we formalize the idea that while the series diverge, its first terms serve as a good approximation?

Let us show that fuzzy logic allows us to come up with a mathematically rigorous formalization of this idea.

Towards a fuzzy solution to the challenge. Since the series (2) diverge, the corresponding sum makes no mathematical sense, so we cannot formulate equation (2) in the literal form. Instead, let us formalize exactly what the physicists are doing: they are claiming, in effect, that, for every k, $x \approx s_k$ with an accuracy proportional to the next ignored term, i.e., to the difference $s_{k+1} - s_k$. In other words, instead of the equation (2), we have a sequence of infinitely many imprecise rules of the type

$$x \approx x_0 + q \cdot x_1 + q^2 \cdot x_2 + \ldots + q^k \cdot x_k, \text{ with accuracy } q^{k+1} \cdot x_{k+1},$$

or, equivalently,

$$x \approx s_k \text{ with accuracy } s_{k+1} - s_k. \tag{4}$$

How to describe the relation "x is approximately equal to a, with accuracy of order σ". We want to describe, for every three real numbers x, a, and $\sigma > 0$, the degree $\mu(x, a, \sigma)$ to which x is approximately equal to a with an accuracy of order σ.

First natural condition. Intuitively, when one of these values changes a little bit, this degree should also change only a little bit. Thus, this function should be *continuous*. to a with an accuracy of order σ.

Second natural condition. This degree should be equal to 1 when $x = a$ and it should strictly decrease to 0 as x increase up from a or strictly decrease to 0 as x decreases down from a.

Third natural condition. We want to apply this function to values of physical quantities. The numerical value of a physical quantity depends on the choice of a measuring unit and on the choice of a starting point. It is reasonable to require that the degree $\mu(x, a, \sigma)$ should not change if we simply re-scale the corresponding physical quantities by using a new measuring unit or a new starting point for measuring this quantity.

Changing units. If we replace a measuring unit by a new unit which is λ times smaller, then the numerical value increases by a factor of λ: $x \to \lambda \cdot x$. For example, if we replace meters with centimeters which are $\lambda = 100$ times smaller, then a length of $x = 2$ m becomes $x' = 200$ cm.

Since accuracy is measured in the same units, in the new units, we have $\sigma' = \lambda \cdot \sigma$. to a with an accuracy of order σ.

So, invariance relative to selection of a measuring unit means that for every $\lambda > 0$, we should have $\mu(\lambda \cdot x, \lambda \cdot a, \lambda \cdot \sigma) = \mu(x, a, \sigma)$.

Changing sign. Sometimes, the sign of a physical quantity is also arbitrary, so it can change $x \to -x$. For example, the direction of a spatial coordinate is a pure convention. It is also a pure convention that we consider electrons to be negatively changed and protons positively changed; all the formulas of electrodynamics remain the same if we simply change the signs of all the electric charges.

Accuracy σ describes the *absolute value* $|x - a|$ of the difference $x - a$, so the value of σ does not change if we simply change the sign of the corresponding physical quantity. to a with an accuracy of order σ.

Changing starting point. Also, if we replace the original starting point with a new starting point which is x_0 units lower, then all numerical values are increased by x_0: $x \to x + x_0$. Since the accuracy σ estimates the value of the difference $x - a$, the value of σ does not change under this transformation. to a with an accuracy of order σ.

Fourth natural condition. Finally, often, we have several estimates of this type; we should be able to combine them into a single estimate. In other words, for every finite set of values a_i and σ_i, we should describe the "and"-combination of all the rules of these types by a single rule of a similar type. We have already argued that algebraic product is a good way to formalize "and". To formulate this requirement in precise terms, we need to take into account that we are usually interested in normalized membership functions, for which $\max_x \mu(x) = 1$, but the product of two or more membership functions is not necessarily normalized. Thus, we need to normalize this product.

Proposition 1. *Let $\mu(x, a, \sigma)$ be a $[0, 1]$-valued continuous function with the following properties:*

- *for any a and σ, $\mu(a, a, \sigma) = 1$;*
- *for any a and σ, the value $\mu(x, a, \sigma)$ strictly decreases for $x \geq a$ and strictly increases for $x \leq a$, and tends to 0 as $x \to \pm\infty$;*
- *for every x, a, σ, and $\lambda > 0$, we have $\mu(\lambda \cdot x, \lambda \cdot a, \lambda \cdot \sigma) = \mu(x, a, \sigma)$;*
- *for every x, a, σ, we have $\mu(-x, -a, \sigma) = \mu(x, a, \sigma)$;*
- *for every x, a, σ, and x_0, we have $\mu(x + x_0, a + x_0, \sigma) = \mu(x, a, \sigma)$;*
- *for every $a_1, \ldots, a_n$, $\sigma_1, \ldots, \sigma_n$, there exist values a, σ, and C for which, for all x, we have $\mu(x, a_1, \sigma_1) \cdot \ldots \cdot \mu(x, a_n, \sigma_n) = C \cdot \mu(x, a, \sigma)$.*

Then, $\mu(x, a, \sigma) = \mu_0 \left(\dfrac{x - a}{\sigma} \right)$, where $\mu_0(z) = \exp(-\beta \cdot z^2)$ for some $\beta > 0$.

Comment. For readers' convenience, all the proofs are placed in the Appendix.

Back to our problem. Proposition 1 shows that a natural interpretation of the phrase "$x \approx a$ with accuracy σ" is provided by a membership function $\mu(x) = \exp\left(-\beta \cdot \dfrac{(x - a)^2}{\sigma^2}\right)$. So, the degree to which each rule is satisfied for a given k is equal to $\exp\left(-\beta \cdot \dfrac{(x - s_k)^2}{(s_{k+1} - s_k)^2}\right)$. The degree to which the 1st

rule is satisfied, *and* the 2nd rule is satisfied, ..., can be found by applying the corresponding "and"-operation (i.e., the product) of the above degrees. If we use the product as an "and"-operation, then the degree to which all the rules are satisfied is equal to

$$\prod_{k=0}^{\infty} \exp\left(-\beta \cdot \frac{(x-s_k)^2}{(s_{k+1}-s_k)^2}\right) = \exp\left(-\beta \cdot \sum_{k=0}^{\infty} \frac{(x-s_k)^2}{(s_{k+1}-s_k)^2}\right).$$

An infinite product can be zero, so we have to consider products corresponding to large N and then tend N to infinity in the resulting formulas. For a finite value N, we get

$$\mu_N(x) = \prod_{k=0}^{N} \exp\left(-\beta \cdot \frac{(x-s_k)^2}{(s_{k+1}-s_k)^2}\right) = \exp\left(-\beta \cdot \sum_{k=0}^{N} \frac{(x-s_k)^2}{(s_{k+1}-s_k)^2}\right). \quad (5)$$

If we need to select a single value x, it is reasonable to select a value for which the corresponding degree of belief is the largest possible. Since the function $\mu_N(x)$ is of the form $\exp(-z_N)$ for some z_N, and the function $\exp(-z)$ is strictly decreasing, its maximum is attained at a point X_N at which z_N attains its minimum: $\sum_{k=0}^{N} \frac{(x-s_k)^2}{(s_{k+1}-s_k)^2} \to \min$. Differentiating this expression with respect to x, we conclude that $X_N = \dfrac{\sum_{k=0}^{N} s_k \cdot (s_{k+1}-s_k)^{-2}}{\sum_{k=0}^{N} (s_{k+1}-s_k)^{-2}}$. The actual solution corresponds to $N \to \infty$. Thus, we arrive at the following conclusion.

The resulting solution: a formal definition. In the general case of a series (convergent or divergent), we select

$$x = \lim_{N\to\infty} \frac{\sum_{k=0}^{N} s_k \cdot (s_{k+1}-s_k)^{-2}}{\sum_{k=0}^{N} (s_{k+1}-s_k)^{-2}}. \quad (6)$$

Comment. It is worth mentioning that if, instead of using fuzzy techniques, we use probabilistic techniques and assume that $x \approx s_k$ with Gaussian approximation error of 0 mean and standard deviation proportional to $|s_{k+1} - s_k|$, we get the exact same estimate.

Analysis of the resulting formula. Let us show that the above covers both the case of a convergent series – in which case it coincides with the limit $\lim s_k$ – and of the divergent series.

Case of divergent series. For *divergent* series, the analysis is simple. If we have

$$|s_1 - s_0| \gg |s_2 - s_1| \gg \ldots \gg |s_k - s_{k-1}| \gg |s_{k+1} - s_k| \ll |s_{k+2} - s_{k+1}| \ll \ldots,$$

then
$$(s_1 - s_0)^{-2} \ll (s_2 - s_1)^{-2} \ll \ldots \ll (s_k - s_{k-1})^{-2} \gg$$
$$(s_{k+1} - s_k)^{-2} \gg (s_{k+2} - s_{k+1})^{-2} \gg \ldots$$

Thus, in the numerator $\mathcal{N}$ of the formula (6), the main term is the term $\mathcal{N} \approx s_k \cdot (s_{k+1} - s_k)^{-2}$ corresponding to the smallest possible difference $|s_{k+1} - s_k|$. Similarly, the denominator $\mathcal{D}$ is approximately equal to $\mathcal{D} \approx (s_{k+1} - s_k)^{-2}$. Thus, the ratio $\dfrac{\mathcal{N}}{\mathcal{D}}$ is approximately equal to s_k – which is exactly what physicists conclude now.

Comment. To be more precise, the physicists use the *next* value s_{k+1}, which, within the accuracy $s_{k+1} - s_k$, is the same as s_k.

Case of convergent series. For a *convergent* series, the result is the same as the limit:

Proposition 2. *When $s_k \to x$, then the expression* (6) *coincides with the limit x.*

Comment. To provide a better understanding of the formula (6), in Appendix D, we provide an example of applying this formula to the diverging geometric series $\sum z^i$ with $|z| \geq 1$.

General comment. The membership function corresponding to the approximation based on the first $N + 1$ sums has the form

$$\mu_N(x) = \exp\left(-\frac{(x - x_N)^2}{\sum\limits_{k=0}^{N} (s_{k+1} - s_k)^{-2}} \right).$$

Thus:

- For the case when $s_k \to x$, the sum $\sum\limits_{k=0}^{N} (s_{k+1} - s_k)^{-2}$ tends to infinity. So, the membership value $\mu_N(x)$ tends to 1 for this x and to 0 for all other x. Thus, in the limit $N \to \infty$, we have a crisp conclusion.
- In contract, in the divergence case, the sum is approximately equal to the value $(s_{k+1} - s_k)^{-2}$ corresponding to the smallest difference $|s_{k+1} - s_k|$. Thus, the limit values $\mu_N(x)$ remain non-zero within a neighborhood of size $\approx |s_k - s_{k+1}|$ – which reflects the fact that in this case, we do not have a precise value of x, we can only determine this value with the accuracy $|s_{k+1} - s_k|$.

3   Fuzzy and Physics: Promising Future

Fuzzy techniques will help to derive new physical equations. In the previous section, we showed that fuzzy techniques can transform informal physical ideas into exact physical equations, and that in this way, we can derive many

known physical equations. From the viewpoint of practical applications, from the viewpoint of being able to predict new physical phenomena, we have not yet achieved anything new: all we did was found a new justification for the already known equations.

However, the fact that the existing fuzzy methodology enables us to transform informal ("fuzzy") description of physical phenomena into well-known physical equations makes us confident that in the future, when new physical phenomena will be discovered, fuzzy methodology may help generate the equations describing these phenomena.

Fuzzy techniques will help to solve physical equations. Similarly, the fact that fuzzy techniques can lead to an explanation of the known heuristic methods for solving physical equations makes us confident that in the future, similarly fuzzy techniques will help to transform informal ideas into new successful mathematical techniques.

Instead of conclusion: the future is fuzzy. People often say "the future is fuzzy" meaning that it is difficult to predict the future exactly. But, based on what we observed, we can claim that "the future is fuzzy" in a completely different sense: that the future will see more and more applications of fuzzy techniques, including applications to areas like theoretical physics and numerical mathematics, areas where, at present, there are not many applications of fuzzy. The future is fuzzy!

References

1. J. Aczel, *Lectures on Functional Equations and Their Applications*, Dover Publ., New York, 2006.
2. A. I. Akhiezer and V. B. Berestetsky, *Quantum Electrodynamics*, Interscience Publishers, New York, 1965.
3. M. del Catillo-Mussot and R. Costa Dias, "Fuzzy sets and physics", *Revista Mexicana de Física*, 1993, Vol. 39, No. 2, pp. 295–303.
4. E. Gutierrez and V. Kreinovich, "Fundamental physical equations can be derived by applying fuzzy methodology to informal physical ideas", *Proceedings of the 30th Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2011*, El Paso, Texas, March 18–20, 2011.
5. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.
6. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Upper Saddle River, New Jersey: Prentice Hall, 1995.
7. V. Kreinovich, C.-C. Chang, L. Reznik, and G. N. Solopchenko, "Inverse problems: fuzzy representation of uncertainty generates a regularization", *Proceedings of the 1992 Conference of the North American Fuzzy Information Processing Society NAFIPS'92*, Puerto Vallarta, Mexico, December 15–17, 1992, NASA Johnson Space Center, Houston, TX, 1992, pp. 418–426.

8. V. Kreinovich, C. Quintana, and L. Reznik, "Gaussian membership functions are most adequate in representing uncertainty in measurements", *Proceedings of the 1992 Conference of the North American Fuzzy Information Processing Society NAFIPS'92*, Puerto Vallarta, Mexico, December 15–17, 1992, NASA Johnson Space Center, Houston, TX, 1992, pp. 618–625.

9. H. T. Nguyen, V. Kreinovich, and B. Bouchon-Meunier, "Soft Computing Explains Heuristic Numerical Methods in Data Processing and in Logic Programming", In: L. Medsker (ed.), *Frontiers in Soft Computing and Decision Systems*, AAAI Press (Publication No. FS-97-04), 1997, pp. 30–35.

10. H. T. Nguyen and E. A. Walker, *First Course on Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.

11. T. J. Ross, *Fuzzy Logic with Engineering Applications*, Wiley, New York, 2010.

12. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.

A   Variational Equations

General derivation. Let us recall how we can transform the Least Action Principle into a differential equation. Let us first do it on the example on Newton-type situation, where we need to find a function $x(t)$ that minimizes the following expression: $S = \int L(x, \dot{x}) \, dt \to \min$. Minimizing means, in particular, that if we take any function $\Delta x(t)$ and consider a function $S(\alpha) = x + \alpha \cdot \Delta x$, then this function must attain its maximum for $\alpha = 0$. Thus, the derivative of $S(\alpha)$ at $\alpha = 0$ must be 0. Differentiating the expression

$$S(\alpha) = \int L(x + \alpha \cdot \Delta x, \dot{x} + \alpha \cdot \Delta \dot{x}) \, dt$$

and equating the derivative to 0, we conclude that

$$\int \left(\frac{\partial L}{\partial x} \cdot \Delta x + \frac{\partial L}{\partial \dot{x}} \cdot \Delta \dot{x} \right) dt = \int \left(\frac{\partial L}{\partial x} \cdot \Delta x \right) dt + \int \left(\frac{\partial L}{\partial \dot{x}} \cdot \Delta \dot{x} \right) dt = 0.$$

Integrating the second term by parts, we conclude that

$$\int \left(\frac{\partial L}{\partial x} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) \right) \cdot \Delta x \, dt = 0.$$

This must be true for every function $\Delta x(t)$, in particular for a function that is equal to 0 everywhere except for a small vicinity of a moment t, For this function, the integral is proportional to the value of the expression $\dfrac{\partial L}{\partial x} - \dfrac{d}{dt} \left(\dfrac{\partial L}{\partial \dot{x}} \right)$ at the point t. Since the integral is 0, this expression must also be equal to 0:

$$\frac{\partial L}{\partial x} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) = 0.$$

The resulting equations are known as *Euler-Lagrange equations*.

Case of Newton's laws. In particular, for the Newton's case, when

$$L = V(x) - \frac{1}{2} \cdot m \cdot \sum_{i=1}^{3} \left(\frac{dx_i}{dt} \right)^2,$$

for each of the components $x_i(t)$, we have $\dfrac{\partial L}{\partial x_i} = \dfrac{\partial V}{\partial x_i}$ and $\dfrac{\partial L}{\partial \dot{x}_i} = -m \cdot \dfrac{dx_i}{dt}$. Thus, Euler-Lagrange's equations lead to $\dfrac{\partial V}{\partial x} + m \cdot \dfrac{d}{dt} \left(\dfrac{dx_i}{dt} \right) = 0$, i.e., to Newton's equations $m \cdot \dfrac{d^2 x_i}{dt^2} = -\dfrac{\partial V}{\partial x_i}$.

General case. In the general case, Euler-Lagrange equations take the form $\dfrac{\partial L}{\partial \varphi} - \sum_{i=1}^{3} \dfrac{\partial}{\partial x_i} \left(\dfrac{\partial L}{\partial \varphi_{,i}} \right) = 0$, where $\varphi_{,i} \overset{\text{def}}{=} \dfrac{\partial \varphi}{\partial x_i}$.

B   Proof of Proposition 1

1°. Let us first apply the condition $\mu(x+x_0, a+x_0, \sigma) = \mu(x, a, \sigma)$ with $x_0 = -a$. Then, we get $\mu(x, a, \sigma) = \mu(x - a, 0, \sigma)$, or, equivalently,

$$\mu(x, a, \sigma) = \mu_1(x - a, \sigma),$$

where we denoted $\mu_1(z, \sigma) \overset{\text{def}}{=} \mu(z, 0, \sigma)$.

2°. In terms of the function μ_1, the condition $\mu(\lambda \cdot x, \lambda \cdot a, \lambda \cdot \sigma) = \mu(x, a, \sigma)$ takes the form $\mu_1(\lambda \cdot (x - a), \lambda \cdot \sigma) = \mu_1(x - a, \sigma)$. Let us apply this condition for $\lambda = \sigma^{-1}$. Then, we conclude that $\mu_1(z, \sigma) = \mu_1 \left(\dfrac{z}{\sigma}, 1 \right)$, or, equivalently, $\mu_1(z, \sigma) = \mu_0 \left(\dfrac{z}{\sigma} \right)$, where we denoted $\mu_0(z) \overset{\text{def}}{=} \mu_1(z, 1)$.

Substituting this expression for $\mu_1(z, \sigma)$ in terms of μ_0 in the expression for μ in terms of μ_1, we conclude that $\mu(x, a, \sigma) = \mu_0 \left(\dfrac{x - a}{\sigma} \right)$.

3°. Substituting the expression for μ in terms of μ_0 into the condition $\mu(-x, -a, \sigma) = \mu(x, a, \sigma)$, we conclude that $\mu_0(-z) = \mu_0(z)$. Thus, $\mu(x, a, \sigma) = \mu_0 \left(\dfrac{|x - a|}{\sigma} \right)$.

4°. For $a_1 = a_2 = 0$ and $\sigma_1 = \sigma_2 = 1$, the fusion condition implies that

$$\mu_0(x) \cdot \mu_0(x) = C \cdot \mu_0 \left(\frac{x - a}{\sigma} \right)$$

for some a, C, and σ. The left-hand side attains its maximum ($= 1$) at $x = 0$, the right-hand side attains its maximum (which is equal to C) for $x = a$. Since these two sides are one and the same function, we conclude that $a = 0$ and $C = 1$, i.e.,

that $\mu_0^2(x) = \mu(k_2 \cdot x)$ for some constant k_2 ($= 1/\sigma$). For an auxiliary function $\ell(x) \stackrel{\text{def}}{=} \ln(\mu_0(x))$ we conclude that $2 \cdot \ell(x) = \ell(k_2 \cdot x)$.

Similarly, if we consider 3, 4, etc. terms, we conclude that $3 \cdot \ell(x) = \ell(k_3 \cdot x)$, $4 \cdot \ell(x) = \ell(k_4 \cdot x)$, etc.

$4°$. The function $\mu_0(x)$ for $x > 0$ is monotonously decreasing from 1 to 0. Therefore, $\ell(x)$ is monotonously decreasing from 0 to $-\infty$. Since μ (and thus, μ_0) is continuous, the function $\ell(x)$ is also continuous, and hence, there exists an inverse function $i(x) = \ell^{-1}(x)$, i.e., such a function that $i(\ell(x)) = x$ for every x.

For this inverse function, the equality $n \cdot \ell(x) = \ell(k_n \cdot x)$ turns into $i(n \cdot \ell(x)) = i(\ell(k_n \cdot x)) = k_n \cdot x = k_n \cdot i(\ell(x))$. So, if we denote $\ell(x)$ by X, we conclude that for every n, there exists a k_n such that $i(n \cdot X) = k_n \cdot i(X)$.

If we substitute $Y = n \cdot X$, we conclude that $i(Y) = k_n \cdot i\left(\dfrac{Y}{n}\right)$, and therefore, $i\left(\dfrac{Y}{n}\right) = \dfrac{1}{k_n} \cdot i(Y)$.

From these two equalities, we conclude that $i\left(\dfrac{m}{n} \cdot X\right) = \dfrac{1}{k_n} \cdot i(m \cdot X) = \dfrac{k_m}{k_n} \cdot i(X)$. So, for every rational number r, there exists a real number $k(r)$ such that $i(r \cdot X) = k(r) \cdot i(X)$. Therefore, the ratio $\dfrac{i(r \cdot X)}{i(X)}$ is constant for all rational r.

$5°$. Since $i(X)$ is a continuous function, and any real number can be represented as a limit of a sequence of rational numbers, we conclude that the ratio $\dfrac{i(r \cdot X)}{i(X)}$ is constant for real values of r as well. Therefore, for every real number r, there exists a $k(r)$ such that $i(r \cdot X) = k(r) \cdot i(X)$.

We have thus arrived at a functional equation for which all monotonis solutions are known: they are $i(X) = A \cdot X^p$ for some A and p; see, e.g., [1]. Therefore, the inverse function $\ell(x)$ ($x > 0$) also takes the similar form $\ell(x) = B \cdot x^m$ for some B and m. Taking into consideration that $\mu_0(x)$ and hence $\ell(x)$ are even functions, we conclude that $\ell(x) = B \cdot |x|^m$ for all x.

$6°$. Now, for every $a_1 > 0$, if we take $a_2 = -a_1$ and $\sigma_1 = \sigma_2 = 1$, then the fusion property implies that $\mu_0(x - a_1) \cdot \mu_0(x + a_1) = C \cdot \mu_0\left(\dfrac{x - a}{\sigma}\right)$ for some a and σ. The left-hand side of this equation is an even function, so the right-hand side must also be even, and therefore $a = 0$. So, $\mu_0(x - a_1)\mu_0(x + a_1) = C \cdot \mu_0\left(\dfrac{x}{\sigma}\right)$. For $x = 0$ we get $\mu_0(a_1) \cdot \mu_0(a_1) = C$. Turning to logarithms, we conclude that for every a_1, there exists a $k(a_1)$ ($= 1/\sigma$) such that $\ell(x - a_1) + \ell(x + a_1) = \ell(k(a_1) \cdot x) + 2 \cdot \ell(a_1)$. If we substitute here $\ell(x) = B \cdot |x|^m$, and divide both sides by B, we conclude that $|x - a_1|^m + |x + a_1|^m = (k(a_1))^m \cdot |x|^m + 2 \cdot a_1^m$.

Let us show that this equality is satisfied only when $m = 2$.

$7°$. When $x > 0$, and a_1 is sufficiently small, then $x + a_1$, x, and $x - a_1$ are all positive, and, therefore, $(x - a_1)^m + (x + a_1)^m = (k(a_1))^m \cdot x^m + 2 \cdot a_1^m$. If

we move $2 \cdot a_1^m$ to the left-hand side, and divide both sides by x^m, we conclude that $\left(1 - \dfrac{a_1}{x}\right)^m + \left(1 + \dfrac{a_1}{x}\right)^m - 2 \cdot \left(\dfrac{a_1}{x}\right)^m = (k(a_1))^m$. The left-hand side of the resulting equality depends only on the ratio $z = \dfrac{a_1}{x}$, the right-hand side only on a_1. Therefore, if we choose any positive real number λ, and take $a_1' = \lambda \cdot a_1$ and $x' = \lambda \cdot x$ instead of a_1 and x, then we can conclude that the left-hand side will be still the same, and therefore, the right-hand side must be the same, i.e., $(k(a_1))^m = (k(\lambda \cdot a_1))^m$. Since λ was an arbitrary number, we conclude that $k(a_1)$ does not depend on a at all, i.e., that $(k(a_1))^m$ is a constant. Let us denote this constant by k.

So the equation takes the form $(1 - z)^m + (1 + z)^m = k + 2 \cdot z^m$. When $z \to 0$, then the left-hand side tends to 2 and right-hand side to k, so from their equality we conclude that $k = 2$, i.e., that $(1 - z)^m + (1 + z)^m = 2 + 2 \cdot z^m$.

The left-hand side is an analytical function of z for z close to 0. Therefore the right-hand side must also be a regular analytical function in the neighborhood of 0 (i.e., it must have a Taylor expansion for $z = 0$). Hence, m must be an integer.

The values $m < 2$ are impossible, because for $m = 0$ our equality turns into a false equality $2 = 3$, and for $m = 1$ it turns into an equality $1 - z + 1 + z = 2 + z$, which is true only for $z = 0$. So $m \geq 2$.

Since both sides are analytical in z, the second derivatives of both sides at $z = 0$ must be equal to each other. The second derivative of the left-hand side at $z = 0$ is equal to $m \cdot (m - 1)$. The second derivative of the right-hand side is equal to $2m \cdot (m - 1) \cdot z^{m-2}$.

If $m > 2$, then this derivative equals 0 at $z = 0$ and therefore cannot be equal to $m \cdot (m - 1)$. So $m \geq 2$, and m cannot be greater than 2. So, $m = 2$. Thus, $\ell(x) = B \cdot x^2$, and hence $\mu_0(x) = \exp(-\beta \cdot x^2)$ for some $\beta > 0$. The proposition is proven.

C Proof of Proposition 2

Let us assume that $s_k \to x$, and let us prove that in this case, the ratios

$$X_N \overset{\text{def}}{=} \frac{\displaystyle\sum_{k=0}^{N} s_k \cdot (s_{k+1} - s_k)^{-2}}{\displaystyle\sum_{k=0}^{N} (s_{k+1} - s_k)^{-2}}$$

also tend to x, i.e., that for every $\varepsilon > 0$, there exists an n for which, for all $N \geq n$, we have $|X_N - x| \leq \varepsilon$.

Since $s_k \to x$, there exists an integer n_0 such that for all $k \geq n_0$, we have $|s_k - x| \leq x + \dfrac{\varepsilon}{2}$. In particular, this means that for such k, we have $s_k \leq x + \dfrac{\varepsilon}{2}$. We can represent the numerator $\mathcal{N}$ of the ratio X_N as

$$\mathcal{N} = \mathcal{N}_0 + \sum_{k=n_0+1}^{n_0} s_k \cdot (s_{k+1} - s_k)^{-2},$$

where $\mathcal{N}_0 \overset{\text{def}}{=} \sum_{k=0}^{n_0} s_k \cdot (s_{k+1} - s_k)^{-2}$. Since $s_k \leq x + \dfrac{\varepsilon}{2}$, we conclude that

$$\mathcal{N} \leq \mathcal{N}_0 + \left(x + \frac{\varepsilon}{2}\right) \cdot \Delta,$$

where we denoted $\Delta \overset{\text{defe}}{=} \sum_{k=n_0+1}^{N} (s_{k+1} - s_k)^{-2}$. Similarly, for the denominator $\mathcal{D}$ of the ratio X_N, we get an expression $\mathcal{D} = \mathcal{D}_0 + \Delta$, where

$$\mathcal{D}_0 \overset{\text{def}}{=} \sum_{k=0}^{n_0} (s_{k+1} - s_k)^{-2}.$$

Thus,

$$X_N = \frac{\mathcal{N}}{\mathcal{D}} \leq \frac{\mathcal{N}_0 + \left(x + \dfrac{\varepsilon}{2}\right) \cdot \Delta}{\mathcal{D}_0 + \Delta}.$$

The right-hand side of this inequality can be represented as

$$\frac{\mathcal{N}_0 + \left(x + \dfrac{\varepsilon}{2}\right) \cdot \Delta}{\mathcal{D}_0 + \Delta} = x + \frac{\varepsilon}{2} + \frac{\mathcal{N}_0 - \mathcal{D}_0 \cdot \left(x + \dfrac{\varepsilon}{2}\right)}{\mathcal{D}_0 + \Delta}.$$

Here, $|s_k - x| \leq \dfrac{\varepsilon}{2}$ and $|s_{k+1} - x| \leq \dfrac{\varepsilon}{2}$ implies that

$$|s_{k+1} - s_k| \leq |s_k - x| + |s_{k+1} - x| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus, $(s_{k+1} - s_k)^{-2} \geq \varepsilon^{-2}$ and so, $\Delta \geq (N - n_0) \cdot \varepsilon^{-2}$. When $N \to \infty$, we have $\Delta \to \infty$ and thus,

$$\frac{\mathcal{N}_0 - \mathcal{D}_0 \cdot \left(x + \dfrac{\varepsilon}{2}\right)}{\mathcal{D}_0 + \Delta} \leq \frac{\varepsilon}{2}$$

for sufficiently large N. For such N, we get $X_n = \dfrac{\mathcal{N}_0}{\mathcal{D}_0} \leq x + \dfrac{\varepsilon}{2} + \dfrac{\varepsilon}{2} = x + \varepsilon$. Similarly, for sufficiently large N, we get $X_N \geq x - \varepsilon$. The proposition is proven.

D   Example: Applying Formula (6) to the Divergent Geometric Series $\sum z^i$ for $|z| \geq 1$

When $|z| > 1$, the series $\sum z^i$ diverges. Here, $s_1 = 1$, $s_2 = 1 + z$, $\ldots$, and, in general, $s_k = 1 + z + \ldots + z^k = \dfrac{z^{k+1} - 1}{z - 1}$. Thus, $s_{k+1} - s_k = \dfrac{z^{k+2} - z^{k+1}}{z - 1} = z^{k+1}$. So, the denominator $\mathcal{D}$ of the formula (6) has the form $\mathcal{D} = \sum_{k=0}^{N} z^{-2 \cdot (k+1)}$. In the limit, when $N \to \infty$, we get $\mathcal{D} \to \dfrac{z^{-2}}{1 - z^{-2}}$.

For the numerator, we similarly have

$$\mathcal{N} = \sum_{k=0}^{N} \frac{z^{k+1} - 1}{z - 1} \cdot z^{-2 \cdot (k+1)} = \frac{1}{z - 1} \cdot \left(\sum_{k=0}^{N} z^{-k+1} - \sum_{k=0}^{N} z^{-2 \cdot (k+1)} \right).$$

In the limit, when $N \to \infty$, we get $\mathcal{N} \to \dfrac{1}{z - 1} \cdot \left(\dfrac{z^{-1}}{1 - z^{-1}} - \dfrac{z^{-2}}{1 - z^{-2}} \right)$. Thus,

$$x = \lim_{N \to \infty} X_N = \lim_{N \to \infty} \frac{\mathcal{N}}{\mathcal{D}} = \frac{\dfrac{1}{z - 1} \cdot \left(\dfrac{z^{-1}}{1 - z^{-1}} - \dfrac{z^{-2}}{1 - z^{-2}} \right)}{\dfrac{z^{-2}}{1 - z^{-2}}} =$$

$$\frac{1}{z - 1} \cdot \left(\frac{z^{-1}}{1 - z^{-1}} - \frac{z^{-2}}{1 - z^{-2}} \right) \cdot \frac{1 - z^{-2}}{z^{-2}}.$$

Here,

$$\frac{1}{z - 1} = \frac{1}{\dfrac{1}{z^{-1}} - 1} = \frac{z^{-1}}{1 - z^{-1}}.$$

Therefore,

$$x = \frac{z^{-1}}{1 - z^{-1}} \cdot \left(\frac{z^{-1}}{1 - z^{-1}} - \frac{z^{-2}}{1 - z^{-2}} \right) \cdot \frac{1 - z^{-2}}{z^{-2}}.$$

Adding two fractions in parentheses, we get

$$\frac{z^{-1}}{1 - z^{-1}} - \frac{z^{-2}}{1 - z^{-2}} = \frac{z^{-1} \cdot (1 + z^{-1}) - z^{-2}}{1 - z^{-2}} = \frac{z^{-1}}{1 - z^{-2}}.$$

Thus,

$$x = \frac{z^{-1}}{1 - z^{-1}} \cdot \frac{z^{-1}}{1 - z^{-2}} \cdot \frac{1 - z^{-2}}{z^{-2}}.$$

The terms z^{-1}, z^{-1}, and z^{-2} cancel each other, as well as the terms $1 - z^{-2}$ in the numerator and in the denominator. Thus, we get $x = \dfrac{1}{1 - z^{-1}}$.

For example, for $z = 2$, we get $x = 1 + 2 + 4 + \ldots = \dfrac{1}{1 - 1/2} = 2$.