

# GROUP-THEORETIC APPROACH AS A GENERAL FRAMEWORK FOR SENSORS, NEURAL NETWORKS, FUZZY CONTROL, AND GENETIC BOOLEAN NETWORKS

Hung T. Nguyen<sup>1</sup>, Vladik Kreinovich<sup>2</sup>, Chitta Baral<sup>3</sup>, Valery D. Mazin<sup>4</sup>

<sup>1</sup>Mathematics, New Mexico State University, Las Cruces, USA

<sup>2</sup>Computer Science, University of Texas, El Paso, USA

<sup>3</sup>Computer Science, Arizona State University, Tempe, USA

<sup>4</sup>State Technical Univ. of St. Petersburg, Russia

**Abstract.** When describing a system of interacting genes, a useful approximation is provided by a Boolean network model, in which each gene is either switched on or off – i.e., its state is described by a Boolean variable.

Recent papers by I. Shmulevich et al. show that although in principle, arbitrarily complex Boolean functions are possible, in reality, the corresponding Boolean networks can be well described by Boolean functions from one of the so-called *Post classes* – classes that are closed under composition. These classes were originally described by E. Post.

It is known that the Boolean model is only an approximate description of the real-life gene interaction. In reality, the interaction may be more complex. How can we extend these results to more realistic continuous models of gene interaction?

In this paper, we show that the Post class approach can be viewed as a particular case of a general group-theoretic framework that has already led to a successful justification of empirical formulas from such areas of signal processing as sensor analysis, neural networks, fuzzy techniques, etc. Because of this relation, we suggest group-theoretic approach as a framework for describing gene interaction in a more realistic way.

**Keywords:** group-theoretic approach, general measurement methodology, fuzzy techniques

## 1. MOTIVATIONS FOR THE MAIN IDEA: WHY NORMAL DISTRIBUTION

Before we start explaining cases covered by the general group-theoretic approach, let us recall a similar case – the use of normal (Gaussian) distribution in signal processing.

How can we describe the probability distribution of a measurement error? From the purely mathematical viewpoint, there are many possible probability distributions, but in reality, for many sensors and for many more sophisticated measuring instruments, normal distribution works just fine. Normal distribution is the usual tool that engineers and scientists use when processing data; see, e.g., [9].

A thorough empirical analysis of actual sensors and measuring instruments have shown that for about half of them, the measurement error is indeed normally distributed, and for many others, the actual distribution is reasonably close to a normal one; see, e.g., [6,7]. Why?

A usual answer to this “Why?” question is that there exists a Central Limit Theorem, according to which, crudely speaking, if we have many error sources, and errors corresponding to each source and small and independent of each other, then the resulting distribution is close to Gaussian. How can we generalize this result to more general signal processing situations?

The usual proofs of the Central Limit Theorem use specific properties of the normal distribution. Therefore, if we want to generalize this result, we must first reformulate the problem in more general terms.

Suppose that we already know that the probability distribution of measurement errors  $\Delta x$  are usually described by distributions from a certain class  $\mathcal{P}$ . How can we then describe this class  $\mathcal{P}$ ? First, the actual distribution depends on the selection of a measuring unit: if we change a unit from, say, a meter to a centimeter, the error remains the same but its numerical value increase 100 times, from  $\Delta x$  to  $100 \cdot \Delta x$ . In general, it is therefore reasonable to assume that if the random variable  $\Delta x$  belongs to the desired class  $\mathcal{P}$ , then, for every positive number  $\lambda$ , the variable  $\lambda \cdot \Delta x$  must also belong to this class.

Second, it often happens that we have several independent sources of error. Thus, if  $\Delta x_1$  and  $\Delta x_2$  are independent variables that are both distributed according to distributions from  $\mathcal{P}$ , then their sum should also be distributed according to one of the distributions from  $\mathcal{P}$ .

If we restrict ourselves to 1-parametric families of distributions, then these families should have the type  $\lambda \cdot \xi_0$  for some standard distribution  $\xi_0$ , and the second property means that a linear combination of such distributions must also be distributed by a similar type distribution. Such probability distributions are called *infinitely divisible* distributions. They have all been classified, Gaussian distribution is one of them, and it is

known that under certain reasonable additional assumption, Gaussian distribution is the only one.

This more general description can already be extended to more general signal processing situations.

## 2. FIRST CASE STUDY: SENSORS

Now that we already have the main idea, let us show how this idea can be applied to other situations. We will start with the sensors.

Sensors usually transform the value of the physical quantity such as stress, temperature, or pressure, into an electric signal (that is later digitized and inputted into signal processing software). The dependence  $y = f(x)$  between the input value  $x$  and the resulting electric signal  $y$  is usually smooth; therefore, in the first approximation, we can safely approximate this dependence by a linear function  $y = a \cdot x + b$ . However, when we start looking for more accurate descriptions, we realize that the dependence of  $y$  on  $x$  is usually non-linear. How can we describe this non-linearity?

Again, in principle, arbitrary non-linearity is possible, but in reality, some non-linear dependencies are more frequent than others. So, a reasonable idea is to describe the class  $\mathcal{F}$  of frequently occurring dependencies. How can we do that?

A sophisticated sensor usually consists of several layers in each of which the signal is transformed. If  $f_1(x)$  is the transformation performed by the first layer and  $f_2(x)$  the transformation performed by the second layer, then, for every input  $x$ , the signal coming out of the second layer is equal to  $f_2(f_1(x))$ . In other words, the transformation corresponding to the two-layer sensor is equal to the composition of the transformations corresponding to individual layers.

It is therefore reasonable to require that if both layers correspond to a transformation from the class  $\mathcal{F}$ , then the transformation corresponding to the 2-layer sensor should also be in the same class  $\mathcal{F}$ , i.e., that this class  $\mathcal{F}$  of “frequently occurring” transformations should be closed under composition. It is also reasonable to require that all linear transformations belong to this class, and that – since most of the corresponding physical processes are reversible – that for every function  $f \in \mathcal{F}$ , its inverse  $f^{-1}$  should also belong to  $\mathcal{F}$ . In other words, it is reasonable to require that the class  $\mathcal{F}$  is a *group* w.r.t. composition, i.e., in other words, it is a *transformation group*.

## 3. WHAT TRANSFORMATION GROUPS ARE POSSIBLE?

What transformation groups are possible? It turns out that there is already an answer to this question, and this answer can be traced back to the work of N. Wiener. Namely, in mid-1940s, Wiener analyzed how we humans recognize an object. According to some physiological studies, there are five clearly distinct levels of recognition:

- When an object is far enough, all we see is a blur. We cannot tell its shape, we cannot tell whether it is a point object or not.
- When we get closer, we can recognize some shape, but we still have trouble telling what shape it is exactly. We may see a circle as an ellipse, a square as a rhombus (diamond).
- As the object gets closer, we can clearly distinguish parallel lines, but we may not yet tell the angles. For example, we are not sure whether what we see is a rectangle or a parallelogram.
- When we get even closer, we can see the shape, but at a large distance, our stereoscopic ability does not work, so we cannot say whether what we see is large or small. For example, we already know this is a square, but we cannot tell whether it is a nearby small square, or a far away large one.
- Finally, when we get really close, we can see both the shape and the size of the object.

In mathematical terms, at each stage (except for the last one), the uncertainty means that we can apply some transformations to the original image without changing the perceived image. So, each stage can be characterized by the group  $G$  of all transformations that are, in this sense, possible on this stage:

- At first, we have the group  $G$  of all possible transformations.
- Then, we get down to the group of all *projective* transformations (that describe *projections* from one plane to another).
- Third, we get the group of all *linear* transformations (also called *affine*).
- Fourth, we get the group generated by all of motions (i.e., translations and rotations) and dilations (transformations from this group are called *homotheties*).
- Finally, we get the group of all motions.

From this physiological observation, N. Wiener in his book [12] (first published in late 1940s), made an interesting conclusion: If there was an intermediate group between, e.g., projective and affine transformations, then in an ideal vision system, it would probably be reasonable to use it in situations intermediate between the situations in which we use these two groups. Since a man is a product of billion years of improving evolution, it is therefore reasonable to assume that whatever transformation groups are possible, they are already used by us humans. Since we only use five different groups, he thus concluded that no other transformation groups exist. To be more precise, he conjectured that the only transformation Lie groups that contain the group of all motions are: the group of all homotheties, the group of all affine transformations, and the group of all projective transformations. Mathematicians were

at first sceptical about this conjecture, but surprisingly, in mid-1960s, papers appeared that, in effect, proved Wiener's hypothesis [11].

For 1-dimensional case, projective transformations are simply fractionally linear, and affine are simply linear. So, the conclusion is that the only non-trivial transformation group that contains all linear transformations is the group of all fractional-linear transformation.

#### 4. SENSORS: CONTINUES

It turns out that fractional-linear functions indeed provide a very good description of how sensors operate. To get a really accurate description of a sensor transformation function  $y = f(x)$ , we may need to use different fractional-linear approximations in different parts of the input range, but still, for many sensors, the overall number of coefficients needed in this piece-wise fractional-linear description is drastically smaller than the overall number of coefficients needed for, e.g., piece-wise linear description [2,3].

#### 5. SECOND CASE STUDY: NEURAL NETWORKS

Let us show that this same idea can explain why the (empirically selected) sigmoid function

$$s_0(x) = \frac{1}{1 + \exp(-x)} \quad (1)$$

is successful in neural data processing.

Indeed, the main application of neural networks is to learn. In the process of learning, we may discover that the original data values have a systematic error  $b$ , so we may want to correct for this error. So, we must be able, knowing  $s(x)$  and  $b$ , easily compute  $s(x - b)$ .

Hence, we are looking for (smooth) activations functions  $s(x)$  for which, for every real number  $b$ , there exists an easy transformation  $s$  such that for every  $x$ ,  $s(x - b) = e(s(x))$ .

How do we formalize easiness? Linear transformations are easy; composition of two easy transformations should be easy, and an inverse of an easy transformation should be easy. Thus, easy transformations form a group that contains all linear ones. If we restrict ourselves to connected Lie groups, then we can conclude that all easy transformations are piece-wise linear.

It is now possible to prove (see, e.g., [4]) that if a smooth monotonic function  $s(x)$  is easily correctable for additive errors, then either  $s(x) = a + b \cdot s_0(K \cdot x + l)$  for some  $a, b, K$  and  $l$ , or  $s(x) = a + b \exp(Kx)$  for some  $a, b$ , and  $K$ . So, if we require (as it is usually done in neural networks) that the range of the function  $s$  is bounded, then we are left with only standard sigmoid.

The proof is as follows: first, we use the above-cited result to show that  $E$  consists of fractional-linear transformations. This means that for every  $a$ , there exist  $A, B, C$ , and  $D$  for which, for every  $a$ , we have

$$s(y + a) = \frac{A(a) + B(a) \cdot s(y)}{C(a) + D(a) \cdot s(y)}. \quad (2)$$

If the function  $s(y)$  is smooth, then it is possible to show that the functions  $A(a)$  through  $D(a)$  are also smooth (see Appendix A). Differentiating both sides of the above equation w.r.t.  $a$  and substituting  $a = 0$ , we arrive at a differential equation of the type  $ds/dy = f(s)$  for some explicit function  $f$ . This equation can be rewritten as  $ds/f(s) = dy$  and integrated explicitly.

#### 6. THIRD CASE STUDY: FUZZY CONTROL

The main idea behind fuzzy control is that in many real-life situations, expert controllers can also describe their experience by using words from natural language, like "if the velocity  $v$  is a little bit too high, hit the brakes for a short time  $t$ ". If we want to automate this control, we must describe these words in precise terms. In fuzzy logic approach, we describe words like "small" by describing, for each value  $x$ , the degree to which an expert believes that this particular  $x$  is small. There are many different ways of specifying this idea into an exact methodology.

- First, in reality, we cannot solicit the expert's degree of belief for all possible values of  $x$ , we have to ask for some values of  $x$  and then interpolate; there are many possible interpolation procedures.
- Another choice comes when we combine different pieces of knowledge, e.g., when we have a rule that says that if  $A$  and  $B$  are true, then we should apply a certain level of control. If we know the degrees of belief in  $A$  and in  $B$ , and we have no information about the possible relation between  $A$  and  $B$ , what degrees should we assign to  $A \& B$ ? There are many possible "and"-operations that intend to solve this problem (they are also called t-norms), and there are many possible "or"-operations (t-conorms).
- Finally, at the end, we describe, for each possible value of control  $u$ , the degree to which this control is reasonable; we must translate this data into a single value that our designed automatic controller will apply. In other words, we must move from the "fuzzy" conclusion to the exact one, "defuzzify" the conclusion. There are many techniques for such defuzzification.

There has been a lot of empirical studies showing, for different control criteria, which versions of the fuzzy control methodology lead to the best results. It turns out [4,5] that most of these choices can be explained by the same group-theoretic idea.

Indeed, there exist different procedures for assigning numeric values that describe uncertainty of the experts' statements. The same expert's degree of uncertainty that he expresses, for example, by the expression "for sure", can lead to 0.9 if we apply one procedure, and to 0.8 if another procedure is used. Just like 1 foot and 12 inches describe the same length, but in different scales, we can say that 0.9 and 0.8 represent the same degree of certainty in two different scales.

From a mathematical viewpoint, one can use any scale, but from the practical viewpoint some of them will be more reasonable to use, and some of them less reasonable. It is reasonable to assume that a composition of two reasonable transformations is reasonable, that the inverse is reasonable, etc. – i.e., that the set of all reasonable transformations forms a group containing all linear ones. Thus, every reasonable transformation is fractionally linear.

This conclusion enables us to justify all the optimal choices in fuzzy control. For example, for each statement  $B$ , we can consider the degree of belief in  $A \& B$  as an indication of degree of belief in  $A$ ; thus, the transformation from  $t(A)$  to  $t(A \& B)$  must be fractionally-linear. This indeed leads to a general family of t-norms that includes all empirically best ones.

## 7. POSSIBLE APPLICATION TO GENETIC NETWORKS

When describing a system of interacting genes, a useful approximation is provided by a Boolean network model, in which each gene is either switched on or off – i.e., its state is described by a Boolean variable; see, e.g., [1].

Recent papers by I. Shmulevich et al. (see, e.g., [10]) show that although in principle, arbitrarily complex Boolean functions are possible, in reality, the corresponding Boolean networks can be well described by Boolean functions from one of the so-called *Post classes* – classes that are closed under composition. These classes were originally described by E. Post.

It is known that the Boolean model is only an approximate description of the real-life gene interaction. In reality, the interaction may be more complex. How can we extend these results to more realistic continuous models of gene interaction?

As we have seen, in many areas of non-linear data processing, in particular, in the areas related to intelligent data processing, it is very fruitful to consider classes of transformations that are closed under composition. We have also seen that such classes have been successful in describing genetic Boolean networks. It is therefore reasonable to expect that this approach may lead to successful extension of genetic Boolean networks results to more realistic models of gene interaction.

## ACKNOWLEDGMENTS

This work was supported in part by NASA under cooperative agreement NCC5-209, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-1-0365, by NSF grants EAR-0112968, EAR-0225670, and EIA-0321328, by the Army Research Laboratories grant DATM-05-02-C-0046, by the Hewlett-Packard equipment grants

89955.1 and 89955.2, and by the Personal Interface AccessGrid awarded by the Education, Outreach and Training Partnership for Advanced Computational Infrastructure EOT-PACI. This work was partly performed during V.K.'s visit to Brazil. This visit was sponsored by the Brazilian funding agency CT-INFO/CNPq.

The authors are thankful to the anonymous referees for valuable suggestions.

## REFERENCES

- [1] S.A. Kaufmann, *The origins of Order: Self-Organization and Selection in Evolution*, Oxford Univ. Press, 1993.
- [2] V.G. Knorrning, Y.Ya. Kreinovich, V.D. Mazin, "Measurement Information: Scales and Conversions", *Measurement Techniques*, **45**(2), 113–115, 2002.
- [3] I.N. Krotkov, V. Kreinovich, V.D. Mazin, "A general formula for the measurement transformations, allowing the numerical methods of analyzing the measuring and computational systems", *Measurement Techniques*, No. 10, 8–10, 1987.
- [4] H.T. Nguyen, V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.
- [5] H.T. Nguyen, V. Kreinovich, "Methodology of fuzzy control: an introduction", In: H. T. Nguyen and M. Sugeno (eds.), *Fuzzy Systems: Modeling and Control*, Kluwer, Boston, MA, 1998, pp. 19–62.
- [6] P.V. Novitskii, I.A. Zograph, *Estimating the Measurement Errors*, Leningrad: Energoatomizdat, 1991 (in Russian).
- [7] A.I. Orlov, "How often are the observations normal?", *Industrial Laboratory*, **57**(7), 770–772, 1991.
- [8] E.L. Post, "The two-valued iterative systems of mathematical logic", *Annals of Mathematical Studies*, **5**, 1–122, 1941.
- [9] S. Rabinovich, *Measurement Errors: Theory and Practice*, American Institute of Physics, N.Y., 1993.
- [10] I. Shmulevich, H. Lähdesmäki, E.R. Dougherty, J. Astola, W. Zhang, "The role of certain Post classes in Boolean network models of genetic networks," *Proc. Nat. Acad. Sci. USA*, **100**(19), 10734–10739, 2003.
- [11] I.M. Singer, S. Sternberg, "Infinite groups of Lie and Cartan, Part 1", *Journal d'Analyse Mathématique*, **15**, 1–113, 1965.
- [12] N. Wiener, *Cybernetics, or Control and Communication in the animal and the machine*, MIT Press, Cambridge MA, 1962.

## APPENDIX A: PROOF FOR NEURAL NETWORKS

Let us first prove that the functions  $A(a)$  through  $D(a)$  are indeed smooth. First of all, we can simplify the expression (2) if we divide both the numerator and the denominator of the fraction (2) by  $C(a)$ . Then, the equation (2) takes a simplified form

$$s(y + a) = \frac{A_1(a) + B_1(a) \cdot s(y)}{1 + D_1(a) \cdot s(y)}, \quad (3)$$

where  $A_1(a) \stackrel{\text{def}}{=} A(a)/C(a)$ ,  $B_1(a) \stackrel{\text{def}}{=} B(a)/C(a)$ , and  $D_1(a) \stackrel{\text{def}}{=} D(a)/C(a)$ . Multiplying both sides of the new equation (3) by the denominator of its right-hand side, we conclude that

$$s(y+a) + D_1(a) \cdot s(y) \cdot s(y+a) = A_1(a) + B_1(a) \cdot s(y).$$

For every  $a$ , we can write this equation for three different values  $y = y_1, y_2, y_3$ , and get a system of three linear equations with three unknowns  $A_1(a)$ ,  $B_1(a)$ , and  $D_1(a)$ . According to Kramer's rule, the solution of this equation can be expressed as a rational – hence smooth – function of the coefficients. Thus, the values  $A_1(a)$ ,  $B_1(a)$ , and  $D_1(a)$  smoothly depend on the coefficients  $s(y_i)$  and  $s(y_i) \cdot s(y_i + a)$ . Since the function  $s(y)$  is smooth, this dependence on  $a$  is smooth, so all the coefficients of the piece-wise linear expression (2) indeed depend smoothly on  $a$ .

Let us now differentiate both sides w.r.t.  $a$ . As a result, we get

$$s'(y+a) = \frac{A_1'(a) + B_1'(a) \cdot s(y)}{1 + D_1(a) \cdot s(y)} - \frac{(A_1(a) + B_1(a) \cdot s(y)) \cdot D_1'(a) \cdot s(y)}{(1 + D_1(a) \cdot s(y))^2}. \quad (4)$$

For  $a = 0$ , the transformation (3) takes the form  $s(y+a) = s(y)$ , hence  $A_1(0) = D_1(0) = 0$  and  $B_1(0) = 1$ . Therefore, for  $a = 0$ , the above equation takes the form

$$s'(y) = a + b \cdot s(y) + c \cdot s^2(y), \quad (5)$$

where  $a \stackrel{\text{def}}{=} A_1'(0)$ ,  $b \stackrel{\text{def}}{=} B_1'(0)$ , and  $c \stackrel{\text{def}}{=} -D_1'(0)$ , or, alternatively,

$$\frac{ds}{dy} = a + b \cdot s + c \cdot s^2. \quad (6)$$

We can separate the variables of this equation if we divide both sides by the right-hand side and multiply both sides by  $dy$ ; as a result, we get

$$\frac{ds}{a + b \cdot s + c \cdot s^2} = dy. \quad (7)$$

We can now integrate both sides of this equation.

If  $c = 0$  and  $b = 0$ , we get  $s/a = y + C$ , so we get a linear function  $s(y)$ .

If  $c = 0$  and  $b \neq 0$ , we get  $d(a + b \cdot s) = b \cdot ds$ , hence

$$\frac{d(a + b \cdot s)}{a + b \cdot s} = b \cdot dy. \quad (8)$$

Integrating, we get  $\ln(a + b \cdot s) = b \cdot y + C$ , hence  $a + b \cdot y = \exp(C) \cdot \exp(b \cdot y)$ , i.e., we get an exponential solution  $s(y)$ .

If  $c \neq 0$ , the integration is somewhat more complex, but still, the integral in the left-hand side is a textbook integral, so we easily describe the generic solution: either  $s(y)$  is fractionally linear, or  $s(y) =$

$(a + b \cdot \tan(ky))/(c + d \cdot \tan(ky))$  for some  $a, b, c, d$ , or  $s(y) = (a + b \cdot \tanh(ky))/(c + d \cdot \tanh(ky))$ . In the first two cases, we do not get monotonicity, and the third function is equivalent to a sigmoid. The statement is proven.

## APPENDIX B: GENERAL PROOF

Let us show that for 2-D transformations, the only finite-parametric smooth family of smooth transformations that contains all linear transformations and that forms a transformation group is the group of all fractional-linear transformations.

### B.1. Infinitesimal transformations and Lie algebras

To prove this result, it is reasonable to consider *infinitesimal* transformations, i.e., functions  $f(x)$  such that for small  $\varepsilon$ , the transformation  $x \rightarrow x + \varepsilon \cdot f(x) + O(\varepsilon^2)$  is a possible transformation. The class of all such functions is called a *Lie algebra* of the original transformation group.

### B.2. Lie algebras are closed under multiplication by a constant

Lie algebras have several useful properties. First, if the function  $f(x)$  belongs to the Lie algebra, then for every constant  $\lambda$ , the function  $\lambda \cdot f(x)$  also belongs to the Lie algebra. Indeed, since the transformation  $x \rightarrow x + \varepsilon \cdot f(x) + O(\varepsilon^2)$  is possible for all  $\varepsilon$ , it is also possible for  $\varepsilon' = \varepsilon \cdot \lambda$ . In other words, the transformation  $x \rightarrow x + \varepsilon \cdot \lambda \cdot f(x) + O(\varepsilon^2)$  is possible. This, by definition of the Lie algebra, means that  $\lambda \cdot f(x)$  also belongs to the Lie algebra.

### B.3. Lie algebras are closed under addition

Let us now prove that the algebra  $A$  is closed under addition, i.e., if  $f(x) \in A$  and  $g(x) \in A$ , then the sum  $f(x) + g(x)$  of these two functions also belongs to  $A$ . Together with the previous property, this would mean that  $A$  is a linear space. Indeed, if two functions  $f(x)$  and  $g(x)$  belong to the Lie algebra, this means that for small  $\varepsilon$ , transformations

$$x \rightarrow y = x + \varepsilon \cdot f(x) + O(\varepsilon^2) \quad (9)$$

and

$$y \rightarrow z = y + \varepsilon \cdot g(y) + O(\varepsilon^2) \quad (10)$$

are both possible. Therefore, the composition of these transformations is also possible. This composition takes the form  $x \rightarrow z$ , i.e.,

$$x \rightarrow (x + \varepsilon \cdot f(x) + O(\varepsilon^2)) + \varepsilon \cdot g(x + \varepsilon \cdot f(x) + O(\varepsilon^2)) + O(\varepsilon^2). \quad (11)$$

If we consider only terms that are linear in  $\varepsilon$ , then we conclude that this composition transformation has the form  $x \rightarrow x + \varepsilon \cdot (f(x) + g(x)) + O(\varepsilon^2)$ , i.e., that the sum  $f(x) + g(x)$  indeed belongs to the Lie algebra.

So,  $A$  is a linear space.

The inverse transformation takes the form

$$x - \varepsilon \cdot f(x) + O(\varepsilon^2),$$

thus we conclude that  $-f(x)$  also belongs to  $A$ .

#### B.4. Lie algebras are closed under Lie product

Let us now consider composition in more detail. After applying  $f$ , we get  $y = x + \varepsilon_1 \cdot f(x) + O(\varepsilon_1)$ . After applying  $g$  to the result  $y$  of the first transformation, we get

$$y + \varepsilon_2 \cdot g(y) + O(\varepsilon_2^2) = x + \varepsilon_1 \cdot f(x) + O(\varepsilon_1^2) + \varepsilon_2 \cdot g(x + \varepsilon_1 \cdot f(x) + O(\varepsilon_1^2)) + O(\varepsilon_2^2).$$

Since the function  $g(x)$  is smooth, we conclude that

$$g(x + \varepsilon_1 \cdot x + O(\varepsilon_1^2)) = g(x) + g'(x) \cdot \varepsilon_1 \cdot f(x) + O(\varepsilon_1^2), \quad (12)$$

hence the result of the two transformations takes the form

$$x + \varepsilon_1 \cdot f(x) + \varepsilon_2 \cdot g(x) + \varepsilon_1 \cdot \varepsilon_2 \cdot f'(x) \cdot g(x) + O(\varepsilon_i^2).$$

Similarly, when we first apply  $g$  and then  $f$ , we get a transformation

$$x + \varepsilon_1 \cdot f(x) + \varepsilon_2 \cdot g(x) + \varepsilon_1 \cdot \varepsilon_2 \cdot f(x) \cdot g'(x) + O(\varepsilon_i^2).$$

If we first apply  $f$  then  $g$ , then inverse to  $f$  and inverse to  $g$ , we thus get a possible transformation

$$x \rightarrow x + \varepsilon_1 \cdot \varepsilon_2 \cdot (f'(x) \cdot g(x) - f(x) \cdot g'(x)) + O(\varepsilon_i^2),$$

so we can conclude that for every two functions  $f(x) \in A$  and  $g(x) \in A$ , the function  $f'(x) \cdot g(x) - f(x) \cdot g'(x)$  also belongs to  $A$ . This function is called a *Lie product* of the functions  $f(x)$  and  $g(x)$  and denoted by  $[f, g]$ .

#### B.5. Our Lie algebra $A$ is closed under the derivative operation

We assume that all linear transformations, including shifts  $x \rightarrow x + h$ , are possible. So, if  $f(x) \in A$ , then, combining an infinitesimal transformation related to  $f(x)$  with shift, we conclude that  $f(x + h) \in A$ . Since  $A$  is a linear space, we conclude that

$$\frac{f(x + h) - f(x)}{h} \in A. \quad (13)$$

As  $h \rightarrow 0$ , this expression tends to  $f'(x)$ . Since the algebra  $A$  is finite-dimensional, it contains the limits of its elements, hence  $f'(x) \in A$ .

#### B.6. The algebra $A$ contains, with each function $f(x)$ , all monomials from its Taylor expansion

Combining infinitesimal transformation  $f(x)$  with scalings  $x \rightarrow \lambda x$ , we conclude that  $f(\lambda \cdot x) \in A$ . Let us expand  $f(x) \in A$  into Taylor series and ignore the first zero terms, then

$$f(x) = a_k \cdot x^k + a_{k+1} \cdot x^{k+1} + \dots \quad (14)$$

Since  $A$  is a vector space, we have

$$\lambda^{-k} \cdot f(\lambda \cdot x) = a_k \cdot x^k + \lambda \cdot a_{k+1} \cdot x^{k+1} + \dots \in A. \quad (15)$$

In the limit  $\lambda \rightarrow 0$ , we conclude that  $a_k \cdot x^k \in A$  – hence that  $x^k \in A$ . From  $f(x) \in A$  and  $a_k \cdot x^k \in A$ , we conclude that

$$f(x) - a_k \cdot x^k = a_{k+1} \cdot x^{k+1} + \dots \in A, \quad (16)$$

so the first non-zero term  $x^l$ ,  $l > k$ , in the expansion of  $f(x)$ , also belongs to  $A$ . By induction, we can prove that all monomials  $x^k$  that are present in the expansion of  $f(x)$  also belong to  $A$ .

#### B.7. The algebra $A$ contains, with each monomial $x^k$ , all smaller powers of $x$

Once  $x^k \in A$ , we have  $(x^k)' = k \cdot x^{k-1} \in A$ , hence  $x^{k-1} \in A$ . By induction, we can conclude that all smaller powers of  $x$  also belong to  $A$ .

#### B.8. The algebra $A$ cannot contain monomials $x^k$ with $k > 2$

From the previous result, we conclude that if  $x^k \in A$  for some  $k \geq 3$ , we have  $x^3 \in A$  and  $x^2 \in A$ . From this, we conclude that  $[x^3, x^2] \sim x^4 \in A$ , hence  $[x^4, x^2] \sim x^5 \in A, \dots$ , and eventually, that  $x^k \in A$  for all  $k$ . So, in this case,  $A$  have infinitely many linearly independent functions  $x^k$ , which contradicts to our assumption that  $A$  is a finite-dimensional space.

This contradiction shows that Taylor expansions of functions  $f(x) \in A$  can only contain 1,  $x$ , and  $x^2$  – i.e., all functions  $f(x)$  are quadratic.

#### B.9. Final result: all possible transformations are fractionally linear

Since every function  $f(x)$  is quadratic, the corresponding infinitesimal transformations have the following form:

$$x \rightarrow x + \varepsilon \cdot a_0 + \varepsilon \cdot a_1 \cdot x + \varepsilon \cdot a_2 \cdot x^2 + O(\varepsilon^2) \quad (17)$$

It is easy to see that this transformation can be represented as fractionally linear transformations:

$$x \rightarrow \frac{(1 + \varepsilon \cdot a_1) \cdot x + \varepsilon \cdot a_0}{1 - \varepsilon \cdot a_2 \cdot x} + O(\varepsilon^2). \quad (18)$$

An arbitrary possible transformation can be obtained as a composition of infinitesimal transformations; a composition of fractionally linear transformations is fractionally linear, so all possible transformations are fractionally linear.

The statement is proven.

---

**Authors:** Hung T. Nguyen, Department of Mathematical Sciences, New Mexico State University, Las Cruces, NM 88008, USA, email hunguyen@nmsu.edu.

Corresponding author: Vladik Kreinovich, Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA, phone 1-915-747-6951, fax 1-915-747-5030, email vladik@cs.utep.edu.

Chitta Baral, Department of Computer Science and Engineering, Arizona State University, Tempe, AZ 85287-5406, USA, email chitta@asu.edu.

Valery D. Mazin, State Technical Univ. of St. Petersburg, Information-Measuring Department, Politechnicheskaya 29, 195251 St. Petersburg, Russia.