International Journal of Theoretical Physics, Vol. 38, No. 6, 1999

# Why Fundamental Physical Equations Are of Second Order

**Takeshi Yamakawa**[1] **and Vladik Kreinovich**[2]

*Received January 25, 1999*

We use a deep mathematical result (namely, a minor modification of Kolmogorov's solution to Hilbert's 13th problem) to explain why fundamental physical equations are of second order. This same result explain why all these fundamental equations naturally lead to nonsmooth solutions like singularities.

## 1. FORMULATION OF THE PROBLEM

Most physical phenomena are described in terms of partial differential equations. These differential equations can be arbitrarily complicated. In particular, they can be of high order; e.g., the equations of elasticity theory (see, e.g., ref. 8) are of fourth order (i.e., involve derivatives of fourth order). However, amazingly, these higher order equations only occur in the description of *nonfundamental* phenomena, i.e., phenomena which (like elasticity) can be reduced to more fundamental forces and fields, while *fundamental* physical equations, i.e., equations which describe the evolution of fundamental fields and forces, are of (at most) second order. Newton's equations are of second order, and so are Maxwell's equations, which describe electrodynamics, Einstein's equations, which describe general relativity, Schrödinger's and Dirac's equations, which describe quantum physics, etc. Why?

[1] Control Engineering and Science, Kyushu Institute of Technology, 680-4, Kawazu, Iizuka, Fukuoka 820-8502, Japan; e-mail yamakawa@ces.kyutech.ac.jp.
[2] Department of Computer Science, University of Texas at El Paso, El Paso, Texas 79968; e-mail vladik@cs.utep.edu.



0020-7748/99/0600-1763$16.00/0 

## 2. TO ANSWER THIS QUESTION, WE REFORMULATE IT IN PHYSICAL TERMS

What do mathematical terms “first order,” “second order,” etc., mean physically?

The fact that a system is described by a differential equation of *first* order means that the state $s(t)$ of this system at a given moment $t$ uniquely determines the rate $\dot{s}(t)$ with which the state changes, and thus uniquely determines the state $s(t + \Delta t)$ of the system in the “next” moment of time $t + \Delta t$. In other words, the state $s(t + \Delta t)$ is a function of a state at the previous moment of time: $s(t + \Delta t) = f(s(t))$. Therefore, to describe the evolution of a system which is described by first-order differential equations, it is sufficient to have a function of one variable which describes how the state changes.

If a system is described by differential equations of *second* order, then it is not enough to know the initial state $s(t)$ to predict the evolution of a system [i.e., to predict the next state $s(t + \Delta t)$]; in addition to the state $s(t)$, we must also know the previous value of the *rate* $\dot{s}(t)$ with which the state changed. This rate is, from a strict *mathematical* viewpoint, a *limit* of the ratio $[s(t) - s(t - \Delta t)]/\Delta t$ when $\Delta t \to 0$. From the *physical* (*practical*) viewpoint, this “limit” means, crudely speaking, that the rate can be defined (within an arbitrary accuracy) as the ratio $[s(t) - s(t - \Delta t)]/\Delta t$ for a sufficiently small $\Delta t$. Therefore, for systems which are described by second-order differential equations, to predict $s(t + \Delta t)$, we must know $s(t)$ *and* the ratio $[s(t) - s(t - \Delta t)]/\Delta t$. Knowing $s(t)$ and the ratio is equivalent to knowing $s(t)$ and $s(t - \Delta t)$. In other words, for such systems, to predict the state of the system in the next moment of time, we must know the state of this system in *two* previous moments of time: $s(t + \Delta t) = f(s(t), s(t - \Delta t))$. Therefore, to describe the evolution of a system which is described by second-order differential equations, it is sufficient to have a function of two variables which describes how the state changes.

Similarly, to describe the evolution of a system which is described by *third*-order differential equations, it is sufficient to have a function of *three* variables which describes how the state changes, and, in general, to describe the evolution of a system which is described by $k$th-order differential equations, it is sufficient to have a function of $k$ variables which describes how the state changes:

$$s(t + \Delta t) = f(s(t), s(t - \Delta t), \ldots, s(t - (k - 1) \cdot \Delta t))$$

Now we are ready to reformulate the above physical phenomenon in precise mathematical terms. The above phenomenon is as follows: *Every time we have a process which is described by kth-order differential equations,* *with* $k \geq 3$, *this process is not fundamental, i.e., it can be decomposed into several more elementary processes each of which is described by equations of first or second order.* We have shown that "a process is described by $k$th-order differential equations" means that its evolution is described by a function of $k$ state variables. Therefore, the above phenomenon can be reformulated as follows: *Every time we have a physical process whose evolution is described by a function of three or more variables, this process is not fundamental, i.e., it can be decomposed into more elementary processes the evolution of each of which is described by a function of one or two variables.*

*Explanation.* We will explain this phenomenon by proving that it is actually a general feature of functions of three or more variables. Namely, we will prove the following result:

*Definition.* Let $m$ be a positive integer.

- By a state space, we mean a set $S = R^m$ of all $m$-tuples $s = (s_1, \ldots, s_m)$.
- By an area $A$ in a state space, we mean a box $A = [a_1, b_1] \times \ldots \times [a_m, b_m]$, i.e., a set of all states $s = (s_1, \ldots, s_m)$ for which $a_1 \leq s_1 \leq b_1, \ldots, a_m \leq s_m \leq b_m$.
- By a state function of $k$ variables, we mean a function $f: A^k \to S$, i.e., a function which transforms every $k$-tuple of states $(s^{(1)}, \ldots, s^{(k)})$ (each of which belongs to an area $A$) into a new states $s = f(s^{(1)}, \ldots, s^{(k)})$.

*Theorem.* Every continuous state function of three or more variables can be represented as a composition of continuous state functions of one or two variables.

*Proof.* For $m = 1$, this result was proven by Kolmogorov[5] as a solution to the conjecture of Hilbert, formulated as the 13th of the 22 problems that Hilbert proposed in 1900 as a challenge to 20th century mathematics.[4]

This problem can be traced to the Babylonians, who found (see, e.g., ref. 1) that the solutions $x$ of the quadratic equation $ax^2 + bx + c = 0$ (viewed as a function of three variables $a$, $b$, and $c$) can be represented as superpositions of functions of one and two variables, namely, arithmetic operations and square roots. Much later, similar results were obtained for functions of five variables $a$, $b$, $c$, $d$, $e$ that represent the solution of the quartic equation $ax^4 + bx^3 + cx^2 + dx + e = 0$. But Galois proved in 1830 that for higher order equations, we cannot have such a representation. This negative result led Hilbert to conjecture that not all functions of several variables can be represented by functions of two or fewer variables. Hilbert's conjecture was refuted by Kolmogorov (see, e.g., ref. 11, Chapter 11) and his student V. Arnold.

It is worth mentioning that Kolmogorov's result is not only of theoretical value: it has been used to speed up actual computations (see, e.g., refs. 3, 2, 6, 7, 13, and 12).

Based on the case $m = 1$, we can now prove the theorem for all $m$, by using the following argument (its idea is similar to ref. 14). Suppose that we have a state function $s = f(s^{(1)}, \ldots, s^{(k)})$ of $k$ state variables $s^{(1)} = (s_1^{(1)}, \ldots, s_m^{(1)}), \ldots, s^{(k)} = (s_1^{(k)}, \ldots, s_m^{(k)})$. For each input $(s^{(1)}, \ldots, s^{(k)})$, the value $s = f(s^{(1)}, \ldots, s^{(k)})$ of this function is a state $f(s^{(1)}, \ldots, s^{(k)}) = (f_1(s^{(1)}, \ldots, s^{(k)}), \ldots, f_m(s^{(1)}, \ldots, s^{(k)}))$, where by $f_i(s^{(1)}, \ldots, s^{(k)})$ we denote the $i$th component of the state $s = f(s^{(1)}, \ldots, s^{(k)})$. Therefore, each state-valued function $f\colon A^k \to S = R^m$ can be represented as $m$ real-valued functions $f_i\colon A^k \to R$, $1 \le i \le m$.

Each of these functions $f_i\colon A^k \to R$ maps $k$ states (i.e., $k \times m$ components) into a real number. Therefore, each of these functions can be represented as a real-valued function of $k \times m$ *real* variables $s_1^{(1)}, \ldots, s_m^{(1)}, \ldots, s_1^{(k)}, \ldots, s_m^{(k)}$. Each of these $m$ functions $f_i$ can be represented (due to Kolmogorov's theorem) as a composition of functions of one and two variables. So, to represent the original state function of $k$ variables as a composition of state functions of one or two variables, we can do the following:

- First, we apply, to each input state $s^{(j)} = (s_1^{(j)}, \ldots, s_m^{(j)})$, $m$ functions $\pi_1(s), \ldots, \pi_m(s)$ of one state variable which transform a state $s = (s_1, \ldots, s_m)$ into corresponding "degenerate" states $\pi_1(s) = (s_1, \ldots, s_1), \ldots, \pi_i(s) = (s_i, \ldots, s_i), \ldots, \pi_m(s) = (s_m, \ldots, s_m)$. When we apply these $m$ functions to $k$ input states, we get $m \times k$ degenerate states $\pi_i(s^{(j)}) = (s_i^{(j)}, \ldots, s_i^{(j)})$, for all $i$ from 1 to $m$ and for all $j$ from 1 to $k$.
- Next, we follow the operations from Kolmogorov's theorem with these degenerate states, and get the "degenerate"-valued functions $F_1(s^{(1)}, \ldots, s^{(k)}) = (f_1(s^{(1)}, \ldots, s^{(k)}), \ldots, f_1(s^{(1)}, \ldots, s^{(k)})), \ldots, F_m(s^{(1)}, \ldots, s^{(k)}) = (f_m(s^{(1)}, \ldots, s^{(k)}), \ldots, f_m(s^{(1)}, \ldots, s^{(k)}))$, as the desired compositions of state functions of one or two variables.
- Finally, we use *combination state functions* $C_2(s, s'), \ldots, C_m(s, s')$ to combine the functions $F_1, \ldots, F_m$ into a single state function $f$. Namely, these functions work as follows:

$$C_2((s_1, \ldots), (s_1', s_2', \ldots, s_m')) = (s_1, s_2', \ldots, s_m')$$

$$\ldots$$

$$C_j((s_1, \ldots, s_{j-1}, s_j, \ldots), (s_1', \ldots, s_{j-1}', s_j', \ldots)) =$$

$$(s_1, \ldots, s_{j-1}, s_j', \ldots, s_m')$$

$$\ldots$$

$$C_m((s_1, \ldots, s_{m-1}, s_m), (s'_1, \ldots, s'_{m-1}, s'_m)) = (s_1, \ldots, s_{m-1}, s'_m)$$

We apply these combination functions to the values produced by the functions $F_1, \ldots, F_m$, to get the results $I_2 = C_2(F_1, F_2)$, $I_3 = C_3(F_2, I_2), \ldots$, $I_j = C_j(F_{j-1}, I_j), \ldots$. As a result, we get

$$I_2 = C_2(F_1, F_2) =$$
$$(f_1(s^{(1)}, \ldots, s^{(k)}), f_2(s^{(1)}, \ldots, s^{(k)}), \ldots, f_2(s^{(1)}, \ldots, s^{(k)}))$$
$$I_3 = C_3(I_2, F_3) =$$
$$(f_1(s^{(1)}, \ldots, s^{(k)}), f_2(s^{(1)}, \ldots, s^{(k)}), f_3(s^{(1)}, \ldots, s^{(k)}), \ldots, f_3(s^{(1)}, \ldots, s^{(k)}))$$
$$\ldots$$
$$I_j = C_j(I_{j-1}, F_j) =$$
$$(f_1(s^{(1)}, \ldots, s^{(k)}), \ldots, f_j(s^{(1)}, \ldots, s^{(k)}), \ldots, f_j(s^{(1)}, \ldots, s^{(k)}))$$

and finally,

$$\begin{aligned} I_m &= C_m(I_{m-1}, F_m) \\ &= (f_1(s^{(1)}, \ldots, s^{(k)}), \ldots, f_m(s^{(1)}, \ldots, s^{(k)})) \\ &= f(s^{(1)}, \ldots, s^{(k)}) \end{aligned}$$

Thus, the function $f(s^{(1)}), \ldots, s^{(k)})$ has been represented as a composition of state functions of one or two variables. The theorem is proven.

## 3. AN INTERESTING SIDE RESULT: NONSMOOTHNESS OF FUNDAMENTAL PHENOMENA

In the above explanation, we used the result that every continuous function of several varuables can be represented as a composition of functions of one or two variables. A natural next question is: If the function of several variables has a certain property (e.g., it is smooth), can we represent it as a composition of functions of one or two variables which have the same property (i.e., are also smooth)? It turns out (see, e.g., refs. 11 and 15) that for a *smooth* function, the answer to this question is no: there exist smooth functions which cannot be represented as a composition of smooth functions of fewer variables.

In other words, if we represent functions of many variables (corresponding to nonfundamental phenomena) as a composition of functions of one or two variables (which correspond to fundamental processes), then in some cases, these functions of one or two variables which correspond to fundamental processes cannot be everywhere smooth. This result provides a *general mathematical explanation of why in different areas of fundamental physics nonsmoothness appears*: infinite proper energy and other divergencies in electrodynamics, singularities in general relativity (see, e.g., ref. 9), "quantum jumps" in quantum mechanics (or, to be more precise, in quantum measurement; see, e.g., ref. 10), etc.

## ACKNOWLEDGMENTS

This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF Grant No. DUE-9750858, by the United Space Alliance, Grant No. NAS 9-20000 (PWO C0C67713A6), 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 No. F49620-95-1-0518, by the National Security Agency under Grant No. MDA904-98-1-0564, and by the Hong Kong Grant RGC4138/97E and UGC Direct Grant 2050148. One of the authors (V.K.) is thankful to Peter Sarnak (Stanford) for proposing this problem.

Part of this work was conducted while one of the authors (V.K.) was visiting the Department of Mechanical and Automation Engineering at the Chinese University of Hong Kong under the support of Grant RGC4138/97E.

## REFERENCES

1. C. B. Boyer and U. C. Merzbach, *A History of Mathematics*, Wiley, New York (1991).
2. H. L. Frisch, C. Borzi, G. Ord, J. K. Percus, and G. O. Williams, Approximate representation of functions of several variables in terms of functions of one variable, *Phys. Rev. Lett.* (1989), **63**, 927–929.
3. R. Hecht-Nielsen, Kolmogorov's mapping neural network existence theorem, *IEEE Int. Conf. Neural Networks* (1987), **2**, 11–14.
4. D. Hilbert, Mathematical Problems, Lecture delivered before the International Congress of Mathematics in Paris in 1900, *Bull. Am. Math. Soc.* (1902), **8**, 437–479.
5. A. N. Kolmogorov, On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition, *Dokl. Akad. Nauk SSSR* (1957), **114**, 369–373.
6. V. Kurkova, Kolmogorov's theorem is relevant, *Neural Computation* (1991), **3**, 617–622.
7. V. Kurkova, Kolmogorov's theorem and multilayer neural networks, *Neural Networks* (1991), **5**, 501–506.
8. L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, Butterworth Heinemann, London (1995).
9. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, Butterworth Heinemann, London (1997).
10. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, Butterworth Heinemann, London (1997).

11. G. G. Lorentz, *Approximation of Functions*, Holt, Reinhart, and Winston, New York (1966).
12. M. Nakamura, R. Mines, and V. Kreinovich, Guaranteed intervals for Kolmogorov's theorem (and their possible relation to neural networks), *Interval Computations* (1993), No. 3, 183–199.
13. M. Ness, Approximative versions of Kolmogorov's superposition theorem, proved constructively, *J. Comput. Appl. Math.* (1993).
14. V. M. Nesterov, Interval analogues of Hilbert's 13th problem, In *Abstracts of the International Conference Interval'94*, St. Petersburg, Russia, March 7–10 (1994), pp. 185–186.
15. A. G. Vitushkin, *Theory of the Transmission and Processing of Information*, Pergamon Press, London (1961).