12-1-2022

Why in MOND -- Alternative Gravitation Theory – a Specific Formula Works the Best: Complexity-Based Explanation

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

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

Comments:
Technical Report: UTEP-CS-22-121

Why in MOND – Alternative Gravitation Theory – a Specific Formula Works the Best: Complexity-Based Explanation

Olga Kosheleva and Vladik Kreinovich

Abstract Based on the rotation of the stars around a galaxy center, one can estimate the corresponding gravitational acceleration – which turns out to be much larger than what Newton's theory predicts based on the masses of all visible objects. The majority of physicists believe that this discrepancy indicates the presence of "dark" matter, but this idea has some unsolved problems. An alternative idea – known as Modified Newtonian Dynamics (MOND, for short) is that for galaxy-size distances, Newton's gravitation theory needs to be modified. One of the most effective versions of this idea uses so-called *simple interpolating function*. In this paper, we provide a possible explanation for this version's effectiveness. This explanation is based on the physicists' belief that out of all possible theories consistent with observations, the true theory is the simplest one. In line with this idea, we prove that among all the modifications which explain both the usual Newton's theory for usual distance and the observed interactions for larger distances, this so-called "simple interpolating function" is indeed the simplest – namely, it has the smallest computational complexity.

1 Formulation of the Problem

What started the whole thing: discrepancy between Newton's gravity, visible masses, and observations. According to Newton's theory of gravitation, bodies with masses m and M attract each other with the force

Olga Kosheleva
Department of Teacher Education, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: olgak@utep.edu

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: vladik@utep.edu

$$F = \frac{G \cdot m \cdot M}{r^2}, \tag{1}$$

where G is a constant and r is the distance between the bodies. In accordance to Newton's Second Law, this force leads to an acceleration a of the mass-m body such that

$$F = m \cdot a. \tag{2}$$

Newton's theory of gravitation provides a very accurate description of all the motions in the Solar system: suffice it to say that the difference between Newton's theory and observations of Mercury's position – the difference that was one of the reasons for replacing this theory with a more accurate General Relativity Theory – amounted to less than a half angular second per year; see, e.g., [1, 6, 8].

In this application of Newton's theory, masses can be determined by observing the motion of celestial bodies. Specifically, if a body of mass m rotates with velocity v around a body of mass M at a distance r, then its acceleration is equal to

$$a = \frac{v^2}{r}. \tag{3}$$

By equating the force (2) corresponding to this acceleration with the gravitational force (1), we conclude that

$$\frac{m \cdot v^2}{R} = \frac{G \cdot m \cdot M}{r^2}, \tag{3}$$

thus we can compute the mass M of the central body as

$$M = \frac{v^2 \cdot R}{G}. \tag{4}$$

From the formula (3), we can also conclude that

$$v^2 = \frac{G \cdot M}{r}, \tag{5}$$

i.e., that for objects rotating around the same central body at different distances r, the square of the velocity is proportional to $1/r$.

Since Newton's theory works so well in the Solar system, a natural idea is to use it to describe other celestial bodies – e.g., rotation of stars in a stellar cluster or in a galaxy. Surprisingly, it turned out that, contrary to the formula (5), when we get to the faraway area, where there are very few visible sources, the velocity stays the same – and does not decrease with r; see, e.g., [7].

The usual explanation is that, in addition to the visible objects, the Universe contains a large amount of cold matter that practically does not emit radiation and is, thus, not visible to our telescopes. This *dark matter* is what most physicists believe in.

MOND – an alternative theory of gravitation. While the presence of dark matter explains a lot of empirical facts, there are also some serious issues with this idea.

One of the main issues is that the main idea behind dark matter is that it should be reasonably independent from the usual matter, there should be very little correlation between the usual matter and the dark matter. However, the actual values of the dark matter mass obtained from the observed velocities seem to be well correlated with the amount of usual (visible) mass.

This correlation led some physicists to believe that there is no such thing as dark matter, but that, instead, Newton's theory needs to be modified. One of the ideas – that got some observable confirmations – is called Modified Newton's Dynamics, MOND, for short. In this theory, instead of the usual Newton's Second Law (2), we have a modified Newton's law

$$F = m \cdot g(a), \tag{6}$$

for an appropriate function $g(a)$; [3, 4, 5]. This theory is consistent with several other astrophysical phenomena; see, e.g., [2] and references therein.

Comment. To be precise, MOND uses an equivalent but slightly different formula $F = m \cdot a \cdot f(a)$ for some function $f(a)$. This is equivalent to our formula (6) if we take $g(a) = a \cdot f(a)$.

Specifics of MOND. In the Solar system, Newton's theory works well, So, it is reasonable to conclude that for reasonable-size accelerations a, we should have $g(a)$ asymptotically equivalent to a. On the other hand, in cluster- and galaxy-size range, where the distances are much larger and accelerations are much smaller, the force is proportional to r^{-2}, while the observed acceleration (3) decreases as r^{-1}. Thus, in this case, the gravitational force is proportional to a^2, so we should gave $g(a)$ asymptotically equivalent to $c \cdot a^2$ for some constant c. So, it is reasonable to impose the following requirement on $g(a)$.

Definition 1. *We say that the functions $f(x)$ and $g(x)$ are* asymptotically equivalent *when x tends to x_0 when*

$$\lim_{xc \to x_0} \frac{f(x)}{g(x)} = 1$$

.

Definition 2. *We say that a function $g(a)$ is* consistent with observations *if:*

- *this function is asymptotically equivalent to a when $a \to \infty$, and*
- *this function is asymptotically equivalent to $c \cdot a^2$ (for some constant c) when $a \to 0$.*

Several such functions have been proposed. At present, one of the most effective is the function

$$f(a) = \frac{a^2}{a + a_0} \tag{7}$$

known as *simple interpolating function*.

Formulation of the problem. How can we explain why, out of all functions that satisfy Definition 2, namely the function (7) is the most efficient?

What we do in this paper. In this paper, we provide a possible explanation for this effectiveness. This explanation is based on the physicists' belief that our of all possible physical theories consistent with observations, the simplest one is the true one [1, 8]. In line with this idea, we show that of all the functions that satisfy Definition 2, the function (7) is the simplest – in the sense that it has the smallest computational complexity.

2 Our Explanation

How can we define computational complexity. In a computer, only arithmetic operations are handware supported, all other computations are reduced to a sequence of arithmetic operations. For example, if we ask a computer to compute $\exp(x)$, what it will actually do is compute the sum of the first few terms of the corresponding Taylor series, i.e., the expression

$$\exp(x) \approx 1 + x + \frac{x^2}{2} + \ldots + \frac{x^N}{N!}.$$

Alternatively, it may compute an approximating fractional-linear expression. In all these cases, each computation consists of several computational steps, each of which is addition $+$, subtraction $-$, multiplication $\times$, or division $/$.

Out of these operations:

- addition and subtraction are the fastest,
- multiplication takes longer time – since multiplication is, in effect, a sequence of additions, and
- division takes the longest time – since the usual way to perform division is to perform several multiplications.

Let us denote the times needed for each arithmetic operation as, correspondingly, $t_\pm$, $t_\times$, and $t_/$. Then, we have

$$t_\pm < t_\times < t_/. \tag{8}$$

Definition 3. *By a* sequence of arithmetic operations*, we means a finite sequence of triples $T_1, \ldots, T_n$, where each triple t has the form $\langle op, v_1, v_2 \rangle$, where:*

- *op is equal to $+$, $-$, $\times$, or $/$, and*
- *for each tuple T_i, each v_j is:*
  - *either a symbol x_j, where j is a positive integer smaller than i,*
  - *or a real number,*
  - *or the original variable a.*

By the result *of applying a sequence to a number a, we mean the value of the last of numbers $x_1, \ldots, x_n$, where each x_i is the result of applying the corresponding operation op to the values v_1 and v_2.*

Examples.

- The expression $(a+1)\cdot(a+2)$ can be computed by the following sequence of arithmetic operations: $T_1 = \langle +, a, 1\rangle$, $T_2 = \langle +, a, 2\rangle$, and $T_3 = \langle \times, x_1, x_2\rangle$. In this case, $x_1 = a+1$, $x_2 = a+2$, and $x_3 = x_1 \cdot x_2$.
- The expression (7) can be computed by the following sequence of arithmetic operations: $T_1 = \langle \times, a, a\rangle$, $T_2 = \langle +, a, a_0\rangle$, and $T_3 = \langle /, x_1, x_2\rangle$. In this case, $x_1 = a \cdot a = a^2$, $x_2 = a + a_0$, and $x_3 = x_1/x_2 = a^2/(a+a_0)$.

Definition 4. *Let t, $t_\times$, and $t_/$ be positive real numbers for which $t_\pm < t_\times < t_/$. By the* computational complexity *of a sequence of arithmetic operations, we mean the sum of the numbers $t_{\rm op}$ corresponding to these operations.*

Examples.

- Computation of $(a+1)\cdot(a+2)$ consists of two additions and one multiplication, so the complexity is $2t_\pm + t_\times$.
- Computation of the expression (7) consists of one addition (to compute $a + a_0$), one multiplication (to compute $a^2 = a \cdot a$), and one division, so the complexity is $t_\pm + t_\times + t_/$.

Proposition. *Out of all functions that are consistent with observations (in the sense of Definition 2), the function (7) has the smallest computational complexity.*

Comment. It should be mentioned that this result holds no matter what values $t_\pm < t_\times < t_/$ we select.

Proof.

1°. Let us first us first prove that a function $g(a)$ that satisfies Definition 2 must contain at least one division.

Indeed, otherwise, it would be a polynomial, and the only way a polynomial is asymptotically equivalent to a for $a \to \infty$ is when it is a linear polynomial – otherwise, the highest power of a will dominate for large a. However, a linear polynomial cannot be asymptotically equivalent to $c \cdot a^2$ when $a \to 0$.

2°. Let us prove that the expression $g(a)$ cannot consist of only divisions and multiplications – and thus, it must contain at leats one addition or subtraction.

Indeed, in the beginning, all we have is a and constants c. Both inputs are expressions of the type $c \cdot a^d$ for some c and integer d: $a = 1 \cdot a^1$ and $c = c \cdot a^0$. One can easily check that if we multiply or divide such expressions, we will still have an expression of this type. However, no expression of this type satisfies Definition 2:

- the first condition from this definition implies that $d = 1$, while
- the second condition is only satisfied when $d = 2$.

3°. So, the desired expression must contain at least one division and at least one addition of subtraction. Depending on how many divisions and multiplications are used, we can consider the following three cases:

- The first case is when this expression contains two (or more) divisions.
- The second case is when this expression contains a single division, and all other operations are additions or subtractions.
- The remaining third case is when this expression contains a single division and at least one multiplication.

Let us consider these cases one by one.

3.1°. Let us first consider the first case, when the expression contains two (or more) divisions.

As we have shown in Part 2 of this proof, the expression must contain at least one addition of subtraction. Thus, the overall computational complexity must be larger than or equal to $2t_/ + t_\pm$. Since $t_/ > t_\times$, this value is larger than the value $t_/ + t_\times + t_\pm$ corresponding to (7). Since the expression (7) satisfies Definition 2, this means that the expression that contains two or more divisions cannot have the smallest computational complexity among all expressions that satisfy this definition.

3.2°. Let us now consider the second case, when the expression contains a single division, and when all other operations are additions or subtractions.

If we simply apply addition and subtraction to a and constants, all we get is expressions of the type $n \cdot a + c$, for some constants n and c. So, all we get after dividing such expressions is an expression of the type

$$\frac{n \cdot a + c}{n' \cdot a + c'},$$

and all we get by further additions are expressions of the type

$$\frac{n \cdot a + c}{n' \cdot a + c'} + n'' \cdot a + c''.$$

If we bring this sum to the common denominator, we get a fraction in which the numerator is quadratic in a and the denominator is linear in a:

$$\frac{n_2 \cdot a^2 + n_1 \cdot a + n_0}{d_1 \cdot a + d_0},$$

for some coefficients n_i and d_i.

If d_1 was equal to 0, then this expression could have been obtained without using division, and we have already shown, in Part 1 of this proof, that this is not possible. Thus, $d_1 \neq 0$. So, we can divide both the numerator and denominator by d_1 and get a simpler expression

$$\frac{N_2 \cdot a^2 + N_1 \cdot a + N_0}{a + a_0}, \text{ where } N_i \stackrel{\text{def}}{=} \frac{n_i}{d_1} \text{ and } a_0 \stackrel{\text{def}}{=} \frac{d_0}{d_1}.$$

If N_2 was equal to 0, then for $a \to \infty$, this expression would be a constant, but, according to Definition 2, it should grow as a. Thus, $N_2 \neq 0$, and for $a \to \infty$, this expression grows as $N_2 \cdot a$. Due to Definition 2, this implies that $N_2 = 1$, so we can further simplify this expression, into

$$\frac{a^2 + N_1 \cdot a + N_0}{a + a_0}.$$

If a_0 was equal to 0, then this expression would be equal to

$$a + N_1 + \frac{N_0}{a},$$

that does not provide the right a^2 asymptotic. Thus, $a_0 \neq 0$. So, for $a \to 0$, the denominator is asymptotically a constant, and thus, the corresponding asymptotic behavior of the ratio is propotional to the numerator.

If $N_0 \neq 0$, then:

- this expression would be tending to a non-zero constant when $a \to 0$,
- while we should have $\sim c \cdot a^2$.

Thus, we have $N_0 = 0$.

In this case, if we had $N_1 \neq 0$, then for $a \to 0$:

- this expression will be asymptotically linear,
- while we assumed it to be quadratic.

Thus, we must have $N_1 = 0$ as well.

In this case, the above expression takes the desired form (7).

3.3°. Finally, let us consider the case when, in addition to division and addition/subtraction, we also have at least one multiplication.

The expression (7) satisfies Definition 2 and contains exactly one of each operations, so it clearly has the smallest computational complexity of all such expressions. So, to complete the proof, we need to show that no other combination of these three operations leads to a function that satisfies Definition 2. Let us consider all such combinations.

4°. Similarly to Part 3.2, we can show that we cannot get any different function $g(a)$ if we simply add multiplication by a constant. Thus, we need at least one multiplication of expressions that contain a.

5°. Let us prove that for functions $g(a)$ that satisfy Definition 2, division cannot be the first operation.

Indeed, in the beginning, all we have is the variable a and constants. So, as a result of division, we get one of the following expressions:

- either divide a by itself, resulting in 1;
- or divide a by a constant c, resulting in $c^{-1} \cdot a$;
- or divide a constant by another constant – resulting in a new constant;
- or divide a constant by a, resulting in $c \cdot a^{-1}$.

In all these cases, we have an expression of the type $c \cdot a^n$ for some integer n. When we add and multiply such expression, all we have is a polynomial in terms of a and a^{-1}, i.e., a linear combination of the terms of this type. We can sort these terms in the increasing order of power and get

$$g(a) = c_1 \cdot a^{n_1} + \ldots + c_m \cdot a^{n_m} \text{ for } n_1 < \ldots < n_m.$$

Here:

- for $a \to \infty$, this expression is asymptotically equivalent to $c_m \cdot a^{n_m}$, so we must have $n_m = 1$,
- but for $a \to 0$, this expression is asymptotically equivalent to $c_1 \cdot a^{n_1}$, so we must have $n_1 = 2$.

This contradicts to the fact that $n_1 < n_m$. The statement is proven.

6°. Let us prove that for 3-step computations consisting of addition (or subtraction), multiplication, and division, addition/subtraction cannot be the last operation.

Indeed, based on Part 2 of this proof, as a result of division and multiplication, we can only get expressions of the type $c \cdot a^n$ for some integer n. If we apply addition ot subtraction, we get a linear combination of such expressions, and we have shown, in Part 5 of this proof, that such linear combination cannot satisfy Definition 2.

7°. Now we know that in the currently considered third case, when the expression contains a single division and at least one multiplication, the fastest computation scheme consists of 3 operations: addition or subtraction, multiplication, and division.

We also know that:

- addition/subtraction cannot be the last operation, so it must be first or second, and
- division cannot be the first operation, so it must be second or third.

In view of these two requirements, let us analyze what are the possible orders of these three operations.

- If addition/subtraction is the first, then division can be second or third, and multiplication must be, correspondingly, the third or the second one. So, in this case, we can have the following sequence of operations: $\pm, /, \times$ or $+, \times, /$.
- If addition/subtraction is the second operation, then division cannot be second, it must be third, and the only remaining place for multiplication is to be the first operation. In this case, we have the following sequence of operations: $\times, \pm, /$.

Let us consider the resulting three possible orders one by one.

8°. Let us first consider the order $\pm, /, \times$.

For this order, for the first operation, we start with a and constant. Thus, we have two possible non-trivial options for the first operation:

- either add or subtract a and a constant, resulting in $c \pm a$, or

- add two a's, resulting in $2a$.

The remaining two possible version of the first operations do not lead to any non-trivial results. Indeed:

- If we add two constants, we get a new constant – so this step is not needed and cannot be part of the shortest computation.
- Similarly, if we subtract a from a, we get a constant 0, which also cannot be a part of the shortest computation.

In both non-trivial operations, after the first step, we get expressions that are linear in terms of a.

After the first step, we divide two expressions A/B, and then perform multiplication. This multiplication must include the result of division – otherwise, this result is not used and we could get a faster computation scheme by skipping this step. So, we:

- either multiply A/B by some other expression C,
- or multiply the expression A/B by itself.

Let us consider these two options one by one.

8.1°. Let us first consider the option when we multiply A/B by C.

For this option, we get $(A/B)\cdot C$, which is equivalent to $(A\cdot C)/B$, i.e., we get a quadratic expression divided by a linear expression. We have already shown – in Part 3.2 of this proof – that in this case, Definition 2 leads to the expression (7).

8.2°. Let us now consider the option when we multiply the fractional-linear expression A/B by itself.

If the denominator was a constant, then we would not need division, and we know, from Part 1 of this proof, that this is not possible. Thus, the denominator is a linear function. In this case, when $a\to\infty$, then A/B tends to a constant. Thus, the product $(A/B)^2$ cannot be asymptotically equivalent to a. So, in this case, we do not get an expression that satisfies Definition 2.

9°. Let us now consider the order $\pm, \times, /$.

In the first step, we get an expression which is linear in a. After the multiplication step, we get an expression which is quadratic in a. We must use this quadratic expression in division – otherwise, multiplication would not be needed at all, and we have already considered this case. There are thus two options here:

- We can divide the quadratic expression by a linear expression. Then, as we have already shown in Part 3.2 of this proof, we get the expression (7).
- Alternatively, we can divide a linear expression by a quadratic expression. Then the ratio cannot grow as a when $a\to\infty$.

So, in this case, we also get only the expression (7).

10°. Finally, let us consider the remaining order $\times, \pm, /$.

For this order, we first multiply. We start with a and constants. Thus, we get the following options:

- We can multiply a constant by a constant – but this will simply lead to a new constant, that we could have gotten from the very beginning. So, this step cannot be part of the computation with the smallest computational complexity.
- We can multiply a by a constant, resulting in $c \cdot a$. In this case, on the second $\pm$ step, we will get linear functions of a, so the ratio obtained on the third computational step will be fractionally linear. And we have already shown that fractionally linear functions do not satisfy Definition 2.
- The only remaining option is to multiply a by a, resulting in a^2. In this option, as a result of $\pm$, we get quadratic functions, so as a result of division, we get either ratio of quadratic and linear function – which we already know leads to (7), or a ratio of two quadratic functions which tends to a constant when $a \to \infty$ and thus, also does not satisfy Definition 2.

So, the only possible case is the expression (7).

11°. So, we have shown that in all possible cases, the only expression $g(a)$ with the smallest computational complexity is indeed the expression (7). The proposition is proven.

Acknowledgments

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

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

References

1. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.
2. P. Kroupa, T. Jerabkova, I. Thies, J. Pflamm-Altenburg, B. Famaey, H. Boffin, J. Dabringhausen, G. Beccari, T. Prusti, C. Boily, H. Haghi, X. Wu, J. Haas, A. H. Zonoozi, G. Thomas, L. Šubr, and S. J. Aarseth, "Asymmetrical tidal tails of open star clusters: stars crossing their cluster's práh challenge Newtonian gravitation", *Monthly Notices of the Royal Astronomical Society*, 2022, Vol. 517, No. 3, pp. 3613–3639, doi:10.1093/mnras/stac2563.
3. M. Milgrom, "A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis", *Astrophysical Journal*, 1983, Vol. 270, pp. 365–370, doi:10.1086/161130

4. M. Milgrom, “A modification of the Newtonian dynamics – Implications for galaxies”, *Astrophysical Journal*, 1983, Vol. 270, pp. 371–383, doi:10.1086/161131.
5. M. Milgrom, “A modification of the Newtonian dynamics – Implications for galaxy systems”, *Astrophysical Journal*, 1983, Vol. 270, pp. 384–389, doi:10.1086/161132.
6. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, W. H. Freeman, New York, 1973.
7. V. C. Rubin and W. K. Ford Jr., “Rotation of the Andromeda Nebula from a spectroscopic survey of emission regions”. *The Astrophysical Journal*, 1970, Vol. 159, pp. 379–403, doi:10.1086/150317.
8. K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.