Ill-Conditioning Provoked by Scaling in Univariate Global Optimization and Its Handling on the Infinity Computer

Dmitri E. Kvasov[1,2,b)], Marat S. Mukhametzhanov[1,2,a)] and Yaroslav D. Sergeyev[1,2,c)]

[1]*University of Calabria, DIMES, 87036 Rende (CS), Italy*
[2]*Lobachevsky State University, IITMM, 603950 Nizhni Novgorod, Russia*

[a)]Corresponding author: m.mukhametzhanov@dimes.unical.it
[b)]kvadim@dimes.unical.it
[c)]yaro@dimes.unical.it

Abstract. Univariate Lipschitz global optimization problems are considered in this contribution. It is shown that in cases, where it is required to solve a scaled problem with very small or very large finite scaling constants, ill-conditioning can be provoked by scaling. It is established that this situation can be avoided using numerical infinities and infinitesimals. For this purpose, a new kind of a supercomputer – the Infinity Computer – that is able to work numerically with finite, infinite and infinitesimal numbers in a unique framework is used. Numerical experiments on benchmark test problems from the literature confirm the obtained results.

INTRODUCTION

Let us consider the following univariate Lipschitz global optimization problem:

$$f^* = f(x^*) = \min f(x), \quad x \in D = [a, b] \subset \mathbb{R}. \tag{1}$$

The objective function $f(x)$ in (1) is supposed to be multiextremal, non-differentiable and Lipschitz-continuous over the interval D, i. e.,

$$|f(x_1) - f(x_2)| \le L|x_1 - x_2|, \quad x_1, x_2 \in D, \tag{2}$$

where L is the finite Lipschitz constant.

The literature dedicated to problems (1), (2) is very wide (see, e. g., [1, 2, 3, 4, 5, 6, 7, 8]). There exists a huge number of methods and techniques for solving global optimization problems of this kind (see, e. g., [9, 10, 11, 12, 13, 14, 15, 16, 17]). For instance, in [18, 19, 20, 21], two classes of widely used methods are discussed and compared: deterministic global optimization algorithms and metaheuristic algorithms. Deterministic algorithms usually have strong convergence properties and are often used by specialists in academia. In their turn, metaheuristic methods usually do not have a proved convergence, but can be applied as meta-optimization and meta-learning techniques by engineers (see, e. g., [22]).

In practice, it can be useful to solve the problem (1), (2) with a scaled objective function $g(x) = \alpha f(x) + \beta$, where α, $\alpha > 0$, and β are the scaling constants (see, e. g., [23, 24, 25]). In [24, 25], the homogeneity and strong homogeneity properties for global optimization algorithms have been introduced. A global optimization algorithm is called *strongly homogeneous* if it generates the same sequences of trials (points where the objective function has been evaluated) while optimizing the functions $f(x)$ and the scaled functions $g(x) = \alpha f(x) + \beta$. Moreover, in [25, 26], the strong homogeneity of several univariate and multidimensional algorithms has been discussed using the Infinity Computer for working numerically with infinite and infinitesimal scales. In this contribution, three univariate methods are debated and tested using the well-known Pintér's class of univariate test functions from [27].

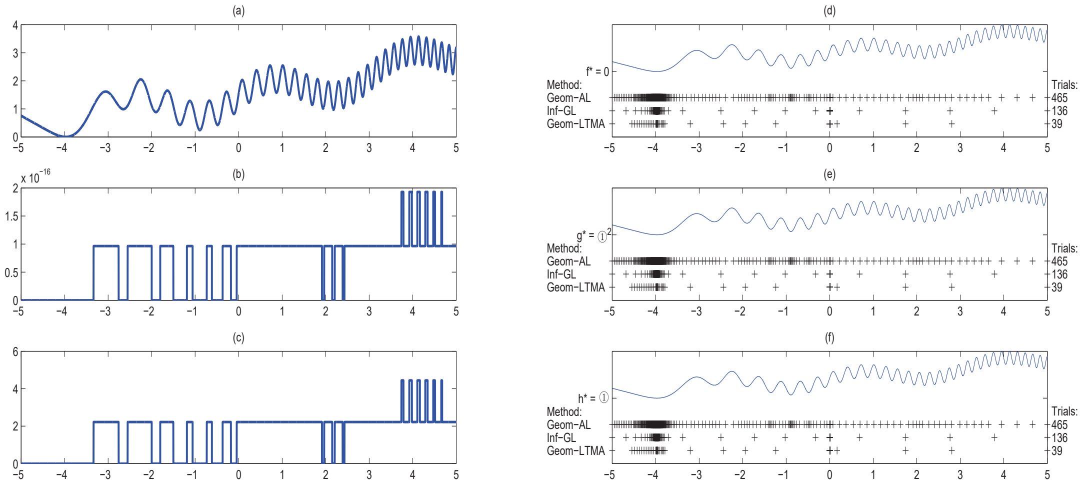

FIGURE 1. The graphs for (a) the test function $f_1(x)$ from (3), (b) the scaled function $\hat{g}_1(x) = 10^{-16} f_1(x) + 1$ in the logarithmic form, (c) the inverted scaled function $\hat{f}_1(x) = 10^{16}(g_1(x) - 1)$; and the experimental results for (d) the original test function $f_1(x)$, (e) the scaled test function $g_1(x) = ① f_1(x) + ①^2$ and (f) the scaled test function $h_1(x) = ①^{-1} f_1(x) + ①$. Trials are indicated by the signs "+" under the graphs of the functions and the number of trials for each method is indicated on the right. The results coincide for each method on all three test functions.

ILL-CONDITIONING PRODUCED BY SCALING

As it has been shown in [26], in practice it is not always possible to scale the objective function $f(x)$ using the finite scales α and β due to overflows and underflows present if traditional computers and numeral systems are used for evaluating the values of $g(x) = \alpha f(x) + \beta$. Let us discuss this problem using Pintér's class of univariate test problems. Each function $f_s(x)$, $1 \leq s \leq 100$, of this class is defined over the interval $[-5, 5]$ and has the following form:

$$f_s(x) = 0.025(x - x_s^*)^2 + \sin^2[(x - x_s^*) + (x - x_s^*)^2] + \sin^2(x - x_s^*), \tag{3}$$

where the global minimizer x_s^*, $1 \leq s \leq 100$, is chosen randomly (and differently for each test function of the class) from the interval $[-5, 5]$. In our tests this has been done by means of the random number generator used in the GKLS-generator of multidimensional test functions (see [28, 29]). It is easy to see that the global minimum f^* is equal to 0 for all test problems of the class (3).

As an example, let us consider the first test function $f_1(x)$ of the class (3) with $x_1^* = -3.9711481954$. Let us take $\alpha = 10^{-16}$ and $\beta = 1$ obtaining the scaled function $\hat{g}_1(x) = 10^{-16} f_1(x) + 1$. As can be seen in Figures 1.a and 1.b, the functions $f_1(x)$ and $\hat{g}_1(x)$ have completely different minimizers due to the used scaling constants in $\hat{g}_1$. As a result, if we wish to invert the function $\hat{g}_1(x)$ trying to establish the original function $f_1(x)$, i. e., to compute the function $\hat{f}_1(x) = 10^{16}(\hat{g}_1(x) - 1)$, then it will not coincide with $f_1(x)$. Figure 1.c shows $\hat{f}_1(x)$ constructed from $\hat{g}_1(x)$ using MATLAB® and the piecewise linear approximations with the step $h = 0.001$. On the one hand, due to underflows taking place in commonly used numeral systems (in this case, the type *double* in MATLAB®), the function $\hat{g}_1(x)$ degenerates over many intervals in constant functions and many local minimizers disappear (see, e. g., the local minimizers that belong to the interval $[0, 2]$ in Figures 1.a and 1.b). On the other hand, due to overflows, several local minimizers become global minimizers of the scaled function $\hat{g}_1(x)$. In particular, one can find from Figures 1.b and 1.c the following global minimizers of the functions $\hat{g}_1(x)$ and $\hat{f}_1(x)$: $(x^*, \hat{g}_1^*) = (-5, 1)$ and $(x^*, \hat{f}_1^*) = (-5, 0)$, while it can be seen from Figure 1.a that the point $x = -5$ is not even a local minimizer of the original function $f_1(x)$.

In [26], it has been shown that the ill-conditioning present in the global optimization problem described above in the traditional computational framework can be avoided in certain cases within the Infinity Computing paradigm (see, e. g., [30, 31]). Finite, infinite, and infinitesimal numbers in this framework are represented using the positional numeral system with the infinite radix ① (called *grossone*) introduced as the number of elements of the set of natural numbers (see [31] for details). This computational methodology has already been successfully applied in optimization (see, e.g., [32, 33]), in particular, in handling ill-conditioning in optimization (see, e.g., [33]), and in a number of other theoretical and applied research areas: cellular automata (see [34]), hyperbolic geometry (see [35]), percolation (see [36]), numerical solution of ordinary differential equations (see [37, 38, 39]), etc.

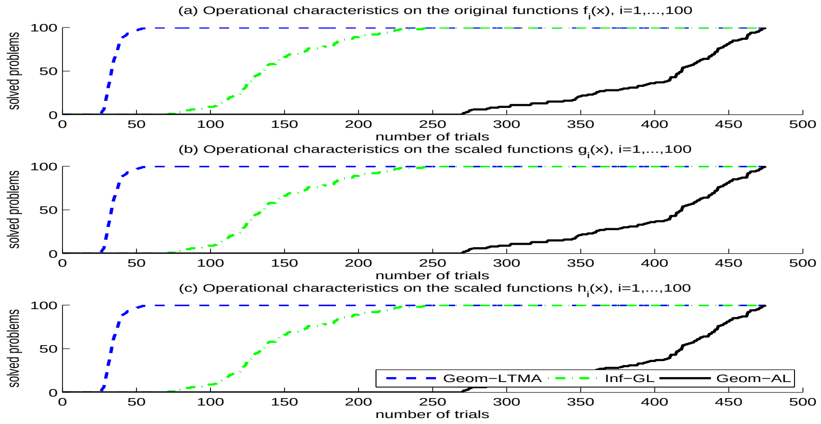

FIGURE 2. Operational characteristics for (a) the original class $f_i(x), 1 \leq i \leq 100$, (b) the scaled class $g_i(x) = ① f_i(x) + ①^2, 1 \leq i \leq 100$, and (c) the scaled class $h_i(x) = ①^{-1} f_i(x) + ①, 1 \leq i \leq 100$. The results coincide for each method on all three test classes.

ALGORITHMS AND NUMERICAL EXPERIMENTS

The following three algorithms being examples of concrete implementations of a general scheme of univariate global optimization algorithms introduced in [26, 40] have been tested in order to illustrate the advantages of using the numerical infinities and infinitesimals for avoiding ill-conditioning provoked by scaling: Geom-AL (Geometric method with an a priori given overestimate of the Lipschitz constant), Inf-GL (Information method with the global estimate of the Lipschitz constant), and Geom-LTMA (Geometric method with the “Maximum-Additive” local tuning).

Parameters of the methods have been chosen as follows. An a priori given overestimate of L from (2) used by the method Geom-AL has been calculated over the 10^{-7}-grid for each test problem and multiplied by α for the scaled functions $g(x)$. In algorithms Geom-LTMA and Inf-GL, the Lipschitz constant L is estimated during the search and the reliability parameter r is used (r was set to 1.1 and 2.0, respectively). The value $\epsilon = 10^{-4}(b - a)$ has been used in the stopping criterion. The algorithms Geom-AL, Inf-GL, and Geom-LTMA have been tested on 300 functions from the following three test classes: $f_i(x)$, from (3); $g_i(x) = ① f_i(x) + ①^2$; and $h_i(x) = ①^{-1} f_i(x) + ①,\ 1 \leq i \leq 100$.

It has been obtained that for each algorithm and for each test problem the results on the original functions $f_i(x)$ and scaled functions $g_i(x)$ and $h_i(x)$ using infinite and infinitesimal scaling constants coincide. For instance, in Figures 1.d, 1.e and 1.f, the results for the first test functions $f_1(x),\ g_1(x)$, and $h_1(x)$ are shown. In Figure 2, the operational characteristics for each algorithm on each test class are presented. An operational characteristic constructed on a class of 100 randomly generated test functions is a graph showing the number of solved problems in dependence on the number of executed evaluations of the objective function (see [20, 41]). Each test problem was considered to be solved if an algorithm generated a point in the ϵ-neighborhood of the global minimizer x^*. It can be seen from Figure 2 that the operational characteristics of the algorithms perfectly coincide for the presented three test classes showing that the ill-conditioning, present in the cases of finite constants α and β, disappear if the Infinity Computer is used.

To conclude, let us recall that traditional floating-point arithmetics does not allow one to work simultaneously with large and small numbers or with the large numbers of a significantly different order due to overflows and underflows. In this case, if finite scaling is needed (for instance, $\alpha = 10^{-16}$ and $\beta = 1$), the infinite and infinitesimal scales can be used in order to avoid ill-conditioning produced by scaling (in our case, $\alpha = ①^{-1}$ and $\beta = ①$). The final value f^* can be easily extracted from the obtained value $g^* = ①^{-1} f^* + ①$ and substituted by the required value $10^{-16} f^* + 1$.

ACKNOWLEDGMENTS

This research was supported by the INdAM-GNCS 2018 Research Project “Numerical methods in optimization and ODEs” (the authors are members of the INdAM Research group GNCS) and by the project “Smart Electronic Invoices Accounting” - SELINA CUP: J28C17000160006 (POR CALABRIA FESR-FSE 2014-2020).

REFERENCES

[1] K. A. Barkalov and R. G. Strongin, Comp. Math. Math. Phys. **42(9)**, 1289–1300 (2002).
[2] C. A. Floudas and P. M. Pardalos, eds., *Encyclopedia of Optimization*, 2nd ed. (Springer, NY, 2009).
[3] V. Gergel, V. Grishagin, and A. Gergel, J. Global Optim. **66**, 35–51 (2016).
[4] J. W. Gillard and A. A. Zhigljavsky, Commun. Nonlin. Sci. Numer. Simulat. **21**, 70–88 (2015).
[5] V. Grishagin, R. Israfilov, and Y. D. Sergeyev, Appl. Math. Comput. **318**, 270–280 (2018).
[6] Y. D. Sergeyev and D. E. Kvasov, SIAM J. Optim. **16**, 910–937 (2006).
[7] Y. D. Sergeyev, D. E. Kvasov, and F. M. H. Khalaf, Optim. Lett. **1**, 85–99 (2007).
[8] R. Strongin and Y. D. Sergeyev, Parallel Computing **18**, 1259–1273 (1992).
[9] K. Barkalov and V. Gergel, J. Global Optim. **66**, 3–20 (2016).
[10] R. Battiti, M. Brunato, and F. Mascia, *Reactive Search and Intelligent Optimization* (Springer, NY, 2009).
[11] V. P. Gergel, V. A. Grishagin, and R. A. Israfilov, Procedia Comput. Sci. **51**, 865–874 (2015).
[12] J. W. Gillard and D. E. Kvasov, Stat. Interface **10**, 59–70 (2017).
[13] D. E. Kvasov and Y. D. Sergeyev, Numer. Algebra Contr. Optim. **2**, 69–90 (2012).
[14] Y. D. Sergeyev, D. Famularo, and P. Pugliese, J. Global Optim. **21**, 317–341 (2001).
[15] Y. D. Sergeyev and D. E. Kvasov, *Deterministic Global Optimization: An Introduction to the Diagonal Approach*, SpringerBriefs in Optimization (Springer, NY, 2017).
[16] Y. D. Sergeyev, Comput. Math. Math. Phys. **35**, 705–717 (1995).
[17] A. A. Zhigljavsky and A. Žilinskas, *Stochastic Global Optimization* (Springer, New York, 2008).
[18] D. E. Kvasov and M. S. Mukhametzhanov, Appl. Math. Comput. **318**, 245–259 (2018).
[19] D. E. Kvasov and M. S. Mukhametzhanov, “One-dimensional global search: Nature-inspired vs. Lipschitz methods,” in *ICNAAM 2015: 13th International Conference of Numerical Analysis and Applied Mathematics*, AIP Conf. Proc., Vol. 1738, edited by T. E. Simos (American Institute of Physics, Melville, NY, 2016).
[20] Y. D. Sergeyev, D. Kvasov, and M. Mukhametzhanov, Scientific Reports **8**, p. 453 (2018).
[21] Y. D. Sergeyev, D. E. Kvasov, and M. S. Mukhametzhanov, Math. Comput. Simulation **141**, 96 – 109 (2017).
[22] Y. Gao, W. Du, and G. Yan, Scientific Reports **5**, p. 9295 (2015).
[23] R. G. Strongin, *Numerical Methods in Multiextremal Problems (Information-Statistical Algorithms)* (Nauka, Moscow, 1978) in Russian.
[24] S. Elsakov and V. Shiryaev, Comput. Math. Math. Phys. **50**, 1642–1654 (2010).
[25] A. Žilinskas, Appl. Math. Comput. **218**, 8131–8136 (2012).
[26] Y. D. Sergeyev, D. Kvasov, and M. Mukhametzhanov, Commun. Nonlin. Sci. Numer. Simul. **59**, 319–330 (2018).
[27] J. D. Pintér, “Global optimization: software, test problems, and applications,” in *Handbook of Global Optimization*, Vol. 2, edited by P. M. Pardalos and H. E. Romeijn (Kluwer Academic Publishers, Dordrecht, 2002), pp. 515–569.
[28] M. Gaviano, D. E. Kvasov, D. Lera, and Y. D. Sergeyev, ACM Trans. Math. Software **29**, 469–480 (2003).
[29] Y. D. Sergeyev, D. Kvasov, and M. Mukhametzhanov, “Emmental-type GKLS-based multiextremal smooth test problems with non-linear constraints,” in *LNCS*, Vol. 10556 (Springer, NY, 2017), pp. 383–388.
[30] Y. D. Sergeyev, *Arithmetic of Infinity* (Edizioni Orizzonti Meridionali, CS, 2003, 2nd ed. 2013).
[31] Y. D. Sergeyev, EMS Surv. Math. Sci. **4**, 219–320 (2017).
[32] S. De Cosmis and R. De Leone, Appl. Math. Comput. **218(16)**, 8029–8038 (2012).
[33] M. Gaudioso, G. Giallombardo, and M. S. Mukhametzhanov, Appl. Math. Comput. **318**, 312–320 (2018).
[34] L. D’Alotto, Appl. Math. Comput. **218(16)**, 8077–8082 (2012).
[35] M. Margenstern, Commun. Nonlin. Sci. Numer. Simul. **21(1–3)**, 3–11 (2015).
[36] D. Iudin, Y. Sergeyev, and M. Hayakawa, Commun. Nonlin. Sci. Numer. Simul. **20(3)**, 861–870 (2015).
[37] F. Mazzia, Y. D. Sergeyev, F. Iavernaro, P. Amodio, and M. S. Mukhametzhanov, “Numerical methods for solving ODEs on the Infinity Computer,” in *Proc. of the 2nd Intern. Conf. “Numerical Computations: Theory and Algorithms”*, AIP Conf. Proc., Vol. 1776, edited by Y. D. Sergeyev, D. E. Kvasov, F. Dell’Accio, and M. S. Mukhametzhanov (American Institute of Physics, Melville, NY, 2016).
[38] P. Amodio, F. Iavernaro, F. Mazzia, M. S. Mukhametzhanov, and Y. D. Sergeyev, Math. Comput. Simulation **141**, 24–39 (2016).
[39] Y. D. Sergeyev, M. S. Mukhametzhanov, F. Mazzia, F. Iavernaro, and P. Amodio, Int. J. Unconv. Comput. **12(1)**, 3–23 (2016).
[40] Y. D. Sergeyev, M. S. Mukhametzhanov, D. E. Kvasov, and D. Lera, J. Optim. Theory Appl. **171**, 186–208 (2016).
[41] V. A. Grishagin, “Operating characteristics of some global search algorithms,” in *Problems of Stochastic Search*, Vol. 7 (Zinatne, Riga, 1978), pp. 198–206, in Russian.