



Acceleration of Global Search by Implementing Dual Estimates for Lipschitz Constant

Roman Strongin , Konstantin Barkalov  , and Semen Bevzuk 

Lobachevsky State University of Nizhni Novgorod, Nizhny Novgorod, Russia
`konstantin.barkalov@itmm.unn.ru`

Abstract. The paper considers global optimization problems with a black-box objective function satisfying the Lipschitz condition. Efficient algorithms for this class of problems require reliable estimates of the Lipschitz constant to be introduced. Various approaches have been proposed to take into account both global and local properties of the objective function. In particular, algorithms using local estimates of the Lipschitz constant have shown their potential. The new approach presented in this paper is based on simultaneous use of two estimates: one is substantially larger than the other. The larger estimate ensures global convergence and the smaller one reduces the total number of trials needed to find the global optimizer. Results of numerical experiments on the random sample of multidimensional functions demonstrate the efficiency of the approach proposed by the authors.

Keywords: Global optimization · Multiextremal problems · Lipschitz constant estimates

1 Introduction

The paper considers global optimization problems of the form

$$\varphi(y^*) = \min \{\varphi(y) : y \in D\}, \quad (1)$$

$$D = \{y \in R^N : a_i \leq y_i \leq b_i, a_i, b_i \in R, 1 \leq i \leq N\}, \quad (2)$$

where the objective function is a black-box function and it is assumed to satisfy the Lipschitz condition

$$|\varphi(y_1) - \varphi(y_2)| \leq L \|y_1 - y_2\|, \quad y_1, y_2 \in D,$$

This research was supported by the Russian Science Foundation, project No.16-11-10150.

© Springer Nature Switzerland AG 2020

Y. D. Sergeyev and D. E. Kvasov (Eds.): NUMTA 2019, LNCS 11974, pp. 478–486, 2020.

https://doi.org/10.1007/978-3-030-40616-5_46

with the constant L , $L < \infty$, unknown a priori.

The assumption of the objective function to be Lipschitzian is typical of many approaches to the development of the global optimization algorithms [3, 11, 12, 17]. Moreover, the adaptive estimate of the unknown Lipschitz constant, based on the obtained search information, is one of the most important problems being solved in these algorithms. The value of the Lipschitz constant affects essentially the convergence rate of the global optimization algorithms. Therefore, the issue of its correct estimate is so important. The underestimation of the real value of this constant may result in losing the convergence of the algorithm to the global solution. At the same time, if the value of the constant estimate for the objective function is too large and does not match its real behavior, this will slow down the convergence of the algorithm to the global minimizer.

Several typical methods of adaptive estimation of the Lipschitz constant are known:

- global estimation of the constant L in the whole search domain D [7, 12, 17].
- local estimations of the constants L_i in different subdomains D_i of the search domain D [9, 10, 15].
- the choice of the estimates of the constant L from a set of possible values [4, 8, 13].

Each of the above approaches has its own advantages and disadvantages. For example, the use of the global estimate over the whole search domain can slow down the convergence of the algorithm to the global minimizer. The use of the local estimates to accelerate the convergence of the method requires an adequate adjustment of the algorithm parameters in order to preserve the global convergence.

In the present work, we consider a new algorithm that uses two global estimates of the Lipschitz constant. One of the two estimates is much greater than the other one. The larger estimate ensures global convergence and the smaller one reduces the total number of trials needed to find the global optimizer. The choice of one of the two estimates in the algorithm is performed adaptively during the search phase.

A rigorous substantiation of the proposed approach goes beyond the present initial publication and will be done in the forthcoming works. Here we present the results of numerical experiments that clearly demonstrate the efficiency of the new algorithm. Several hundred multiextremal test problems of various dimensionalities have been solved in the course of numerical experiments.

2 Global Search Algorithm and Dimensionality Reduction

The adaptation of the efficient algorithms that solve one-dimensional problems to solve multidimensional problems is a typical method to construct global optimization algorithms, see, for example, the diagonal partitions method in [13] or the simplicial partitions method in [18].

In this paper, we follow the approach based on the idea of reducing the dimension with the use of the Peano-Hilbert curves [16,17], which continuously and unambiguously map the unit interval $[0, 1]$ onto the N -dimensional cube D from (2). By using this kind of mapping, it is possible to reduce the multidimensional problem (1) to a univariate problem

$$\varphi(y^*) = \varphi(y(x^*)) = \min \{ \varphi(y(x)) : x \in [0, 1] \},$$

where the function $\varphi(y(x))$ will satisfy a uniform Hölder condition

$$| \varphi(y(x_1)) - \varphi(y(x_2)) | \leq H |x_1 - x_2|^{1/N}$$

with the Hölder constant H linked to the Lipschitz constant L by the relation $H = 2L\sqrt{N + 3}$ and $y(x)$ is a Peano-Hilbert curve from $[0, 1]$ onto D . Note that theoretically the Peano-Hilbert curve $y(x)$ is defined as a limit object. Therefore, in practical implementation, only some approximation to the true space-filling curve can be constructed. Some methods for constructing this type of approximations (called evolvents) are considered in [16,17]. In this case, the accuracy of the evolvent approximation to the true curve $y(x)$ depends on the density of the evolvent m (which is a parameter for constructing the evolvent) and is of the order of 2^{-m} for each coordinate.

Let us call the process of computing a function value (including the construction of the image $y = y(x)$) as a *trial*, and the pair $\{x, z = \varphi(y(x))\}$ as the outcome of the trial.

The Divide-The-Best global search algorithm used in this paper (according to [17]) can be formulated as follows. The first two trials are executed at the points $y^0 = y(0), y^1 = y(1)$. The choice of the point $y^{k+1}, k \geq 1$, for the next $(k + 1)^{\text{th}}$ trial is defined by the following rules.

1. Renumber the preimages of all the points $y^i = y(x^i)$ from the trials already performed by subscripts in the increasing order of their coordinates, i.e.

$$0 = x_0 < x_1 < \dots < x_k = 1, \tag{3}$$

and associate these with the values $z_i = \varphi(y(x_i)), 0 \leq i \leq k$, computed at these points.

2. Compute the maximum absolute value of the first divided differences

$$M = \max_{1 \leq i \leq k} \frac{|z_i - z_{i-1}|}{\Delta_i},$$

where $\Delta_i = (x_i - x_{i-1})^{1/N}$ and let

$$\mu = \begin{cases} 1, & \text{if } M = 0, \\ M, & \text{if } M \neq 0. \end{cases} \tag{4}$$

3. For each interval $(x_{i-1}, x_i), 1 \leq i \leq k$, calculate the value $R(i)$ called the *characteristic* of the interval

$$R(i) = \Delta_i + \frac{(z_i - z_{i-1})^2}{r^2 \mu^2 \Delta_i} - 2 \frac{z_i + z_{i-1} - 2z^*}{r \mu}, \tag{5}$$

where

$$z^* = \min_{0 \leq i \leq k} z_i \quad (6)$$

and the real number $r > 1$ is a reliability parameter of the algorithm.

4. Select the interval (x_{t-1}, x_t) corresponding to the maximum characteristic

$$R(t) = \max_{1 \leq i \leq k} R(i). \quad (7)$$

5. Carry out the next trial at the point $x^{k+1} \in (x_{t-1}, x_t)$ calculated using the following formula

$$x^{k+1} = \frac{x_t + x_{t-1}}{2} - \text{sign}(z_t - z_{t-1}) \frac{1}{2r} \left[\frac{|z_t - z_{t-1}|}{\mu} \right]^N. \quad (8)$$

The algorithm terminates if the condition $\Delta_t < \epsilon$ is satisfied where t is from (7), and $\epsilon > 0$ is the predefined accuracy.

The theory of convergence of this algorithm is provided in [17]. The algorithm can be efficiently parallelized for shared and distributed memory [6] and for accelerators [1].

3 Algorithm with Dual Lipschitz Constant Estimates

The global search algorithm presented in the previous section is intended for solving the multiextremal problems, in which the objective function satisfies the Lipschitz condition. It is not necessary to define the value of the constant for the algorithm convergence. The estimation of the constant is performed in the course of global search based on available search information. According to the theorem from [17], the sequence of the trial points $\{y^k\}$ will converge to the global minimizer y^* if the condition

$$r\mu > 2^{3-1/N} L\sqrt{N+3} \quad (9)$$

is satisfied. Thus, an appropriate choice of the parameter r from (5) allows using the value $(r\mu)/(2^{3-1/N}\sqrt{N+3})$ as an estimate of the Lipschitz constant for the objective function $\varphi(y)$.

Satisfying the condition (9) will be guaranteed if we choose a large enough value of the parameter r . However, in this case the method will perform a large number of trials until the stop condition is satisfied. The choice of a small value of the parameter r (that corresponds to the lower estimate of the Lipschitz constant) would considerably reduce the number of trials but may violate the convergence to the global extremum.

An approach, in which two estimates of the Lipschitz constant are used in the rules of the algorithm, seems quite promising. This approach implies the use in the algorithm of two parameters r_{glob} and r_{loc} , where $r_{glob} > r_{loc} > 1$. When using the parameter r_{loc} we shall deal with the smaller estimate of the Lipschitz constant, and using the parameter r_{glob} will correspond to the larger one.

The rules of the algorithm with two estimates of the Lipschitz constant reproduce the ones of the global search algorithm completely except Rule 3 (the computation of the characteristic) and Rule 4 (search for the interval with the maximum characteristic).

The new rule for calculating the characteristic $R(i)$ of the interval (x_{i-1}, x_i) will consist of the following operations:

- Calculate the value $R_{glob}(i)$ corresponding to the larger estimate of the Lipschitz constant

$$R_{glob}(i) = \Delta_i + \frac{(z_i - z_{i-1})^2}{r_{glob}^2 \mu^2 \Delta_i} - 2 \frac{z_i + z_{i-1} - 2z^*}{r_{glob} \mu}.$$

- Calculate the value $R_{loc}(i)$ corresponding to the smaller estimate of the Lipschitz constant

$$R_{loc}(i) = \Delta_i + \frac{(z_i - z_{i-1})^2}{r_{loc}^2 \mu^2 \Delta_i} - 2 \frac{z_i + z_{i-1} - 2z^*}{r_{loc} \mu}.$$

- Determine the characteristic $R(i)$ as

$$R(i) = \max\{\rho R_{loc}(i), R_{glob}(i)\}, \text{ where } \rho = \left(\frac{1 - 1/r_{glob}}{1 - 1/r_{loc}} \right)^2, \quad (10)$$

The new rules for finding the interval with the maximum characteristic will be as follows:

- Select the interval (x_{t-1}, x_t) corresponding to the maximum characteristic $R(t) = \max_{1 \leq i \leq k} R(i)$.
- Fix the value $r = r_{loc}$ if $\rho R_{loc}(t) > R_{glob}(t)$, otherwise fix $r = r_{glob}$.
- Use this value of r in Rule 5 of the algorithm in the computing of the next trial point.

This method for computing the characteristic can be substantiated as follows. Each search iteration will yield an interval with the current minimum value z^* from (6) at one of its boundaries. In the final phase of the search, this interval will correspond to the interval containing the global minimum, i.e. it will be the best one in terms of conducting further trials within it.

Let the current minimum value of z^* from (6) be achieved at the left point of the i^{th} interval, i.e. $z^* = z_{i-1}$. As proven in [17], according to the rule (5) the following inequality will be true:

$$R(i) \geq \Delta_i (1 - 1/r)^2.$$

Therefore, for the estimates of the characteristics $R_{loc}(i)$ and $R_{glob}(i)$ calculated with different parameters r_{loc} and r_{glob} , the following relation will hold:

$$\Delta_i (1 - 1/r_{glob})^2 > \Delta_i (1 - 1/r_{loc})^2.$$

Thus, when choosing the largest of the characteristics in accordance with

$$R(i) = \max\{R_{loc}(i), R_{glob}(i)\}$$

the characteristic $R_{glob}(i)$ corresponding to the higher estimate of the Lipschitz constant will be chosen; the lower estimate (which speeds up the process of refining the current solution) will not be used. However, if we multiply the lower estimate for the characteristic $R_{loc}(i)$ by the coefficient ρ in accordance with (10), then such lower estimates will be equal, thus the choice of the characteristic $R_{loc}(i)$ corresponding to the lower estimate of the Lipschitz constant will become more likely.

4 Numerical Experiments

A numerical comparison of the algorithms was carried out by solving several series of problems generated by the GKLS generator [5]. This generator of multiextremal functions is widely used to compare global optimization methods (see, for example, [2, 13, 14]). In this study, six series each containing 100 problems of dimensions $N = 3, 4, 5$ were solved. For each dimension, *Simple* and *Hard* problems were generated, differing in the size of the attraction regions for local extremums and global extremum. The problem was considered solved if the method conducted the trial at a point that was in the δ -neighborhood of the global minimizer y^* , i.e. $\|y^k - y^*\| < \delta \|b - a\|$, where a and b are the boundaries of the search domain D .

Each series of problems has been solved by the original global search algorithm (GSA) and by the method with two estimates of the Lipschitz constant (GSA-DL). The evolvent constructed using the parameter $m = 10$ was used for the dimensionality reduction in both algorithms. The relative accuracy of the solution search was $\delta = 0.01$. The maximum allowable number of iterations per problem was $K_{max} = 10^6$.

The averaged numbers of iterations performed by the algorithms are presented in Table 1. For the GSA method the values of the parameter $r = 4.8$ when solving the problems of the *Simple* class and $r = 5.6$ when solving the problems of the *Hard* class were used. These values are the minimum ones (with the accuracy 0.1), at which all problems have been solved successfully. When solving the problems from the above classes by the GSA-DL method, the value of the parameter r specified above was selected for the upper estimate of the Lipschitz constant, i.e. the value $r_{glob} = r$ was set, which was complemented by the values $r_{loc} = 1.8$, $r_{loc} = 2.1$ and $r_{loc} = 2.4$. The number of unsolved problems is specified in brackets.

Table 1. Average number of iterations

	$N = 3$		$N = 4$		$N = 5$	
	<i>Simple</i>	<i>Hard</i>	<i>Simple</i>	<i>Hard</i>	<i>Simple</i>	<i>Hard</i>
AGS	2444	5345	28415	77470	25220(1)	126138(4)
AGS-DL, $r_{loc} = 1.8$	1372	2632	13273	37715	12702	94296(1)
AGS-DL, $r_{loc} = 2.1$	1502	2805	14826	38843	15213	90792(2)
AGS-DL, $r_{loc} = 2.4$	1567	2868	19447	40342	18239	106438(2)

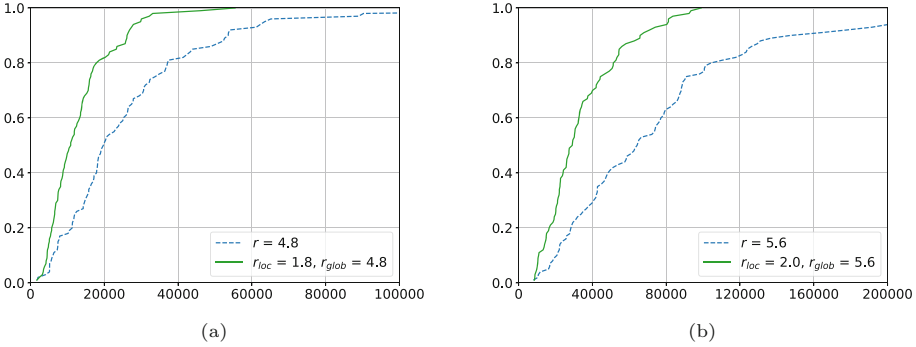


Fig. 1. Operational characteristics for GKLS *Simple* (a) and *Hard* (b) classes, $N = 4$.

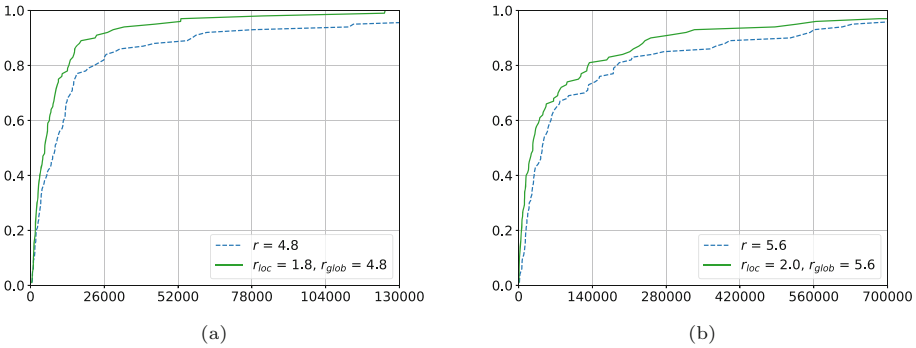


Fig. 2. Operational characteristics for GKLS *Simple* (a) and *Hard* (b) classes, $N = 5$.

The advantages of the GSA-DL algorithm over its prototype are also confirmed by the operational characteristics of the algorithms as well. Assume a series of test problems to be solved. The results of solving the series can be presented by a function $p(k)$ featuring the fraction of the total number of problems solved in k iterations. Such a function will be called the *operational characteristic* of the algorithm.

The operational characteristics for the GSA and GSA-DL methods obtained when solving the *Simple* and *Hard* problem series with the dimensionalities

$N = 4$ and $N = 5$ are presented in Figs. 1 and 2, respectively. The values of the parameters r used for estimating the Lipschitz constant are given in the figures.

The lower curves in Figs. 1 and 2 feature the characteristics of the GSA method whereas the upper ones, those of the GSA-DL. Such relative positions of the curves show the algorithm with two estimates of the Lipschitz constant is much faster on average when solving the problem series than the algorithm using a single estimate of the constant. Note that to solve problems of all classes (except for the *Hard* class at $N = 5$) the GSA-DL method requires about half as many trials as the GSA. The deterioration of the results in the case of the *Hard* class at $N = 5$ is explained by the complexity of this class' functions, which have a large attraction region of local minima and a narrow attraction region of the global minimum. To correctly solve such problems, the GSA-DL method often uses a higher estimate of the Lipschitz constant, thus reducing the speed difference of the GSA and GSA-DL algorithms.

References

1. Barkalov, K., Gergel, V., Lebedev, I.: Solving global optimization problems on GPU cluster. In: AIP Conference Proceedings, vol. 1738, p. 400006 (2016)
2. Barkalov, K., Strongin, R.: Solving a set of global optimization problems by the parallel technique with uniform convergence. *J. Glob. Optim.* **71**(1), 21–36 (2018)
3. Evtushenko, Y., Posypkin, M.: A deterministic approach to global box-constrained optimization. *Optim. Lett.* **7**, 819–829 (2013)
4. Gablonsky, J.M., Kelley, C.T.: A locally-biased form of the DIRECT algorithm. *J. Glob. Optim.* **21**(1), 27–37 (2001)
5. Gaviano, M., Kvasov, D.E., Lera, D., Sergeyev, Y.D.: Software for generation of classes of test functions with known local and global minima for global optimization. *ACM Trans. Math. Softw.* **29**(4), 469–480 (2003)
6. Gergel, V.P., Strongin, R.G.: Parallel computing for globally optimal decision making. In: Malyshkin, V.E. (ed.) PaCT 2003. LNCS, vol. 2763, pp. 76–88. Springer, Heidelberg (2003). <https://doi.org/10.1007/978-3-540-45145-7-7>
7. Horst, R., Tuy, H.: *Global Optimization - Deterministic Approaches*. Springer, Berlin (1996). <https://doi.org/10.1007/978-3-662-02598-7>
8. Jones, D., Perttunen, C., Stuckman, B.: Lipschitzian optimization without the Lipschitz constant. *J. Optim. Theory Appl.* **79**(1), 157–181 (1993)
9. Kvasov, D.E., Pizzuti, C., Sergeyev, Y.D.: Local tuning and partition strategies for diagonal GO methods. *Numer. Math.* **94**(1), 93–106 (2003)
10. Lera, D., Sergeyev, Y.D.: An information global minimization algorithm using the local improvement technique. *J. Glob. Optim.* **48**(1), 99–112 (2010)
11. Paulavičius, R., Žilinskas, J., Grothey, A.: Investigation of selection strategies in branch and bound algorithm with simplicial partitions and combination of Lipschitz bounds. *Optim. Lett.* **4**(2), 173–183 (2010)
12. Pinter, J.D.: *Global Optimization in Action (Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications)*. Kluwer Academic Publishers, Dordrecht (1996). <https://doi.org/10.1007/978-1-4757-2502-5>
13. Sergeyev, Y.D., Kvasov, D.E.: Global search based on efficient diagonal partitions and a set of Lipschitz constants. *SIAM J. Optim.* **16**(3), 910–937 (2006)

14. Sergeyev, Y.D., Kvasov, D.E.: A deterministic global optimization using smooth diagonal auxiliary functions. *Commun. Nonlinear Sci. Numer. Simul.* **21**(1–3), 99–111 (2015)
15. Sergeyev, Y.D., Mukhametzhanov, M.S., Kvasov, D.E., Lera, D.: Derivative-free local tuning and local improvement techniques embedded in the univariate global optimization. *J. Optim. Theory Appl.* **171**(1), 186–208 (2016)
16. Sergeyev, Y.D., Strongin, R.G., Lera, D.: *Introduction to Global Optimization Exploiting Space-Filling Curves*. Springer Briefs in Optimization. Springer, New York (2013). <https://doi.org/10.1007/978-1-4614-8042-6>
17. Strongin, R.G., Sergeyev, Y.D.: *Global Optimization with Non-convex Constraints. Sequential and parallel Algorithms*. Kluwer Academic Publishers, Dordrecht (2000). <https://doi.org/10.1007/978-1-4615-4677-1>
18. Žilinskas, J.: Branch and bound with simplicial partitions for global optimization. *Math. Model. Anal.* **13**(1), 145–159 (2008)