Local Tuning in Peano Curves-based Global Optimization Scheme

Vladislav V. Sovrasov

State University of Nizhny Novgorod, Nizhny Novgorod, Russia
sovrasov.vlad@gmail.com

Abstract
This paper considers one of the methods to account for the local information on the objective function in Lipschitz global optimization problems. In the course of solving such problems, an issue arises regarding estimating the Lipschitz constant of the objective function arises. According to the classic scheme, this constant is a single estimate for the whole search domain. The method of accounting for local properties is based on building estimates of Lipschitz constants for the search subdomains, and has previously been investigated for the one-dimensional case. For solving multidimensional problems, dimension reduction is applied. A multidimensional optimization problem is reduced to a one-dimensional problem, for which the objective function satisfies the Hölder condition. In this paper, the application of building local Hölder constant estimates in reduced multidimensional optimization problems using the same scheme as is used in one-dimensional Lipschitzian problems is considered.

Keywords: advanced numerical methods, deterministic global optimization, speedup of convergence, derivative-free algorithms

1 Introduction

One-dimensional-characteristical algorithms represent one class of widely-used optimization algorithms [1]. A number of schemes to reduce multidimensional problems to one or more one-dimensional problems, i.e. the *dimension-reduction* schemes have been developed [2], [3]. All these schemes can be efficiently parallelized [4].

In recent years, when building global optimization algorithms, increased attention has been paid to accounting for the local properties of the objective function. In the case of characteristical schemes, this approach allows a mixed algorithm to be developed [5] and different estimates of Lipschitz constants to be used in different search domains [6]. Originally, the latter approach had been proposed for one-dimensional problems, and this has also been examined when carrying out a nested optimization scheme [7]. In all these cases, this approach has proven to be efficient, having essentially accelerated the convergence of the optimization methods. This paper considers generalizing the method for building local estimates of Lipschitz constants in the case of Hölder functions obtained when reducing multidimensional optimization problems.

2 Problem Statement

The global optimization problem can be formulated as follows: to find the global minimum of an N-dimensional function $\varphi(y)$ in a hyperinterval $D = \{y \in R^N : a_i \leqslant x_i \leqslant b_i, 1 \leqslant i \leqslant N\}$:

$$\varphi(y^*) = \min\{\varphi(y) : y \in D\}.$$

In order to obtain an estimate of the global minimum from a finite number of function value computations, $\varphi(y)$ is required to satisfy Lipschitz condition.

$$|\varphi(y_1) - \varphi(y_2)| \leqslant L\|y_1 - y_2\|, y_1, y_2 \in D, 0 < L < \infty$$

The use of the evolvents $y(x)$ i.e. the curves filling the space are a classic dimension-reduction scheme for global optimization algorithms [2].

$$\{y \in R^N : -2^{-1} \leqslant y_i \leqslant 2^{-1}, 1 \leqslant i \leqslant N\} = \{y(x) : 0 \leqslant x \leqslant 1\}$$

Such a mapping allows the reduction of a problem stated in a multidimensional space to solving a one-dimensional problem at the expense of worsening its properties. In particular, the one-dimensional function $\varphi(y(x))$ is not a Lipschitzian but a Hölderian function:

$$|\varphi(y(x_1)) - \varphi(y(x_2))| \leqslant H|x_1 - x_2|^{\frac{1}{N}}, x_1, x_2 \in [0, 1]$$

where the Hölder constant H is related to the Lipschitz constant L by the relation

$$H = 4Ld\sqrt{N}, d = \max\{b_i - a_i : 1 \leqslant i \leqslant N\}$$

Therefore, not limiting the generality, one can consider the minimization of the one-dimensional function $f(x) = \varphi(y(x))$, with $x \in [0, 1]$ satisfying Hölder condition. The issues of numerically building the mapping like a Peano curve and the corresponding theory have been considered in detail in [2]. Here we would note that an evolvent built numerically is an approximation to the theoretical Peano curve with a precision of the order 2^{-m} where m is the building parameter of the evolvent.

3 Description of the Algorithm

The algorithm considered for solving the stated problem implies generating a sequence of points x_k, in which the values of the minimized function $z_k = f(x_k)$ are computed. Let us call the process of computating the function value (including calulating an image $y^k = y(x^k)$) a trial, and the pair (x^k, z^k) — the result of the trial. A set of the pairs $\{(x^k, z^k)\}, 1 \leqslant k \leqslant n$ makes up the search information accumulated by the method after executing n steps.

At the first iteration of the method, a trial is executed at an arbitrary internal point x^1 within the interval $[0, 1]$. Let us assume $n \geqslant 1$ iterations of the method to be completed, during the course of which the iterations in $k = k(n)$ points $x_i, 1 \leqslant i \leqslant k$ have been performed. Then, the point x^{k+1} of the search trial for the next $(k + 1)$th iteration is determined in accordance with the rules:

Step 1. Renumber the points in the set $X_k = \{x^1, \dots, x^k\} \cup \{0\} \cup \{1\}$, which includes the boundary points of the interval $[0, 1]$ as well as the points of preceding trials, by the lower indices in order of increasing coordinate values i.e.

$$0 = x_0 < x_1 < \ldots < x_{k+1} = 1$$

Step 2. Assuming $z_i = f(x_i), 1 \leqslant i \leqslant k$, compute the values

$$\mu = \max_{1 \leqslant i \leqslant k} \frac{|z_i - z_{i-1}|}{\Delta_i}, M = \begin{cases} r\mu, \mu > 0 \\ 1, \mu = 0 \end{cases} \tag{1}$$

where $r > 1$ is a predefined parameter for the method, and $\Delta_i = (x_i - x_{i-1})^{\frac{1}{N}}$.

Step 3. For each interval $(x_{i-1}, x_i), 1 \leqslant i \leqslant k+1$, compute the characteristics according to the formulae

$$R(1) = 2\Delta_1 - 4\frac{z_1}{M}, R(k+1) = 2\Delta_{k+1} - 4\frac{z_k}{M}, \tag{2}$$

$$R(i) = \Delta_i + \frac{(z_i - z_{i-1})^2}{M^2 \Delta_i} - 2\frac{z_i + z_{i-1}}{M}, 1 < i < k+1. \tag{3}$$

Step 4. Select the interval (x_{t-1}, x_t) such that

$$t = \arg\max_{1 \leqslant i \leqslant k+1} R(i), \tag{4}$$

i.e., the interval with the maximal characteristic.

Step 5. Execute a new trial at point x_{k+1} computed according to the formulae

$$x_{k+1} = \frac{x_t + x_{t-1}}{2}, t = 1, t = k+1,$$

$$x_{k+1} = \frac{x_t + x_{t-1}}{2} - \mathrm{sign}(z_t - z_{t-1}) \frac{1}{2r} \left[\frac{|z_t - z_{t-1}|}{\mu} \right]^N, 1 < t < k+1. \tag{5}$$

The algorithm is terminated if the condition $\Delta_t \leqslant \varepsilon$ is fulfilled; here $\varepsilon > 0$ is the predefined accuracy. As an estimate of the global optimum solution of the problem the values

$$f_k^* = \min_{1 \leqslant i \leqslant k} f(x_i), x_k^* = \arg\min_{1 \leqslant i \leqslant k} f(x_i) \tag{6}$$

are selected. The theoretical substantiation of this method is presented in [8], chapter 8.

4 Local Adaptive Estimate of Hölder Constant

As it is seen from the scheme of the algorithm, regardless of the local properties of the optimized one-dimensional function, the same estimate of the Hölder constant (1) is used to compute the characteristic of all intervals (2), (3). In [6], it has been proposed to use different values of M (M is the same as in (1), (2)) for each interval, and also the efficiency of this approach has been shown in the case of optimizing one-dimensional functions that satisfy the Lipschitz condition. In [7], the application of the adaptive estimates of Lipschitz constants in the multidimensional nested optimization scheme has been considered. For each interval, the local estimate of the constant is an additive convolution of the "global" and "local" constants (γ and λ, correspondingly):

$$\lambda_i = \max\{H_{i-1}, H_i, H_{i+1}\}$$
$$H_i = \frac{|z_i - z_{i-1}|}{\Delta_i}$$
$$H^k = \max\{H_i : i = 2, \dots, k\}$$
$$\gamma_i = H^k \frac{\Delta_i}{\Delta^{max}}$$
$$\Delta^{max} = \max\{\Delta_i : i = 2, \dots, k\}$$

$$M_i = r \cdot \max\{H_i, \frac{1}{2}(\lambda_i + \gamma_i), \xi\} \tag{7}$$

A small ξ is chosen to prevent the function from being constant over the search interval. This variant of convolution does not depend on the parameter r, however, the adaptive convolution:

$$M_i = r \cdot \max\{H_i, \frac{\lambda_i}{r} + \frac{r-1}{r}\gamma_i, \xi\} \tag{8}$$

has been introduced in [9] and detailed considered (in case of solving one-dimentional Lipschitzian problems) in [10] as well. If it is known a priori that the optimized function has a complex shape with multiple local minima, then the initial value of r is specified to be large enough to result in the dominance of the "global" component γ in the adaptive convolution.

5 Hypothesis about Convergence

In [10], a theorem is given on the convergence of the method in the case of a Lipschitz objective function. However, as a rule, such statements are true for Hölder metrics as well. Therefore, the following hypothesis is presumably true:

Hypothesis 1. *Assume the objective function $f(x)$ to satisfy Hölder condition with finite constant $H > 0$, and let x be a limit point of $\{x_k\}$ generated by the algorithm. Then, the following assertions hold:*

1. *If $x \in (0;1)$, then convergence to x is a bilateral one i.e. there exists two infinite subsequences of $\{x_k\}$ converging to x: one from the left, the other from the right;*

2. *$f(x_k) \geqslant f(x)$ for all trial points $x_k, k \geqslant 1$;*

3. *If there exists another limit point $x^* = x$, then $f(x) = f(x^*)$;*

4. *If the function $f(x)$ has a finite number of local minima in $[0,1]$, then the point x is a local optimum;*

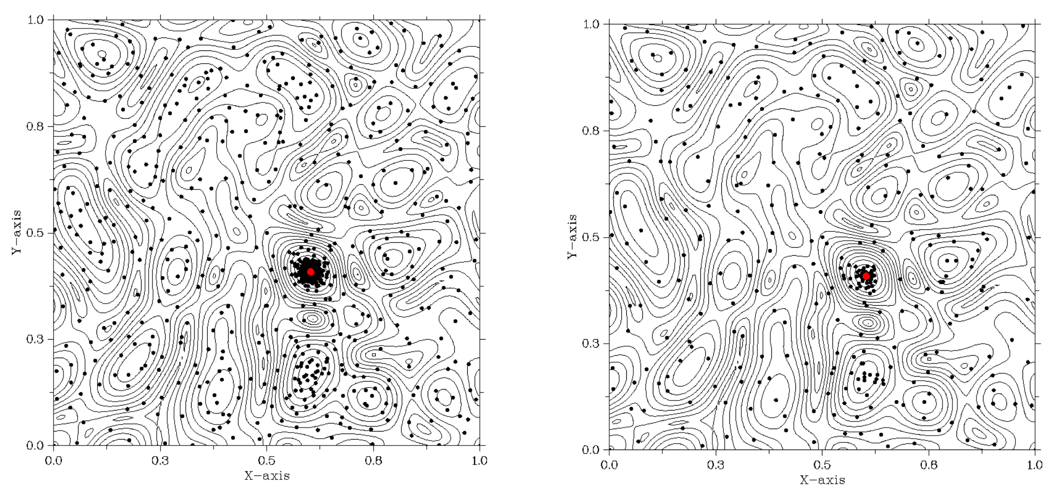
5. *(Sufficient conditions for convergence to a global minimizer). Let x^* be a global minimizer of $f(x)$. If there exists an iteration number k^* such that for all $k > k^*$ the inequality $M_j(k) > H_j(k)$ holds, where $H_j(k)$ is the Hölder constant for the interval $[x_{j(k)-1}, x_{j(k)}]$ containing x^*, and $M_{j(k)}$ is its estimate. Then, the set of limit points for the sequence $\{x_k\}$ coincides with the set of global minimizers for the function $f(x)$.*

The proof of this hypothesis requires further theoretical studies. It was not performed as part of this present work. The convergence has only been established numerically.

6 Experimental Results

The experiments to evaluate the efficiency of the method with a local adaptive estimate of the Hölder constant have been carried out using the two-dimensional classes of Grishagin (F_{GR}) [11] and GKLS problems [12]. Each class includes 100 multiextremal functions. In all experiments, the evolvent was built with the density $m = 12$, the parameter ε in the termination criterion was set to 10^{-3}. The parameter r was selected as low as possible, at which the given method solves all problems of the class. The search step for r was 0.1.

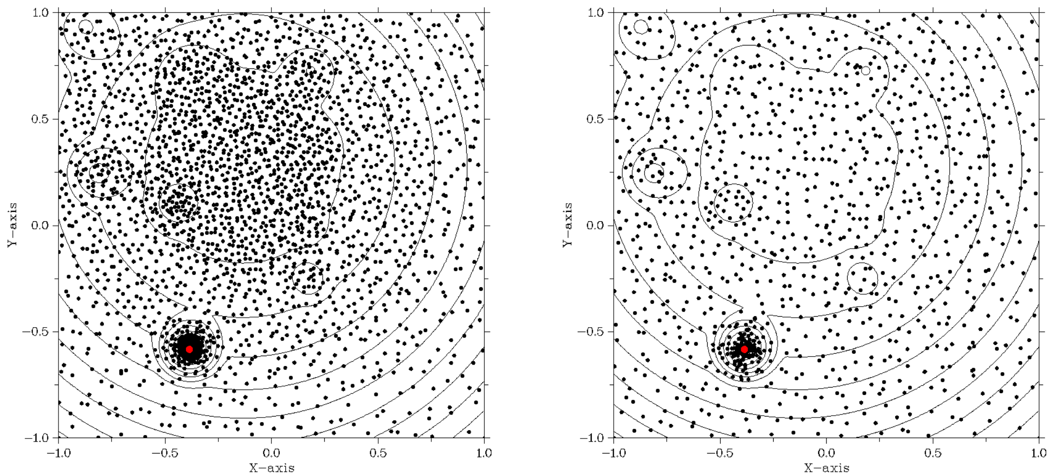
For clarity in illustrating the advantages of the local adaptive scheme for estimating the constant H, let us consider a particular example of the results obtained by the method. In Figure 1, the level lines of a function from the F_{GR} class and the trial points executed by the method with a global estimate of the Hölder constant, and with the one estimated according to formula (7) are shown. As can be seen from the figures, the method using a global estimate of the constant executed a large number of trials in the vicinity of the global minimizer (a total of 1,086 trials were executed) before the termination condition was satisfied, whereas the method with the local adaptive estimate converged much faster (a total of 385 trials were executed). The same situation occurs when optimizing a function from the GKLS class (Fig. 2). The method using the global estimate of the constant executed 2,600 trials while the method using the local adaptive estimate only executed 1,190 trials.



(a) global H estimation (b) local-adaptive H estimation

Figure 1: The level lines of a function from F_{GR} class

Next, let's compare the various alternatives for the method in the problem classes considered. In order to evaluate the efficiency of an algorithm, we will use the operating characteristics [11], which are defined by a set of points on the (K, P) plane where K is the average number of search trials conducted before satisfying the termination condition when minimizing a function from a given class, and P is the proportion of problems solved successfully. If at a given K, the operating characteristic of a method goes higher than one from another method, it means that at fixed search costs, the former method has a greater probability of finding the solution. If some value of P is fixed, and the characteristic of a method goes to the left from that of another method, the former method requires fewer resources to achieve the same reliability.



(a) global H estimation

(b) local-adaptive H estimation

Figure 2: The level lines of a function from GKLS class

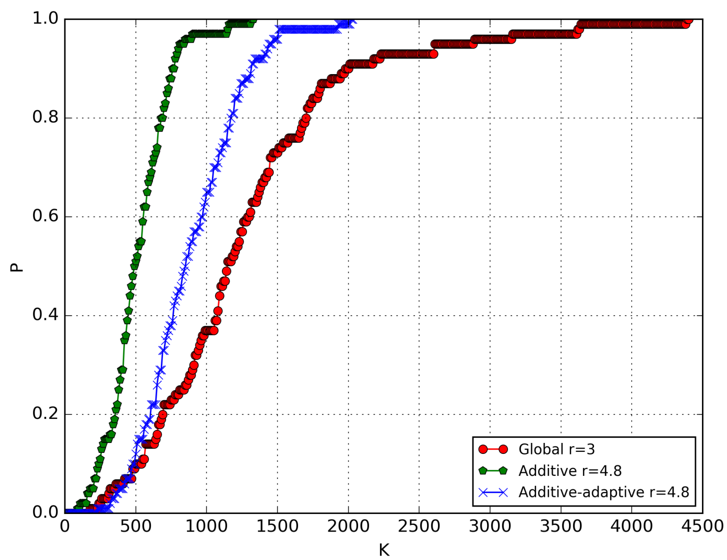


Figure 3: Operating characteristics of the methods compared on F_{GR} problems class

As can be seen from the operating characteristics in Figures 3 and 4, both methods using a local adaptive estimate of the Hölder constant have demonstrated an essential advantage. However, both methods require setting a higher value for the parameter r. If one compares the methods with both the adaptive and non-adaptive convolution (7)(8), the advantage of the latter can clearly be seen, although it requires a higher value for the reliability parameter r to solve problems from the GKLS class.

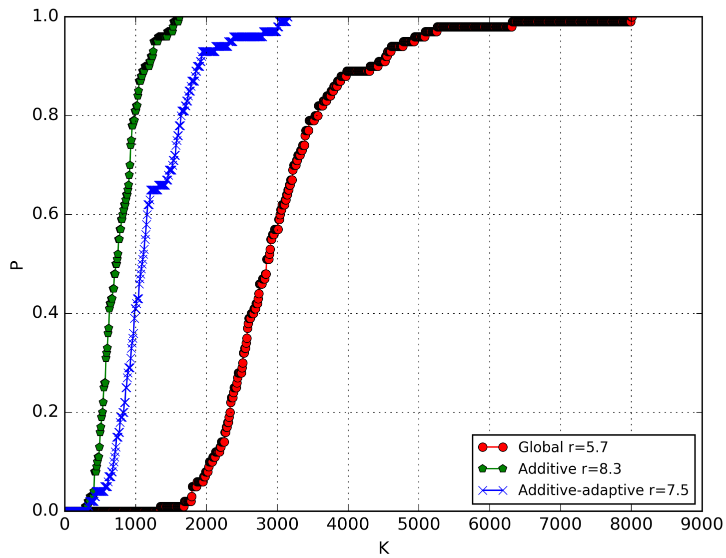


Figure 4: Operating characteristics of the methods compared on GKLS problems class

7 Conclusions

This present work considers global optimization problems and the deterministic methods for solving them. The relevance of solving this class of problems is defined by the importance of the applications (optimal design, problems of parameter identification, pattern recognition, signal processing, etc.). The multiextremal (global) optimization methods are extremely computationally costly, since these require constructing and using adaptive models of the objective function behavior according to the trial results accumulated while executing the algorithm.

This paper considers the application of a method to account for the local behavior of the objective function within the multidimensional multiextremal optimization method. This method has previously been applied to one-dimensional problems only. Taking into account the local properties are implemented in using different estimates of Hölder constant within different search domains. Thus, the process of solving the problem is accelerated significantly by means of reducing the number of search trials.

The comparison of two different schemes for constructing estimates of the Hölder constant (7), (8) has revealed the advantage of scheme (7). However, a larger value for the reliability parameter r in the search method might be required to use this scheme for solving certain problems. When solving complex problems with a small area of attraction for the global extremum, scheme (8) is preferable for the greater reliability of this method. The efficiency of the considered approach has been confirmed by solving the series of multidimensional problems from two classes.

8 Acknowledgements

This research was supported by the Russian Science Foundation, project No 16-11-10150 "Novel efficient methods and software tools for the time consuming decision- making problems with using supercomputers of superior performance."

References

[1] Horst, R., Pardalos, P.M. *Handbook of Global Optimization.* Kluwer Academic Publishers, Dordrecht, 1995.

[2] Sergeyev, Y.D., Strongin, R.G., Lera, D. *Introduction to Global Optimization Exploiting Space-Filling Curves.* Springer Briefs in Optimization, Springer, New York, 2013.

[3] Gergel, A.V., Grishagin, V.A., Strongin, R.G. Development of the parallel adaptive multistep reduction method. *Vestn. Lobachevsky State Univ. Nizhni Novgorod*, 6:216–222, 2013.

[4] Strongin, R.G., Gergel, V.P., Grishagin, V.A., Barkalov, K.A. *Parallel Computing for Global Optimization Problems.* Moscow University Press, Moscow (in Russian), 2013.

[5] Markin, D.L., Strongin, R.G. A method for solving multi-extremal problems with non-convex constraints, that uses a priori information about estimates of the optimum. *Computational Mathematics and Mathematical Physics*, 27(1):33–39, 1987.

[6] Sergeev, Ya.D. An information global optimization algorithm with local tuning. *SIAM Journal on Optimization*, 5(4):390–407, 1995.

[7] Gergel, V., Grishagin, V., Israfilov, R. Local tuning in nested scheme of global optimization. *Procedia Computer Science*, 51:865—-874, 2015.

[8] Strongin R.G., Sergeyev Ya.D. *Global optimization with non-convex constraints. Sequential and parallel algorithms.* Kluwer Academic Publishers, Dordrecht, 2000.

[9] Gergel, A.V. Parallel methods of multiextremal optimization with adaptive descision rules. In *Proceedings of conference of high perfomance computing on cluster systems (HPC-2010)*, pages 152–159, Perm, 2010. PGTU publishers (In Russian).

[10] Sergeyev, Y.D., Mukhametzhanov, M.S., Kvasov, D.E. Derivative-free local tuning and local improvement techniques embedded in the univariate global optimization. *J Optim Theory Appl*, pages 1–23, 2016.

[11] V.A. Grishagin. Operating characteristics of some global search algorithms. *Problems of Statistical Optimization*, 7:198–206, 1978 (In Russian).

[12] Gaviano, M., Kvasov, D.E., Lera, D., Sergeev, Ya.D. Software for generation of classes of test functions with known local and global minima for global optimization. *ACM Transactions on Mathematical Software*, 29(4):469–480, 2003.