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

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

1 Introduction

The one-dimensional characteristic algorithms represent one of the classes of the optimization algorithms used widely [?]. A number of schemes of reduction of the multidimensional problems to one or several one-dimensional ones, i. e. the dimension reduction schemes have been developed [?], [?].

In recent years, when building the global optimization algorithms, an increased attention is paid to accounting for the local properties of the optimized function. In the case of the characteristic schemes, this approach allows developing a mixed algorithm [?] as well as using different estimates of Lipschitz constant in different search domain [?]. Originally, the latter approach has been proposed for the one-dimensional problems, also the investigation of this one when using the nested optimization scheme has been carried out [?]. In all these cases, this approach have demonstrated itself to be an efficient one, having allowed accelerating the convergence of the characteristic optimization methods essentially. In the present paper, a generalization of the method for building of the local estimates of Lipschitz constants onto the case of Hölder functions obtained when reducing the multidimensional optimization problems is considered.

2 Problem Statements

One of the global optimization problem statements can be formulated as follows: to find the global minimum of an N-dimensional function $\varphi(y)$ in a hyperinterval $D = \{y \in R^N : a_i \leqslant x_i \leqslant b_i, 1 \leqslant i \leqslant N\}$. In order to obtain an estimate of the global minimum from a finite number of computations of the function values, $\varphi(y)$ is required to satisfy Lipschitz condition.

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

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

The use of the evolvents i. e. the curves filling the space is a classic scheme for the dimension reduction for the global optimization algorithms [**?**].

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

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

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

where Hölder constant H is related to 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 $f(x) = \varphi(y(x))$, $x \in [0; 1]$ satisfying Hölder condition. The issues of numerical building of the mapping like Peano curve and the corresponding theory have been considered in details in [**?**]. Here we would note that an evolvent built numerically is an approximation to the theoretical Peano curve with the precision of the order 2^{-m} where m is the building parameter of the evolvent.

3 Description of the Algorithm

The considered algorithm of solving the stated problem implies the building of a sequence of points x_k, which the values of the minimized function $z_k = f(x_k)$ are computed in. Let us call the process of computation of the function value (including the building of 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 the search information accumulated by the method after executing n steps.

At the first iteration of the method, a trial is executed in an arbitrary internal point x_1 of the interval $[0, 1]$. Let us assume $n \geqslant 1$ iterations of the method to be completed, in 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 of the next $(k + 1)$th iteration is determined in accordance with the rules:

Step 1. To renumber the points of 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 the order of increasing the values of coordinate i. e.

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

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

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

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

Step 3. For each interval $(x_{i-1}, x_i), 1 \leqslant i \leqslant k+1$, to 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. To execute new trial in the 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 a predefined precision. 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 [**?**].

4 Local Adaptive Estimate of Hölder Constant

At it is seen from the scheme of the algorithm, regardless to the local properties of the optimized one-dimensional function, the same values of estimate of Hölder constant is used to compute the characteristic of all intervals (1, 2, 3). In [**?**], it has been proposed to use different values of M for each interval, and also the efficiency of this approach in the case of the optimization of the one-dimensional functions satisfying Lipschitz condition has been shown. In [**?**], 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» ones (γ and λ, correspondingly):

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

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

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

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

has been considered in [**?**] 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 that results in the dominance of the «global» component γ in the adaptive convolution.

5 Conditions of Convergence

In [**?**], a theorem on the convergence of the method in the case if the objective function is Lipschitz one is given. However, as a rule, such statements are true in Hölder metrics as well, therefore, presumably, the following theorem is true:

Theorem 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 exist 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 one;*

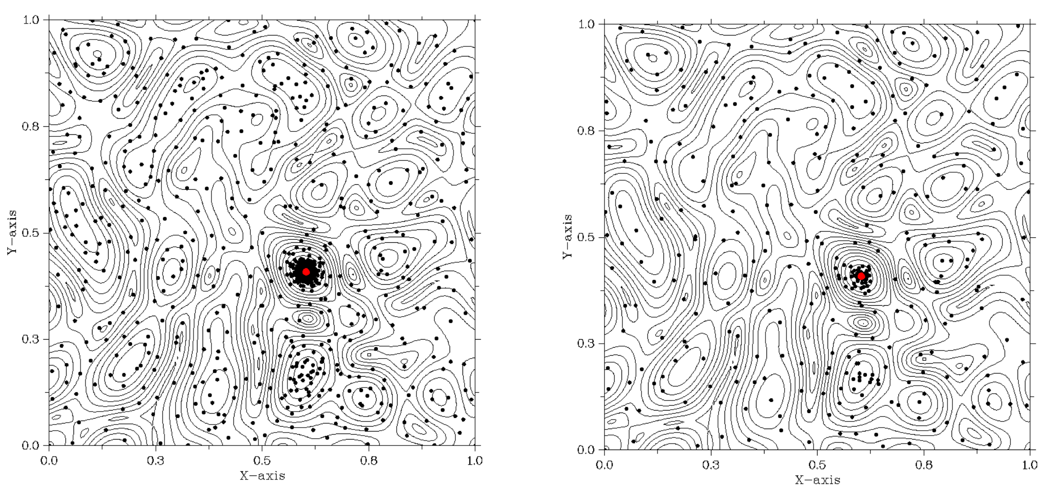
5. *(Sufficient conditions for convergence to a global minimizer). Let x^* be a global minimizer of $f(x)$. If there exists an iteration number k^* such that for all $k > k^*$ the inequality $M_j(k) > H_j(k)$ holds, where $H_j(k)$ is Hölder constant for the interval $[x_{j(k)-1}, x_{j(k)}]$ containing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

The proof of this theorem requires further theoretical studies. Within the frames of the present work, it is not performed. The convergence has been established numerically only.

6 Results of Experiments

The experiments on the evaluation of the efficiency of the method with the local adaptive estimate of Hölder constant have been carried out using the two-dimensional classes of Grishagin problems (F_{GR}) [?] and GKLS Simple 2d ones [?]. Each class includes 100 multiextremal functions. In all experiments, the evolvent was built with the density $m = 12$, parameter ε in the termination criterion was equal 10^{-3}. The parameter r was selected as low as possible, which given method solves all problems of the class at. The search step for r was 0.1.

For clarity of illustration of the advantages of the local adaptive scheme of estimating the constant H, let us consider a particular example of the results obtained by the method. In Fig. 1, the contour plots of a function from F_{GR} class and the trial points executed by the method with the global estimate of Hölder constant and with the one estimated according to formula (7) are shown. As it is seen from the figures, the method with global estimate of the constant executes a large number of trials in the nearness of the global minimizer (total 1086 trials have been executed) before the termination condition is satisfied, whereas the method with the local adaptive estimate converges much faster (total 385 trials have been executed). The same situation takes place when optimizing a function from GKLS Simple 2d class (Fig. 2). The method with the global estimate of the constant had executed 2600 trials while the method with the local adaptive estimate had executed 1190 trials only.

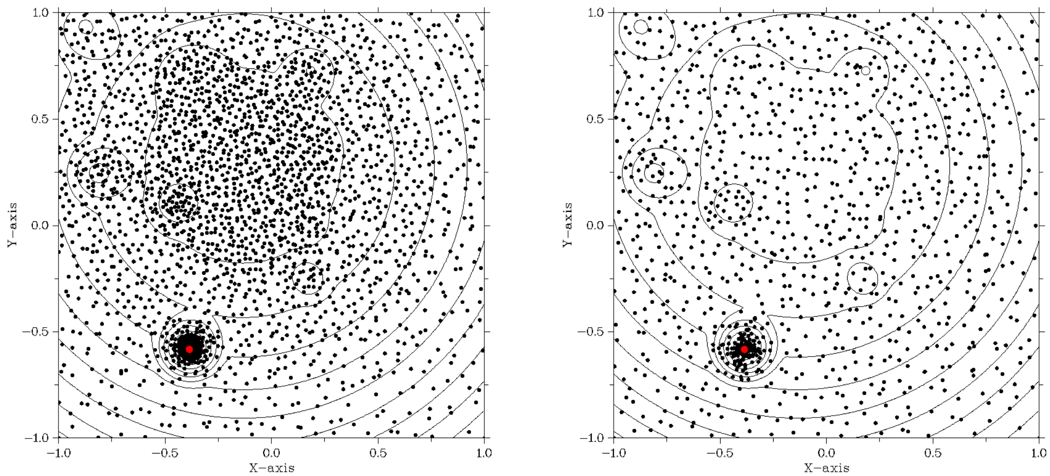


(a) global H estimation

(b) local-adaptive H estimation

Figure 1: The contour plots of a function from F_{GR} class

Next, let us go to the comparison of various variants of the method for the problem classes considered. In order to evaluate the efficiency of an algorithm, we will use the operating characteristics [?], which is defined by a set of points on the (K, P) plane where (K) is the averaged number of the search trials before the satisfaction of the termination condition when minimizing a function from given class, and (P) is the statistical probability of finding the global extremum is found with given precision to the moment of termination. If at given (K) the operating characteristic of a method goes higher than the one of another method, it means that at fixed costs of the search the former method will find the solution with a greater statistical probability. If some value of (P) is fixed, and the characteristic of a method goes to the left



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

Figure 2: The contour plots of a function from GKLS Simple 2d class

from the one of another method, the former method requires fewer resources to achieve the same reliability.

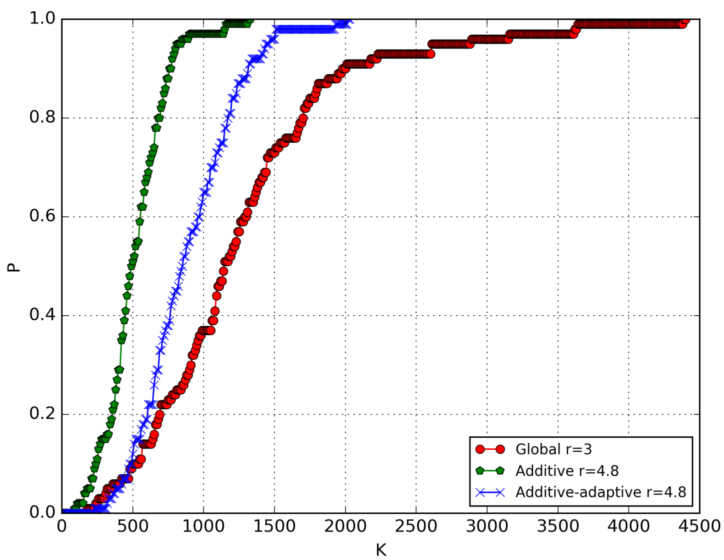


Figure 3: Operating characteristics of the methods compared on F_{GR} problems class

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

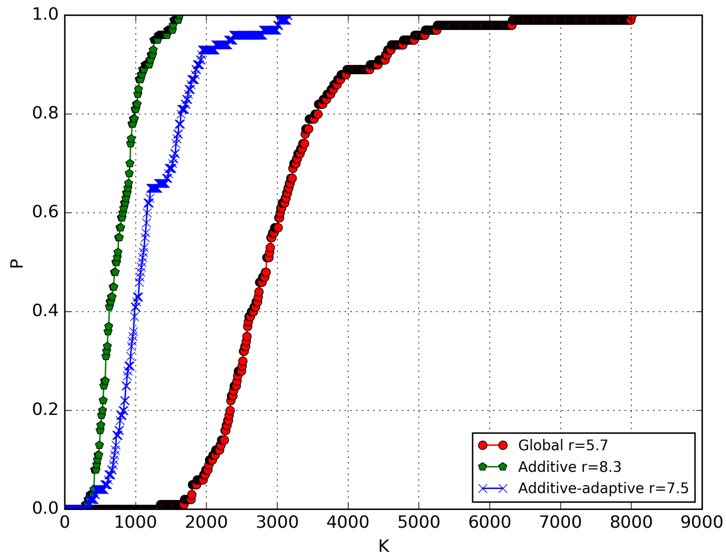


Figure 4: Operating characteristics of the methods compared on GKLS Simple 2d problems class

7 Conclusions

In the present paper, the application of the methods of accounting for the local behavior of the optimized function proposed earlier for the one-dimensional problems in the multidimensional multiextremal optimization methods has been considered. Taking into account the local properties is expressed in the use of different estimates of Hölder constant in different search domains. The efficiency of the considered approach has been confirmed by solving the essentially multiextremal problems form two test 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] Butz, A.R. Space-filling curves and mathematical programming. *Inform. Control*, 12(4):314–330, 1968.

[2] 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.

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

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

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

[6] 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.

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

[8] 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.

[9] R.G. Strongin. *Numerical Methods in Multiextremal Problems (Information-StatisticalAlgorithms).* Moscow: Nauka (In Russian), 1978.

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