Commun Nonlinear Sci Numer Simulat 21 (2015) 99–111

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

journal homepage: www.elsevier.com/locate/cnsns

A deterministic global optimization using smooth diagonal auxiliary functions ☆

Yaroslav D. Sergeyev [a,b], Dmitri E. Kvasov [a,b,*]

[a] DIMES – University of Calabria, Via P. Bucci 42C, 87036 Rende (CS), Italy
[b] Software Department, N.I. Lobachevsky State University, Nizhni Novgorod, Russia

ARTICLE INFO

Article history:
Available online 7 September 2014

Keywords:
Global optimization
Deterministic methods
Lipschitz gradients
Unknown Lipschitz constant

ABSTRACT

In many practical decision-making problems it happens that functions involved in optimization process are black-box with unknown analytical representations and hard to evaluate. In this paper, a global optimization problem is considered where both the goal function $f(x)$ and its gradient $f'(x)$ are black-box functions. It is supposed that $f'(x)$ satisfies the Lipschitz condition over the search hyperinterval with an unknown Lipschitz constant K. A new deterministic 'Divide-the-Best' algorithm based on efficient diagonal partitions and smooth auxiliary functions is proposed in its basic version, its convergence conditions are studied and numerical experiments executed on eight hundred test functions are presented.

1. Introduction

In many important applied problems, some decisions should be made by finding the global optimum of a multiextremal objective function subject to a set of constrains (see, e.g., [9,17,18,31,32,34,36,45,46,50,51,54] and the references given therein). Frequently, especially in engineering applications, the functions involved in optimization process are black-box with unknown analytical representations and hard to evaluate. Such computationally challenging decision-making problems often cannot be solved by traditional optimization techniques based on strong suppositions (such as, e.g., convexity) about the problem.

Because of significant computational costs involved, normally a limited number of functions evaluations are available for a decision-maker (physicist, biologist, economist, engineer, etc.) when he/she optimizes this kind of functions. Hence, the main goal is to construct fast global optimization algorithms that generate acceptable solutions with a relatively small number of functions evaluations.

Usually, the methods applied in this context are subdivided in stochastic and deterministic. Stochastic approaches (see, e.g., [2,9,18,31,32,53,54]) can often work in a simpler manner than the deterministic algorithms. They can be also suitable for the problems where the functions evaluations are corrupted by noise. However, solutions found by some stochastic algorithms (e.g., by popular heuristic methods like evolutionary algorithms, simulated annealing, etc.; see, e.g., [9,11,32,38,52]) can be only local solutions that can be located far away from the global one. This fact can preclude these methods from their usage in practice if an accurate estimate of the global solution is required. Therefore, our attention in the paper is focused on deterministic approaches.

☆ This work was supported by the INdAM–GNCS 2014 Research Project of the Italian National Group for Scientific Computation of the National Institute for Advanced Mathematics "F. Severi".

* Corresponding author at: DIMES – University of Calabria, Via P. Bucci 42C, 87036 Rende (CS), Italy. Fax: +39 0984 494713.
E-mail addresses: yaro@si.dimes.unical.it (Y.D. Sergeyev), kvadim@si.dimes.unical.it (D.E. Kvasov).

http://dx.doi.org/10.1016/j.cnsns.2014.08.026
1007-5704/

Deterministic global optimization is an important applied field (see, e.g., [9,18,36,45,51,55]). As a rule, deterministic methods exhibit (under certain conditions) rigorous global convergence properties and allow one to obtain guaranteed estimations of global solutions. However, deterministic models can still require too many function evaluations to achieve adequately good solutions for the considered global optimization problems.

One of the simplest techniques in this framework is represented by the so-called derivative-free (or direct) approach (see, e.g., [4,5,22,30,37]), often used for solving important applied problems (see, e.g., pattern search methods [22], the DIRECT method [20], the response surface, or surrogate model methods [21], etc.). Unfortunately (see, e.g., [6,26,35,44]), these methods either are designed to detect only stationary points or can demand too high computational effort for their work. As observed, e.g., in [18,49], if no particular assumptions are made on the objective function, any finite number of function evaluations cannot guarantee getting close to the global minimum, since the function may have very deep and narrow peaks.

The Lipschitz-continuity assumption is one of the natural suppositions on the objective function (in fact, in technical systems the energy of change is always limited). The problem involving Lipschitz functions is said to be the Lipschitz global optimization problem (see, e.g., [7,18,27,36,45,46,51,54]).

The global optimization problem with a differentiable objective function having the Lipschitz gradient (with an unknown Lipschitz constant) is an important class of Lipschitz global optimization problems. Formally, this class of problems can be stated as follows:

$$f^* = f(x^*) = \min_{x \in D} f(x), \tag{1}$$

$$\| f'(x') - f'(x'')\| \leqslant K\|x' - x''\|, \quad x', x'' \in D, \quad 0 < K < \infty, \tag{2}$$

where

$$D = [a, b] = \{x \in R^N : a(j) \leqslant x(j) \leqslant b(j)\}. \tag{3}$$

It is assumed here that the objective function $f(x)$ can be black-box, multiextremal, its gradient $f'(x) = \left(\frac{\partial f(x)}{\partial x(1)}, \frac{\partial f(x)}{\partial x(2)}, \ldots, \frac{\partial f(x)}{\partial x(N)}\right)^T$ (which can be itself an expensive black-box vector-function) can be calculated during the search, and $f'(x)$ is Lipschitz-continuous with some unknown constant $K, 0 < K < \infty$, over D. Problem (1)–(3) is frequently met in engineering (see, e.g., [36,45,51]), for instance, in electrical engineering design (see, e.g., [42,51]).

There are known several methods for solving this problem that can be distinguished with respect to the way the Lipschitz constant K from (3) is estimated in their computational schemes. There exist algorithms using an a priori given estimate of K (see, e.g., [1,39]), its adaptive estimates (see, e.g., [13,14,39]), and adaptive estimates of local Lipschitz constants (see, e.g., [39,45]). Recently, methods working with multiple estimates of K chosen from a set of possible values have been also proposed (see [25,27]).

This paper is devoted to developing a new global optimization algorithm for solving problem (1)–(3). The new method adaptively estimates the unknown Lipschitz constant K from (2) during the search and is based on efficient diagonal partitions of the search domain proposed recently. Theoretical background for the introduced method is drawn in Section 2. The algorithm is presented and analyzed theoretically in Section 3. Section 4 presents results of numerical experiments executed on several hundreds of multiextremal test functions. Finally, Section 5 concludes the paper.

2. Theoretical background

In this Section, some important theoretical results, necessary for introducing the new algorithm, are briefly described.

2.1. 'Divide-the-Best' algorithms

As known, many global optimization methods (of both stochastic and deterministic types) have a similar structure. Therefore, several approaches to the development of a general framework for describing global optimization algorithms and providing their convergence conditions in a unified manner have been proposed (see, e.g., [9,16,18,36]). The '*Divide-the-Best*' approach (see [40,45]) is one of such unifying schemes. It generalizes both the schemes of adaptive partition [36] and characteristic [16,45,51] algorithms, which are widely used for constructing and studying global optimization methods.

In a 'Divide-the-Best' algorithm (its generic iteration is represented by the flow chart in Fig. 1), given a vector q of the method parameters, an adaptive partition of the search domain D from (3) into a finite set $\{D_i^k\}$ of robust subsets D_i^k is considered at every iteration k. In Step 1, the Lipschitz constant (K from (2), for the objective function gradient $f'(x)$; or L in the case of the Lipschitz objective function $f(x)$) is estimated in some way. Basing on the previously obtained information X^k, Z^k about the objective function, the 'merit' (called *characteristic*) R_i of each subset (see Step 2 in Fig. 1) is estimated for performing a further, more detailed, investigation (see Steps 3 and 4 in Fig. 1). The best (in a predefined sense) characteristic achieved over a hyperinterval D_t^k corresponds to a higher possibility to determine the global minimum point within D_t^k (see Step 3). The hyperinterval D_t^k is then partitioned at the next iteration of the algorithm. More than one 'promising' hyperintervals can be subdivided at each iteration.