INFORMATICA, 1992, Vol.3, No.2, 275–279

ON CONVERGENCE OF ALGORITHMS FOR BROAD CLASSES OF OBJECTIVE FUNCTIONS

Antanas ŽILINSKAS

Institute of Mathematics and Informatics and
Vilnius Pedagogic Institute
2600 Vilnius, Akademijos St.4, Lithuania

Abstract. In some recent papers a discussion on global minimization algorithms for a broad class of functions was started. An idea is presented here why such a case is different from a case of Lipshitzian functions in respect with the convergence and why for a broad class of functions an algorithm converges to global minimum of an objective function iff it generates an everywhere dense sequence of trial points.

Key words: global optimization, convergence, Lipshitzian functions, adaptation.

1. Introduction. In Pinter (1983) the convergence of global optimization methods is considered for a class of continues functions and in Pinter (1986a, 1986b) for $\mathcal{F}^l = \cup_{L>0} F_L$ where F_L is a class of Lipshitzian function with Lipshitz constant L. It is stated there that the algorithms satisfying some requirements (axioms) generate the sequences of trial points which only limit points are the points of global minima of the objective functions $f(x)$, $x \in A \subset R^n$.

In the notes of Žilinskas (1989a, 1989b) it is shown that for such broad classes of functions as $\mathcal{F}^l$ or a class of continues functions the method converges to global minimum iff it generates an everywhere dense sequnce of trial points x_i. In Pinter (1991) this statement is interpreted incorrectly. Some correct examples are presented there considering class F_L but not $\mathcal{F}^l$. No critics towards the type of convergence in case of minimization of Lipshitzian functions with known constant was in the notes of Žilinskas (1989a, 1989b) which are formulated as a mathematical theorem without an interpretation. Since type of convergence is one of the most important problems in global optimization theory it seems neccessary to present to the comunity of global optimization the main arguments of Žilinskas (1989a, 1989b) to avoid their incorrect interpretation.

2. Paradigm of designing the numerical algoritms. The numerical algorithms (integration, search for a zero, minimization etc.) normally are designed for a class of functions $\mathcal{F}$ which may be defined qualitatively (e.g., convex, quadratic etc.) or qualitatively and quantitatively (e.g., Lipshitzian functions with Lipshitz constant L). Sometimes an algorithm is defined heuristically without specifying a class of functions explicitly, but depending on some parameters which specify a favourable class of functions implicitly. For a given class of functions an optimal algorithm may be defined or complexity of a specific algorithm estimated (see Traub and Wozniakowski, 1980).

In a general case a deterministic algorithm $d(I_{\mathcal{F}})$ is a sequence of functions $d_i(x_j, y_j,\ j = 1, ..., i-1, I_{\mathcal{F}}),\ i = 1, 2, ...,$ where $x_1 = d_1 \in A,\ x_i = d_i(x_j, y_j,\ j = 1, ..., i-1, I_{\mathcal{F}}) \in A$ and $I_{\mathcal{F}}$ is a vector of parameters, depending on information on $\mathcal{F}$. The variable y_j may be substituted by a vector containing not only the function values but also the derivatives at point x_j, but such a generalization is not essential for this consideration. The algorithm $d = (d_1, d_2, ...)$ may be defined, e.g., maximizing an optimality criterion in respect with $\mathcal{F}$ or as mentioned before heuristically without explicit specifying of $\mathcal{F}$.

To apply an algorithm to a specific problem the objective function should be embedded into a class of functions $\mathcal{F}$ choosing a vector $I_{\mathcal{F}}$. We would like to stress *class* here because a generalization of the results which are correct, e.g., for F_L to $\mathcal{F}^l$ may be not trivial. On the other hand the arbitrarily different choice of parameters for different functions makes not impossible an algorithm finding a global minimum in the first iteration.

Some broad classes (continues functions or $\mathcal{F}^l$) are defined qualitatively without parameters. The corresponding algorithms should also be without any parameters. If they are introduced then the (maybe implicit) orientation to a subclass of a considered class is meant there. When $I_{\mathcal{F}}$ is fixed, a deterministic algorithm generates the same sequence of trial points x_i also for different functions $f_1(\cdot)$, $f_2(\cdot)$ for which however $f_1(x_i) = f_2(x_i)$, $i = 1, 2, \ldots$. The idea of Žilinskas (1989a, 1989b) was that if the sequence of the points x_i is not everywhere dense in A then in a broad class of functions, e.g., continues or $\mathcal{F}^l$, there exist different functions with different global minima which coincide at the points x_i. Therefore an algorithm should generate an everywhere dense sequnce of points x_i to guarantee the convergence for every function of the considered broad class.

3. Adaptive methods. In practical problems very often the qualitative information on an objective function corresponds to that used to justify an optimization algorithm, but the quantitative information is not available. For example, the practical continues problems are, as a rule, Lipshitzian, but the Lipshitz constant is unknown, what corresponds to the assumption $f(\cdot) \in \mathcal{F}^l$. There exist two possibilities: to choose a more or less justified constant (and to use an algorithm for F_L) or construct a specific algorithm (theoretically without parameters in this case) for a class $\mathcal{F}^l$. The third possibility is *a constructive adaptation:* to try to estimate L from the data obtained in the course of optimization, i.e., $L_j = L_j((x_i, y_i),\ i = 1, \ldots, j)$. For the statistical models in control theory and optimization such an adaptation is justified by the statistical features of an estimate (Torn and Žilinskas, 1989). We will not discuss here the justification of the estimates of parameters for deterministic models (it does not seem possible if the sequence x_i is not everywhere dense in A) but only mention that in this case x_i depends only on x_j, y_j, $j = 1, \ldots, i-1$. Therefore an adaptive algorithm for a class $\mathcal{F}^l$ generates the same sequence for the different functions which coincide only at x_i.

4. Type of convergence. If the points x_i are dense everywhere in A then continues functions coinciding at x_i are identic. If they are not dense everywhere in A then in a broad class of functions (continues or $\mathcal{F}^l$) it is possible to construct a function which coincides with a given function at x_i but differs from the latter arbitrarily much at the point which is not a limit point of x_i. As shown above the algorithms (also adaptive ones) designed for the broad class of functions use only information on (x_i, y_i) and generate the same sequences x_i for different functions (also with different points of global minima) if they only coincide at x_i. This contradicts to the statement of Pinter (1983, 1986a, 1986b, 1991), that only limit points of x_i are the points of global minimum of an objective function.

The latter type of convergence is characteristic for the case of more narrow than $\mathcal{F}^l$ class of functions, e.g., Lipshitzian (with known constant) one. This case is considered in the specific algorithms used as the examples in papers of Pinter (1983, 1986a, 1986b, 1991). It seems that namely this case (contrary to the formulation) is meant in the cited papers since it is mentioned by Pinter (1991) that for an objective function the corresponding Lipshitz constant $L = L(f)$ is meant. But this means that the Lipshitz constant is known and the usual Lipshitzian case but not a generalization for the case of $\mathcal{F}^l$ is considered. For the class of continues functions or $\mathcal{F}^l$ therefore a *safe* constant for the algorithm could not be choosen. The statement of Žilinskas (1989a, 1989b) on the everywhere dense convergence is valid only for a broad class of functions (continues or $\mathcal{F}^l$) but not for F_L. If the Lipshitz constant may be estimated automatically (e.g., by means of interval arithmetic) then the adaptation in the initial *Lipshitzian* class of functions F_L is possible: to use different smaller than L constants for subsets of A, e.g., Torn and Žilinskas (1989).

Concluding our arguments we would like to repeat that an algorithm converges to the global minimum of every objective function from the broad class (continues or $\mathcal{F}^l$) iff it generates the everywhere dense in A sequence x_i.

REFERENCES

Pintér, J. (1983). *A Unified Approach to Globally Convergent One-dimensional Optimization Algorithms*. Research Report IAMI, Institute of Applied Math-

ematics and Informatics CNR, Milan.
Pintér, J. (1986a). Globally convergent methods for n-dimensional multiextremal optimization. *Optimization*, **17**, 187–202.
Pintér, J. (1986b). Extended univariate algorithms for n-dimensional global optimization. *Computing*, **36**, 91–103.
Pintér, J. (1991). Global convergence revisitied: reply to A.Žilinskas. *Computing*, **46**, 87–91.
Törn, A., and A. Žilinskas (1989). *Global Optimization*. Springer, Berlin.
Traub, J.F., and H. Wozniakovski (1980). *A General Theory of Optimal Algorithms*. Academic Press, New York.
Žilinskas, A. (1989a). Note on Pintér's paper. *Optimization*, **19**, 195.
Žilinskas, A. (1989b). A Note on "Extended univariate algorithms" by J.Pintér. *Computing*, **41**, 275–276.

Received April 1992

A. Žilinskas received the Degree of Candidate of Technical Sciences from the Kaunas Polytechnic Institute, Lithuania, 1973 and the Degree of Doctor of Physical and Mathematical Sciences from the Petersburg University, 1985. He is a senior researcher at the Department of Optimal Decision Theory, Institute of Mathematics and Informatics; Professor and a head of the Department of Informatics at Vilnius Pedagogical Institute.