Journal of Global Optimization **14:** 437–447, 1999.

Stochastic Global Optimization: Problem Classes and Solution Techniques

A. TÖRN[1], M.M. ALI[2] and S. VIITANEN[1]
[1]*Department of Computer Science, Åbo Akademi University, Datacity, 20520 Åbo, Finland (e-mail: atorn@abo.fi);* [2]*Centre for Control Theory and Optimization, Department of Computational & Applied Mathematics, The University of Witwatersrand, Johannesburg, Republic of South Africa*

(Received 12 November 1996; accepted in revised form 6 August 1998)

Abstract. There is a lack of a representative set of test problems for comparing global optimization methods. To remedy this a classification of essentially unconstrained global optimization problems into unimodal, easy, moderately difficult, and difficult problems is proposed. The problem features giving this classification are the chance to miss the region of attraction of the global minimum, embeddedness of the global minimum, and the number of minimizers. The classification of some often used test problems are given and it is recognized that most of them are easy and some even unimodal. Global optimization solution techniques treated are global, local, and adaptive search and their use for tackling different classes of problems is discussed. The problem of fair comparison of methods is then adressed. Further possible components of a general global optimization tool based on the problem classes and solution techniques is presented.

Key words: Global optimization, Problem features, Problem classes, Test problems, Solution techniques

1. Introduction

In this paper we discuss essentially unconstrained global optimization problems, i.e., find $\hat{f}^* = f(\hat{x}^*)$, where $\hat{x}^* \in A \subset R^n$ so that $|\hat{f}^* - f^*| \leqslant \epsilon$, where f^* is the global minimum obtained in the interior of A. The region A is assumed to be either a box or some other region easy to sample. Our only requirement on the problem is that $f(x)$ can be computed for any $x \in A$.

The methods we consider for solving such problems are those containing some probabilistic technique, which explore the search region A by evaluating f for points sampled in A. Based on this minimal requirement, we are able to classify problems into degrees of difficulty. In practice, other information may be available and may of course be utilized in solving the problem at hand, and thus can make a given problem easier. The word minimum in the text that follows normally means the minimum value but sometimes also the pair (minimizer, minimum), and refers to any minimum, local or global.

When presenting a new method authors commonly illustrate the working of their method and compare its performance with that of some other algorithms on some test problems. In many cases the choice of test problems is quite random with the only systematic selection being over different values of n. Many of the test problems often used in the literature are trivially easy to solve and some of them are even unimodal and could thus be solved by applying a local optimization method from a single starting point.

Of course the choice of test problems should be systematic (for constrained problems see (Floudas and Pardalos 1990)) so that they represent different types of problems ranging from easy to difficult to solve. The failure in making such a choice may partly be caused by the lack of a suitable classification of problems according to some complexity measure and partly by the fact that the features of the test problems are not known.

This discussion shows that it would be important to be able to classify global optimization problems in order to test methods more systematically. This can then lead to a characterization of algorithms which is important in order to choose a suitable method given a problem with known features. Also such a classification could be the base for constructing an optimization tool that could characterize the problem at hand and then choose a suitable set of methods to apply.

Choosing a suitable method from a set of methods implies that the methods can be characterized and compared. We will therefore also address the problem of comparing methods.

2. Problem features and solution techniques

We here discuss global optimization problem features and their contribution to problem complexity. We also recognize different techniques that are used in global optimization methods.

2.1. PROBLEM FEATURES

When solving (finding the global minimum of) a global optimization problem the outcome is dependent on the complexity of the problem. We postulate that the complexity is dependent on the following features of the problem:

- the size p^* of the region of attraction of the global minimum,
- the affordable number of function evaluations N_f,
- embedded or isolated global minimum,
- the number of local minimizers.

The region of attraction $S(x_m) \subset A$ of a local minimizer x_m is defined as the largest region containing x_m such that when starting an infinitely small step strictly descending local minimization from any point in $S(x_m)$ then the minimizer x_m will

be found each time. The region of attraction of a minimum m is the union of the regions of attraction of all minimizers x for which $f(x) = m$.

If the region of attraction of f^* is large then this region is easy to detect when sampling in A and such a problem is of course easier to solve than a problem with smaller such region.

The value of the expression $(1-p^*)^{N_f}$ is the chance that the region of attraction of f^* is missed when sampling N_f points at random in A. If the function f is cheap to evaluate then N_f is large and the probability that even a very small region of attraction is missed becomes small. However, if only a small number of function evaluations can be performed then the probability to miss the region of attraction of the global minimum for small p^* is large.

Embedded global minimum means that there exist non-global minimizers near a global minimizer, so that exploration (e.g. sampling) near these leads to detecting better and better minima and eventually the global minimum. A simplified picture of such a case is that these minima are at the surface of a bowl with the global minimum at the bottom. If the global minimum is embedded this means that the region of attraction of f^* may be found by such exploration even if the size of the region of attraction is very small. We say that the global minimum is isolated as the opposite to being embedded.

The number of minimizers and the size of the region of attraction of the global minimum are normally not independent of each other. One would expect that the size is a decreasing function of the number of minimizers. However, it is an important feature on its own because local search will become increasingly ineffective for an increasing number of local minimizers.

Of course there are other features which have an influence like the size of A, the dimensionality n, and unique or several global minimizers. When formulating a problem, A should be specified as small as possible because if on two domains $A \subseteq \bar{A}$, f has the same global minimizer then working with the smaller domain is better. The way A influences the complexity is explained by the following example. Let the smallest A be the box $[0,1]^n$. If instead A equal $[0,2]^n$ is used then the volume of A grows with a factor 2^n. Sampling in the larger box will then increase $1-p^*$ to maximally $1-p^*/2^n$ so this is reflected in the chance to miss the region of attraction of f^*. The increase is maximal if the region of attraction is in the interior of $[0,1]^n$. Larger n may mean smaller p^* either because of more minima or larger A. A problem with several global minimizers is of course harder if all minimizers are to be found and may also affect efficiency because of convergence problems for some methods.

2.2. SOLUTION TECHNIQUES

Global techniques, local techniques, and adaptive techniques are all used in global optimization.

Table 1. Problem classes and solution techniques

		Problem features			Solution techniques		
Class	Complexity	$(1-p^*)^{N_f}$	Embeddedness	#mins	glob	local	adapt
U	unimodal	0	none	1	(+)	++ld	+
E_1	easy	small	any	few	+	+ld	+
E_2		small	any	many	+	+gd	+
M_1	moderate	large	embedded	few	+	++ld	+
M_2		large	embedded	many	+	++gd	+
D_1	difficult	large	isolated	few	++	+ld	–
D_2		large	isolated	many	++	+gd	–

The global technique is responsible for exploring the whole area of interest A and in this way ensures that a point in the region of attraction of the global minimum is found.

The local technique is used to find better points in the vicinity of some (good) point in order to improve the accuracy of a solution. In the case of an embedded global minimum the local technique could also be responsible for finding better minima near promising minima.

The effect of an explicit local technique can also be achieved in another way. By an adaptive technique we mean that the global technique is gradually sampling more points in the regions where good points have already been found. This technique is used in methods where no explicit local technique is used so that the global technique gradually turns into a local technique. Another motive for using adaptation is based on the assumption that the global minimum is embedded. If this is the case then it is rewarding to focus the search on neighborhoods of promising points.

Other techniques used are single working point techniques (like Simulated Annealing, SA (Dekkers and Aarts 1991)), working set techniques (Controlled Random Search, CRS (Price 1987)), multistart, and clustering, just to mention a few. These techniques are not generally applied but rather relate to some specific methods.

3. Problem classes and solution techniques

In Table 1 we present the relation between problems and solution techniques. There are seven classes presented in increasing order of complexity.

In the technique part of the table + means that the technique is usable, ++ means that the main effort should be in using this technique. In the column for

Table 2. Used test problems by classes

Class	Function	$(1-p^*)^1$	#mins	Reference
U	Kowalik	0.00	1	(Jansson and Knüppel 1995)
	Powell	0.00	1	(Jansson and Knüppel 1995)
E_1	Branin	0.00	3	(Törn and Žilinskas 1989)
	Goldprice	0.55	4	(Törn and Žilinskas 1989)
	Shekel5	0.65	4	(Törn and Žilinskas 1989)
	Shekel7	0.65	7	(Törn and Žilinskas 1989)
	Shekel10	0.65	10	(Törn and Žilinskas 1989)
	Hartman3	0.30	4	(Törn and Žilinskas 1989)
	Hartman6	0.30	4	(Törn and Žilinskas 1989)
	Hosaki	0.35	2	(Bekey and Ung 1974)
	Si(B): P3	0.66	3 ?	(Ali, Storey and Törn 1997)
E_2	Levy10	0.15	10^{10}	(Jansson and Knüppel 1995)
	Shubert3	0.65	$\approx 5^3$	(Dekkers and Aarts 1991)
	Si(B): P4	0.89	∞ ?	(Ali, Storey and Törn 1997)
	Si(B): P5	0.95	∞ ?	(Ali, Storey and Törn 1997)
M_2	Griewank2	0.99	≈ 500	(Törn and Žilinskas 1989)
	Griewank10	0.93	$O(10^3)$	(Törn and Žilinskas 1989)
	Shubert5	0.95	$\approx 15^5$	(Dekkers and Aarts 1991)
D_2	Si(B): P6	.997	∞ ?	(Ali, Storey and Törn 1997)

the local technique the notation *ld* means local descent and the notation *gd* means global descent, i.e., local improvement for instance by sampling so that the local technique may escape inferior minima.

For difficult problems where the global minimum is isolated, applying an adaptive technique (concentration of the search in promising regions which have been found) will in general increase the chance of missing the global minimum. This occurs because new regions should continue to be explored; this fails to happen when we use adaption.

Table 2 shows some examples of problem members of each class. For each problem are given the number of minimizers, #mins, and the relative size of the region of attraction of the global minimum, p^* (actually $(1-p^*)^1$, the probability to miss the region of attraction of the global minimum by sampling one point at random in A). The values in the column $(1-p^*)^1$ have been obtained by applying a local descent algorithm to 1000 points randomly distributed over A. For Si(B) and Griewank 10000 points were used.

It is not possible in general to specify N_f because it depends on the environment in which the optimization takes place. Attributes of the environment, important in this respect, are computing power and possible time constraints. However, for $1 - p^* \leqslant 0.95$ sampling just 100 points at random in A makes the chance to miss the global minimum less than 0.0006, so for most problems above, $(1 - p^*)^{N_f}$ (the chance to miss), will be rather small in any application.

The characterization of the sample problems in Table 2 is partly based on experiences made in solving global optimization problems with CRS techniques (Ali, Törn and Viitanen 1997) and some other stochastic methods (Ali, Storey and Törn 1997). Because N_f will increase as the computers become more powerful, problems may switch to lower complexity classes with time.

3.1. UNIMODAL PROBLEMS

Unimodal problems are local optimization problems and therefore actually no global technique needs to be applied. The (+) in the column for the global technique indicates the possibility to use several starting points in order to guarantee that the local optimization method really converges to the global minimum. Problems used as global optimization test problems belonging to this class are Powell, which analytically can be proved to have only one minimizer, and Kowalik, which in all our experiments was found to have a single minimizer. Possibly the unimodal feature of these problems was not known when they were used. The only reason for using local optimization problems as test problems for global optimization methods is to find out how they manage to solve such problems. This can be important as it is not generally known if the problem is multimodal or not.

3.2. EASY PROBLEMS

The easy problems are characterized by a small chance to miss the region of attraction of the global minimum. This means that either the region of attraction is large or that enough points can be sampled in order for the chance to miss the region of attraction to be small. For large regions of attraction the strategy could be to sample a small number of global points uniformly in A and then to start local optimizations from some promising points. Most global optimization methods should work. For few minimizers methods designed to find all local minimizers (e.g. clustering techniques) could be used. There are many test problems belonging to this class, the best known are the "standard test problems" Branin, Goldstein-Price, Shekel 5,7,10 and Hartman 3,6. The problems of Hosaki and Si(B): P3 also belong to this class.

For problems with many minimizers global descent rather than local descent should be used. Examples of problems belonging to this class are Levy10 and Shubert3. Also for smaller regions of attraction of the global minimum more effort has to be devoted to the global part.

3.3. MODERATELY DIFFICULT PROBLEMS

There seems to be a lack of problems in the literature belonging to the class M_1, i.e., problems with embedded global minimum and few minima, with small region of attraction of the global minimum. We therefore leave out the class M_1 from Table 2.

Moderately difficult problems are characterized by an embedded global minimum and a high probability of missing the region of attraction of the global minimum (either because the region is very small or because the function is so expensive to evaluate that the number of affordable function evaluations is small). Embedded means that given a promising point (a point with relatively small function value) in the region of attraction of some local minimum, there will be regions of attraction of even better local minima and of the global minimum nearby. This means that an adaptive technique works well. The CRS method CRS(q, β) which uses quadratic approximation for generating trial points accompanied by sampling from the β-distribution for local improvement (Ali, Törn and Viitanen 1997) could be recommended.

The Griewank problems belong to this class. They have many local minimizers with the global minimum in the origin and many other nearby. In our experiments we found that the 2-dimensional problem is more difficult than the 10-dimensional. This was not expected. We also note that the problems listed could very well be classified as belonging to the class E_2 because a very large number of function evaluations can be performed on the powerful computers of today.

3.4. DIFFICULT PROBLEMS

We have not found problems in the literature belonging to the class D_1, i.e., problems with isolated global minimum and few minima, with small region of attraction of the global minimum. We have therefore left out the class D_1 from Table 2.

The difficult problems are characterized by a large chance to miss and isolated global minimum. This means that the detection of a point in the region of attraction of the global minimum must totally rely on sampling in A. There is no reward in using an adaptive technique for the global part because the global minimum is isolated and thus adaptive sampling will increase the chance to miss the global minimum. The number of minimizers could either be small or large and this will affect which local technique is to be used. For a small number of minimizers it will be effective to use a local descent method when a promising point is found by the global technique. Here clustering techniques or other techniques which try to find all local minimizers could be applied.

Many of these problems are probably unsolvable with affordable effort and one can only hope to find a good local solution. Examples of difficult problems are Many-body Problems (Ali, Storey and Törn 1997). There are three similar problems described in (Ali, Storey and Törn 1997). We give one of them, Si(B), as an example here. These become increasingly difficult with increasing n because of

the increasing number of minimizers and growing $1 - p^*$. Some of these ($P3$, $P4$) are solvable by using some thousands of function evaluations, and could be characterized as rather easy, others ($P6$, $P7$, ...) require several more times of function evaluations for a serious solution effort and can be characterized as difficult.

4. Comparing methods

For comparing methods we propose that test problems should represent different classes of problems, for instance Branin, Shekel5, Shubert3, Griewank, and Many-body problems or other problems from class D.

For a method the outcome of a solution effort (the quality of the solution) is dependent on the stopping condition used because this will determine how much work is used to find the solution. We therefore touch upon the stopping problem below.

Comparing methods is complicated by the fact that two goals are to be achieved: high quality and small effort. How to compare methods properly in this two dimensional goal space will also be discussed.

4.1. STOPPING CONDITIONS AND CONVERGENCE

Every method must use some stopping condition. Stopping conditions are sometimes based on theoretical convergence properties. Because a global optimization problem generally cannot be solved for sure the convergence properties are at best probabilistic which means that some method will find the global minimum with a probability that approaches one as the algorithm runs on. Finite convergence is dependent on that a point in the region of attraction of the global minimum is found in a finite number of steps and that a local algorithm that successfully finds the minimum is started from such a point. Even if a method converges in probability it is not normally possible to estimate the probability that the global minimum is found when the algorithm stops after a finite number of function evaluations. An exception is multistart if we assume that the local technique is always successful and we know p^* or a lower bound of it (Törn and Žilinskas 1989).

It is trivial to modify any method that can be made to run indefinitely to converge with probability one by adding some random sampling element that is always applied but possibly with a probability decreasing over time in order not to degrade the efficiency of the method. As an example consider the CRS methods which do not have any theoretical convergence properties. Add the following to the CRS procedure: "Each time the number of function evaluations reaches N_0 (let initially N_0 be e.g. $10N_f$), sample a point at random in the search region, update the working set, and double N_0". For parameter values making CRS to run forever this modified version will converge to the global minimum with probability one. However, the modified CRS method and the original CRS method are identical in any application.

This means that knowing that a method has theoretical convergence properties in a probabilistic sense is not enough from a practical point of view. Some other additional measures are needed. One such measure could be the number of uniformly distributed points sampled in A during the solution process. From this the chance to miss the region of attraction of the global minimum, under some assumption on the lower bound of its size, could then be calculated.

4.2. EMPIRICAL COMPARISON

Methods are normally compared on their efforts needed to find the global minimum (e.g. number of function evaluations N_f or cpu time). A method is considered better than another method if it solves the problem with fewer function evaluations or in less cpu time.

When making empirical comparisons between methods it is important to realize that a probabilistic global optimization method M applied to a problem P is a mapping $(M, P) \rightarrow (E, q)$, where E is the effort applied in the solution process and $q \in [0, 1]$ is the probability that the global minimum is found. The effort could either be measured as cpu time and/or as the number of function evaluations N_f. This means that when comparing two methods M_1 and M_2 for a problem P the pairs (E_1, q_1) and (E_2, q_2) are to be compared. If then either one pair dominates the other or $q_1 = q_2$ or $E_1 = E_2$ such a comparison is possible otherwise not. The pair (E_1, q_1) dominates (E_2, q_2) if they are not equal and $E_1 \leqslant E_2$ and $q_1 \geqslant q_2$. In order to compare two methods the easiest way is thus to fix E and apply the methods repeatedly in solving the same problem, calculating the averages for q_1 and q_2 and comparing these. This should then be done for a range of E. Alternatively one could fix q and try to find such parameters for the methods that q is achieved on the average and then compare the efforts E. Whichever of these approaches is chosen a lot of experimenting is needed in order to find parameter values that makes the methods comparable.

In conclusion we think that comparisons of methods reported in the literature are normally not fair in the way explained above. Some are rather heuristic based on incompatible stopping conditions giving an outcome which as a rule seems to be in favor of the new method.

5. Elements of a global optimization tool

Given a global optimization problem the features of the problem are usually not known. What strategy should then be applied to solve the problem? We think that a characterization of the problem is part of the solution and therefore the solution strategy should be to apply methods that would reveal the features of the problem. We next address exploring the features of a given problem, see Table 1.

The method to use for finding #mins is multistart, i.e., starting a local minimization algorithm from N randomly sampled starting points in A. If most of the

minimizations arrive at different solutions then many minimizers can be expected, otherwise few. Other information obtained is the variation of $f(x)$ over A and the effort needed (function evaluations) for local minimization. Further the affordable number of function evaluations in solving the problem can be estimated. We can of course not state anything about the embeddedness of the global minimum. By calculating the distances from the best minimizer to nearby minimizers some indication of the embeddedness of the best minimum found can be obtained.

The feature p^* can of course not be estimated because this means that we know for sure we have solved the problem which might not normally be possible to assure. One possible way to proceed is to make some assumption about the lower limit of p^*. Such an assumption could be based on practical information, i.e., the usefulness of a solution might be dependent on whether the solution is stable. By stable we mean that variations in the decision variable x inside the tolerance lead to small variations in the function values.

Solving an unknown global optimization problem in practice is of course not an one shot effort with efficiency in the driving seat. Rather a lot of experimenting is undertaken until confidence in the solution (within affordable effort) is obtained. We are here advocating for some systematical approach in this experimenting. Our recommendation is:

1. Sample N points uniformly in A and evaluate f giving $P_i = (x_i, f_i)$, $i \in \{1, N\}$. Record $f_{\min}$, $f_{\max}$.
2. Start local minimization from each point and record the minima, the relative sizes of their regions of attraction, their minimizers, and the embeddedness of the best minimum.
3. Print a characterization report on the collected information.
4. Based on analysis of the report choose a suitable global optimization method (methods) and tailor to meet characteristics by specifying parameters and make a run. Based on results adapt and run again until satisfied.

The methods to include for solving global optimization problems should allow the user to adapt the mode of exploration to what he knows or what he learns about the problem features, e.g. switch between local and global, and utilize adaptive techniques if deemed favorable. It is not possible to here give a definite recommendation on which methods to use, we just mention one which has proved usable.

A possible candidate to include in such a pool could be a recent version of Price's Controlled Random Search method, CRS(q, β) (Ali, Törn and Viitanen 1997). This method does not use a local optimizer, so it can be used even if there are many minimizers and will help in finding embedded global minima. The method, as all CRS methods, also has a very natural and easy to understand stopping condition. The method was found competitive in a comparison of several global optimization methods (Ali, Storey and Törn 1997).

Based on the ideas presented above, it should be possible to construct a global optimization tool that would assist in obtaining a solution by offering several global

optimization methods and also an overall strategy in solving global optimization problems.

References

Ali, M., Storey, C. and Törn, A. (1997), 'Application of some Recent Stochastic Global Optimization Algorithms to Practical Problems', in *Journal of Optimization Theory and Applications* 95, No. 3, December 1997, pp. 545–563.

Ali, M., Törn, A. and Viitanen, S. (1997), 'A Numerical Comparison of Some Modified Controlled Random Search Methods', *Journal of Global Optimization* 11, pp. 377–385.

Bekey, G.A. and Ung, M.T. (1974) 'A Comparative Evaluation of Two Global Search Algorithms', *IEEE Trans. Syst. Man, Cybern.*, SMC-4, No. 1, pp. 112–116.

Černy, V. (1984), 'Minimization of Continuous Functions by Simulated Annealing', Research Institute for Theoretical Physics, University of Helsinki, Preprint No. HU-TFT-84-51.

Dekker, A. and Aarts, E. (1991), 'Global Optimization and Simulated Annealing', *Mathematical Programming* 50, pp. 367–393.

Floudas, C.A. and Pardalos, P.M. (1990), 'A Collection of Test Problems for Constrained Global Optimization Algorithmst', Lecture Notes in Computer Science 455, Springer-Verlag.

Jansson, C. and Knüppel, O. (1995), 'A Branch and Bound Algorithm for Bound Constrained Optimization', *Journal of Global Optimization* 7, pp. 297–331.

Price, W.L. (1987), 'Global Optimization Algorithm for a CAD Workstation', *Journal of Optimization Theory and Applications* 55, pp. 133–146.

Törn, A. and Žilinskas, A. (1989), *Global Optimization*, Lecture Notes in Computer Science 350, Springer-Verlag, 255 p.