Papers, Volume 1 1962-1985

Thoft-Christensen, Palle

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

CHAPTER 5

RANDOM SEARCH OPTIMIZATION: STOPPING RULES[1]

P. Thoft-Christensen, Aalborg University, Denmark

INTRODUCTION

During the last twenty years or so the theory of non-linear optimization has undergone a development making it more usable for engineers, scientists etc. The interest in application of modern optimization techniques is considerable, primarily as a result of the development of modern computers. Only very simple optimization problems can be solved by classical methods without rather extensive numerical calculations.

Most optimization techniques to day are of the so-called iterative type that is the optimum point is approached by a step-by-step method. As usual, when using a numerical method, the intention is to get an approximate solution and if possible to evaluate the degree of approximation. In almost all methods used it is necessary to specify a way to terminate the iterative process. It turns out that the stopping rule or termination criterion is of great importance for the effectiveness of an iterative method, so it is strange that stopping rules seems to be a neglected research area. In particular for multimodal functions where more than one optimum point exists a good choice of termination criterion is essential.

With the non-linear optimization techniques available now unimodal problems can be handled with a satisfactory result in most cases although it can be difficult for a user to choose between the great numbers of methods. Only in some very special cases problems arises.

But in the case of multimodal functions serious problems are obvious and only some of these problems are partly solved. Further it is often impossible for a user of optimization theory to make out whether his objective function is unimodal or multimodal.

In this paper stopping rules will be discussed with reference to a recently presented very simple optimization strategy based on random search.

---

[1] International Conference on "Mathematical Models for Environmental Problems", Southampton, England, September 1975.

OPTIMIZATION TECHNIQUES

In this paper the discussions will for the sake of simplicity be restricted to unconstrained parameter optimization problems that is the problem of finding minimum (or maximum) for the so-called *criterion function*

$$F:\ R^n \rightarrow R \tag{1}$$

The ordered set of n parameters $\bar{x}_0 \in R^n$ is called a *global minimum* for the function *F*, if

$$F(\bar{x}) \geq F(\bar{x}_0) \ \text{ for all } \ \bar{x} \in R^n \tag{2}$$

If there exists a neighbourhood $\omega(\bar{x}_0)$ of the point $\bar{x}_0$ so that

$$F(\bar{x}) \geq F(\bar{x}_0) \ \text{ for all } \ \bar{x} \in \omega(\bar{x}_0) \tag{3}$$

then $\bar{x}_0$ is called a *local minimum* for the function *F*.

It is outside the intention with this paper to give a review of methods of optimization. Excellent presentations and discussions of widely used methods are given in the books by Lavi & Vogl [1], Lootsma [2], Jacoby, Kowalik & Pizzo [3], and in many others. As mentioned earlier most methods have one very serious drawback, namely that they only work satisfactory if the criterion function F is unimodal. If the starting point for an iterative method is outside of the field of influence of the global optimum for a multimodal function there exists a risk that methods based on gradient techniques only will disclose a local optimum. As the main objective by most methods is to arrive in the nearest minimum as fast as possible the iteration has to be initiated from a number of different starting points if the criterion function is multimodal. Few research workers have only attacked this local-global problem. It is usually believed that random search methods in this connection have some potential not yet fully investigated. The strategy by Thoft-Christensen & Hartmann [4] is a creeping random search method based on a combination of local and global search technique. The method is compared with other random search methods in a paper by Thoft-Christensen [5].

The basic principle in creeping random search is very simple. A starting point $\bar{x}_0$ is chosen and a new point

$$\bar{x}_1 = \bar{x}_0 + \Delta\bar{x}_0$$

is determined in a way depending of the strategy used. If $F(\bar{x}_1) < F(\bar{x}_0)$ the step $\Delta\bar{x}_0$ is called a *success* and the point $\bar{x}_1$ is used as starting point for the next step, etc. If $F(\bar{x}_1) \geq F(\bar{x}_0)$ the step $\Delta\bar{x}_0$ is called a *failure* and it is replaced by a new step until a success takes place. In the strategy by Thoft-Christensen & Hartmann [4] two types of steps are used, namely local and global steps. The intention with the local steps is to obtain a high degree of convergence to the nearest minimum not depending on whether it is a global or a local minimum. So the step size for a local step is "rather" small. On the contrary the global steps are not limited in size as their purpose only is to escape from a local minin1um by making sufficiently large steps. The global steps must be arranged so that any point in the feasible region can be reached.

The local part of the algorithm is arranged in the following way. First a starting point is chosen random. Then some trial steps are taken with different step-sizes and

with a directional adaptation so that the direction of the last successful step is given a preference. When a prescribed number of successful steps are carried out a global trial step follows. Here a stochastic local search has been employed but almost any of the non-random methods mentioned above can be used. The choice of termination criterion for this local part will be discussed later.

As the intention with the global steps is to be able to reach any point in the feasible region no information concerning a "good" direction can be given in general. Therefore the step size and the direction for a global step are chosen random. In this way a prescribed number of global steps are made until an improvement may happen. In such a case the new point is used as starting point for a new set of local steps. Otherwise further local steps are made from the same point but now with new directions and step sizes. The procedure described above continues until a pre-described number of successful steps have occurred or until no further improvements by local steps can be found within the stopping rule used.

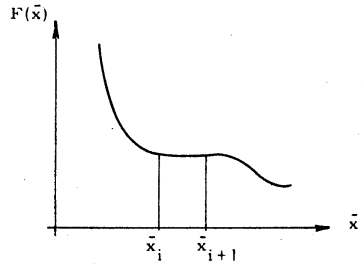
The algorithm sketched above has with some success been used on several test functions with a considerable number of parameters. One of the conclusions based on these experiments is that the choice of stopping rules for both the local and the global part is a crucial one. It was therefore decided to look closer at the problems connected with stopping rules. In the next sections a discussion of stopping rules will be given in relation to the algorithm above. The results are only preliminary, as the investigation is not closed. It is therefore not possible to make final conclusive statements concerning these important questions.

STOPPING RULES

It is emphasized by several authors that slight changes in the termination criteria make a big difference in the performance of an algorithm, see e. g. Himmelblau [6]. In the paper by Himmelblau [6] it is also emphasized that more than one stopping rule usually must be used to take care of specific characteristics by some functions. Below some very often-used stopping rules are presented and some of their advantages and disadvantages are discussed.

Stopping rules based on function values $F(\bar{x})$:

In the literature one will often find that the fractional change in the function values $F(\bar{x})$ is used as a stopping rule. In an iterative process the function values $F(\bar{x})$ are evaluated at a sequence of points $\bar{x}_1, \bar{x}_2, ..., \bar{x}_i, ...$ It is therefore natural to use stopping rules of the following types:

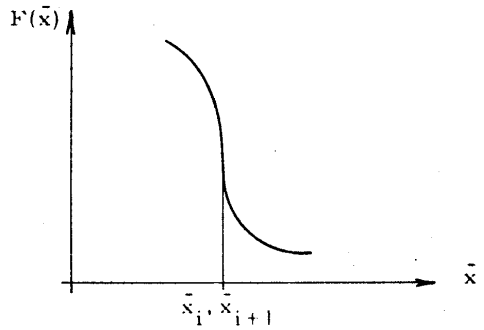


$$\left|F(\bar{x}_{i+1}) - F(\bar{x}_i)\right| < \varepsilon \tag{4}$$

$$\left|\frac{F(\bar{x}_{i+1}) - F(\bar{x}_i)}{F(\bar{x}_i)}\right| < \varepsilon \tag{5}$$

Unfortunately such stopping rules will terminate prematurely on a flat plateau. The same situation will arise by any stopping rule based on differences between function values $F(\bar{x}_{i+1}) - F(\bar{x}_i)$. Further in case of a flat minimum the algorithm may terminate too far away from the exact minimum point $\bar{x}$.

Stopping rules based on the parameters $\bar{x}$:



A simple way to remove the disadvantages by using the stopping rules defined in equations (4) and (5) is to use stopping rules of the following types

$$\left|\bar{x}_{i+1} - \bar{x}_i\right| < \varepsilon \tag{6}$$

$$\frac{\left|\bar{x}_{i+1} - \bar{x}_i\right|}{\left|\bar{x}_i\right|} < \varepsilon \tag{7}$$

But such stopping rules based on $\left|\bar{x}_{i+1} - \bar{x}_i\right| < \varepsilon$ will terminate prematurely on a very steep slope. A combination of stopping rules based on function values and the parameters $\bar{x}$ may be a usable solution.

Stopping rules based on the gradient vector $\nabla F(\bar{x})$:
If the gradient vector

$$\nabla F(\bar{x}) = \left(\frac{\partial F(\bar{x})}{\partial x_1}, \ldots, \frac{\partial F(\bar{x})}{\partial x_n}\right) \tag{8}$$

exists, it is likely that it can be used as a stopping rule because a necessary condition for $\bar{x}$ to be a local minimum point for the differentiable function *F* is, that $\nabla F(\bar{x}) = \bar{0}$. The stopping rule

$$\left|\nabla F(\bar{x})\right| < \varepsilon \tag{9}$$

is often used in connection with the stopping rules (6) or (7), see e. g. Sargent & Sebastian [7]. If the stopping rule (9) is used solely it has the same defiance's as the stopping rules (4) and (5).

Stopping rules based on the number of function value calculations:
The number of function value calculations can be of decisive importance in some optimization problems when the objective function is very complicated. To keep the computer-time down it is in such cases necessary to stop the iterative process when a prescribed number of function value calculations have taken place. It is important to have in mind that such a stopping rule can be effective far-away from the minimum point, so it must be used together with a stopping rule, which can secure the minimum.

It may be a good idea when attacking a new optimization problem to use stopping rules based on the number of function value calculations to encircle a good starting point for the iterative process.

For an optimization strategy based on random search it is in most cases necessary to use a stopping rules of this type.

Stopping rules based on the number of successful steps:

When using an optimization strategy based on random search a great number of trial steps will be failures. The number of successful steps will therefore of course be less than the number of function value calculations. So a particular stopping rule based on the number of successful steps is not relevant as a protection against using too much

computer-time if a stopping rule based on the number of function value calculations already is used.

However, when both local and global search are used, as in the algorithm presented above, it may be appropriate to aim at a certain number of successful local steps before a global step is taken. The interaction between the local and global steps will be treated more detailed in the next chapter.

RANDOM SEARCH

In the last chapter a number of stopping rules was presented and some of their advantages and disadvantages were discussed. When the objective function is unimodal and a non-random optimization method is used the main problem in relation to stopping rules is to choose the values for the numbers ε in the termination criteria. At least two stopping rules must be used to prevent the algorithm to terminate prematurely. It is in general possible to get a good solution of the optimization problem by choosing small numbers for ε but the numerical calculations may then be rather expensive. So the ε-values must be chosen as a compromise between the desire to get a good solution and the limitations in computer time.

However in the case of a multimodal function it is clear that the use of small ε-values will not guarantee that the algorithm will end up with a good solution. It will only secure that a good local minimum is determined. In this chapter stopping rules for random search methods will be discussed in relation to the specific strategy sketched above. The special idea behind this strategy is the use of local and global steps. But some of the considerations here will also be of interest for other random methods and for non-random methods when used on multimodal functions.

The main problem concerning the local steps is the following. The local steps should be chosen in such a way that the convergence to the local minimum is fast. But it is a waste of computer time to get very close to the minimum point if it is only a loca1 minimum. If the local search is very close to the minimum point it may be difficult to escape from it. On the other hand it is inexpedient to leave a local minimum if it is the global minimum as it may be difficult to come back to it again. So the local search should be so arranged that it terminate relative close to the local minimum point with respect to the function value. That is the stopping rules based on function values should be used with not too small ε-values. Later when the global minimum is encircled one can always use a smaller ε-value and combine the stopping rule with one based on the parameters $\overline{x}$.

As mentioned earlier the intention with the global steps is to be able to escape from a local minimum. The global part of the algorithm should be arranged in such a way that any point of the feasible region can be inspected. The number of trial global steps will therefore depend of the type of the feasible region to secure a reasonable covering of the region. So it is reasonable to let the search by global steps terminates by a stopping rule based on the number of trial steps that is the number of function value calculations. It is important to remember that in general it is impossible to be sure that the global minimum point is disclosed.

A number of tests with the strategy by Thoft-Christensen & U. Hartmann [4] have been carried out with combinations of stopping rules based on the number of function value calculations and stopping rules based on the number of successful steps. In a typical experiment with the function

$$F(x,y,z) = 2 - \exp\left(-\frac{1}{4\pi}\sqrt{x^2+y^2+z^2}\right)\cos x \quad \cos y \quad \cos z \tag{10}$$

the starting point was (-1, 10, 10). After 18 successful steps the search was trapped by a local minimum (F = 1.458). A successful global step was then taken and a better local minimum (F = 1.296) was disclosed in 11 successful steps. Finally a global step made it possible to reach the global minimum (F = 1.000) in 30 steps. The experience from several experiments with this strategy shows clearly that the number of global trial steps compared with the number of successful local steps is of very great importance. As the examination of this problem is not finished yet it is not possible to go into further details here. But it seems to be advantageous to use it combination of stopping rules based on the number of function value calculations and stopping rules based on the number of successful steps and perhaps finish the search with stopping rules based on function values and parameters $\bar{x}$.

CONCLUSIONS

As the problem of choosing relevant and effective stopping rules not is solved yet there is a great need for more research in this area. First of all it is important when random optimization methods are used to get more experimental experience. The strategy by which the random steps are chosen is important in this connection. Therefore the combination of probability distributions and stopping rules must be theoretical and experimental investigated.

REFERENCES

[1] Lavi, A. & T.P.Vogl. *Recent Advances in Optimization Techniques*. John Wiley and Sons, N. Y.,1966.

[2] Lootsma, F. A. *Numerical Methods for Non-linear Optimization*. Academic Press, London, 1972.

[3] Jacoby,S.L.S., J.S.Kowalik & J.T.Pizzo. *Iterative Methods for Nonlinear Optimization Problems.* Prentice-Hall, Englewood Cliffs, 1972.

[4] Thoft-Christensen, P. & U, Hartmann. *A Strategy of Optimization Based on Random Search.* Stochastic Control Symposium, Budapest, Hungary, September 1974, pp. 297-303.

[5] Thoft-Christensen, P. *Application of Stochastic Methods in Optimization*. VII International Congress on the Application of Mathematics in Engineering, Weimar, DDR, June 1975.

[6] Himmelblau, D. M. (1972) *A Uniform Evaluation of Unconstrained OptimizationTtechniques*. pp. 69 -97 in [2].

[7] Sargent, R. W. H. & D. J. Sebastian. *Numerical Experience with Algorithms for Unconstrained Minimization*. pp. 45-68 in [2].