Exploring the Search Space With Intervals

Marek W. Gutowski

Institute of Physics, Polish Academy of Sciences, Warsaw, Poland
email: gutow@ifpan.edu.pl

Abstract. The term *global optimization* is used in several contexts. Most often we are interested in finding such a point (or points) in many-dimensional search space at which the objective function's value is optimal, i.e. maximal or minimal. Sometimes, however, we are also interested in *stability* of the solution, that is in its robustness against small perturbations. Here I present the original, interval-analysis-based family of methods designed for exhaustive exploration of the search space. The power of interval methods makes it possible to reach all mentioned goals within a single, unified framework.

1 Preliminaries

We start with brief introduction to interval calculus. Next the necessary tools, useful for search space exploration, are presented. In sequel, the sketch of the procedure good for the most often encountered optimization problems is outlined. We finish with discussion on possible extensions for less regular objective functions.

1.1 Intervals and interval computations

An **interval** is a connected[1] subset of real numbers: $\mathbf{x} = [\underline{x}, \overline{x}] := \{x \in \mathbb{R}\colon \underline{x} \leqslant x \leqslant \overline{x}\}$. We refer to the two real numbers, $\underline{x} \leqslant \overline{x}$, as to interval's endpoints. The ordinary real numbers may be identified with degenerate intervals (called also *thin intervals* or *singletons*) having identical endpoints. Shortly: $\mathbb{R} \subset \mathbb{IR}$, where $\mathbb{IR}$ denotes the set of all intervals.

It is possible to define the interval counterparts of ordinary arithmetic operations, acting on $\mathbb{IR} \times \mathbb{IR}$, in such a manner that they would produce exactly the same results when applied to thin intervals as their originals do when acting on real numbers. For truly interval operands those operators produce interval results containing all the possible outcomes of the ordinary addition, subtraction, multiplication or division. This way an uncountable infinity of regular arithmetic operations can be "performed" simultaneously by few manipulations with endpoints of their interval arguments. For example:

$$\mathbf{x} - \mathbf{y} := \left[\underline{x} - \overline{y},\ \overline{x} - \underline{y}\right], \quad [3,5] - [1,2] = [1,4] \tag{1}$$

[1] Reminder: a set is *connected* when any two its elements can be connected by a path fully contained in the same set.

In computer evaluation of interval expressions we have to remember that every intermediate result has to be properly rounded before attempting the next operation. It is the *outward rounding* what makes interval results guaranteed.

It is worth noticing that simple arithmetic operations on intervals always produce *tight* results. This is not always the case when we deal with more complicated arithmetic expressions, with one or more interval variables (or constants) appearing more than once. In particular we have:

$$\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) \ \subseteq \ \mathbf{xy} + \mathbf{xz} \tag{2}$$

We say that evaluation of interval expressions may lead to the *overestimation*.

It is natural to extend the notion of intervals onto such mathematical objects as vectors and matrices. We can also think of interval functions, of one or more variables, taking interval arguments and producing interval results. For more detailed discussion the reader is referred to the literature [1, 2]. For our purposes it is enough to know that the interval extensions of real-valued functions are not unique and that they generally overestimate the true results when evaluated on arguments from $\mathbb{IR}^n$. Multidimensional interval arguments (interval vectors) for obvious reasons are usually called *boxes*.

1.2 Some concepts related to energy landscape investigation

Let f be the objective function of $n \geqslant 1$ variables: $f\colon \mathbb{R}^n \supset \mathcal{D} \to \mathbb{R}$ and let the real number E (called *energy* from now on) be given. The **lower level set** $M_<(E)$ is defined as:

$$M_<(E) := \{\mathbb{R}^n \ni x = (x_1, x_2, \ldots, x_n)\colon \ f(x_1, x_2, \ldots, x_n) \leqslant E\} \tag{3}$$

Similarly, we define the **upper level set** corresponding to the same number E, $M_>(E)$ (please reverse the direction of the inequality in Eq. 3). Both sets are generalizations of the familiar *level set* (contours) of f which is usually formally written as $f^{-1}(E)$, or more precisely as $M_<(E) \cap M_>(E)$. It is easy to see (and to prove) that

$$E_1 \leqslant E_2 \implies M_<(E_1) \subset M_<(E_2) \quad \text{but} \quad M_>(E_1) \supset M_>(E_2) \tag{4}$$

(think of the lake area when the water level rises). What is *not* obvious is that $M_<(E)$ **need not to be a connected set**, i.e. it may be the union of several disjoint components which, in turn, are connected sets. In particular, in Eq. 4 we may have to deal with two-component $M_<(E_1)$ and a single-component $M_<(E_2)$. This observation is crucial for further considerations. In what follows we will assume that the domain $\mathcal{D}$ of function f is a simply connected[2] set and that f is continuous. An example of the 2-dimensional energy landscape is given as Fig. 1.

Of course $M_<(E)$ is empty whenever energy E is lower than the global minimum of f, and $M_<(E) = \mathcal{D}$ if E exceeds the global maximum of f. On the other hand, the lower level set and the upper level set may be considered complementary, since their union $M_<(E) \cup M_>(E) = \mathcal{D}$, is always equal to the domain of function f.

1.3 Cover sets for $M_<(E)$

It is usually not a proper thing to ask interval methods to determine the lower level set exactly. Nevertheless we will be usually able to find the other set $\mathcal{M}$, approximating

[2]The set is *simply connected* when all the paths linking any fixed pair of its elements can be continuously mapped onto a single path. Less precisely: simply connected set "has no holes."

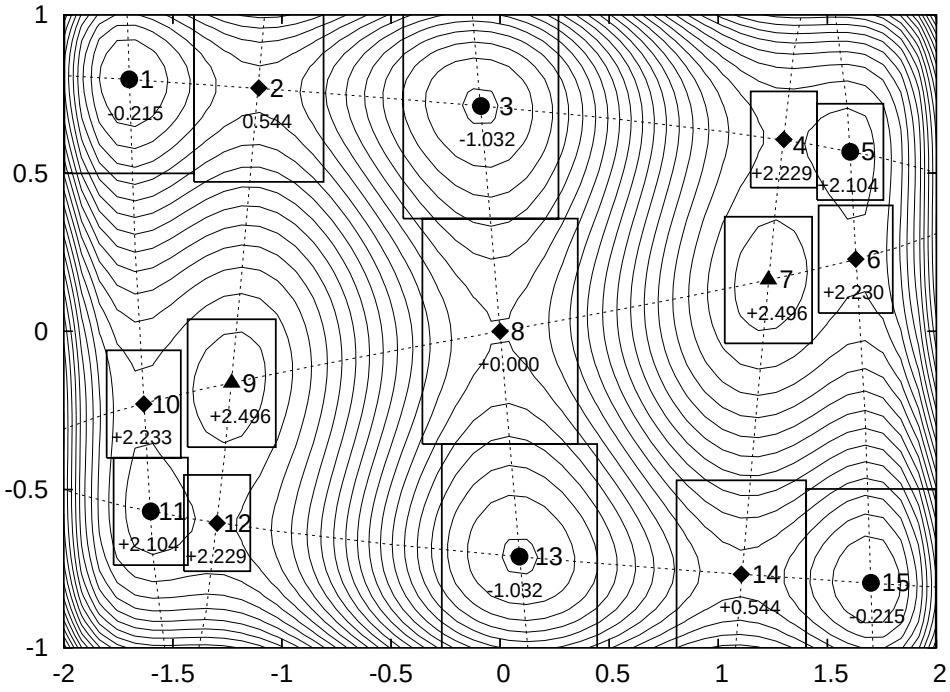

Figure 1. A contour map of the function of two variables $f(x_1, x_2) = \left(4 - 2.1x_1^2 + x_1^4/3\right) x_1^2 + x_1x_2 + 4\left(x_2^2 - 1\right) x_2^2$ with its approximate values at critical points. Dotted lines show where gradient's component(s) are null. Critical points, surrounded by L_∞ balls as ε-neighborhoods, are: minima marked as circles (3 and 13 – global), maxima (7 and 9) – triangles, and saddles – diamonds. Outside the domain shown the function increases to $+\infty$. Chemists are trying to discover the most energetically favourable path connecting minima 11 and 3, without prior detailed knowledge of this map (other than the location of the mentioned minima [3]). The problem is that the saddle 12, which necessarily has to be a part of a such path, is "inactive" during $M_<(E)$ evolution (see text). The other "inactive" saddle is 4.

$M_<(E)$ and containing it at the same time. The overestimation of this cover will be limited by the available computer resources and, eventually, by our patience.

The suitable algorithm, presented below, is rather typical for interval calculations. We start with two lists of boxes: $\mathcal{L}$, containing only one element: a box containing the domain of interest, $\mathcal{D} \subset \mathbb{R}^n$, (ideally: *only* this domain), and $\mathcal{M}$ – the cover set, initially empty. Below is the outline of the algorithm. The box is regarded *small* if its longest edge is shorter than the prescribed accuracy. *Bisection* means splitting a given box into two smaller ones; this is done by splitting the longest edge of the parent box.

Finding the set $\mathcal{M}$ - a cover of set $M_<(E)$.

1. get the next box $\mathbf{x}$ from the list $\mathcal{L}$
2. **if** $\mathbf{x} \cap \mathcal{D} = \varnothing$ **then** // – $\mathbf{x}$ outside of the domain of interest
   - discard $\mathbf{x}$
   - **if** $\mathcal{L} \neq \varnothing$ **then goto** 1 **else** return

   endif
3. evaluate interval extension of the objective $f(\mathbf{x}) = \left[\underline{f(\mathbf{x})},\ \overline{f(\mathbf{x})}\right]$
4. **if** $\underline{f(\mathbf{x})} > E$ **then** // – certainly $\mathbf{x} \not\subset M_<(E)$
   - discard $\mathbf{x}$
   - **if** $\mathcal{L} \neq \varnothing$ **then goto** 1 **else** return

   endif
5. **if** $\overline{f(\mathbf{x})} < E$ **then** // – certainly $\mathbf{x} \subset M_<(E)$
   - append $\mathbf{x}$ to the end of list $\mathcal{M}$
   - **if** $\mathcal{L} \neq \varnothing$ **then goto** 1 **else** return

   endif
6. **if** $\mathbf{x}$ is *small* **then**
   - move $\mathbf{x}$ to the list $\mathcal{M}$ // – not sure, but ...
   - **if** $\mathcal{L} \neq \varnothing$ **then goto** 1 **else** return

   endif
7. bisect $\mathbf{x}$ obtaining $\mathbf{x}_1$ and $\mathbf{x}_2$ such that $\mathbf{x}_1 \cup \mathbf{x}_2 = \mathbf{x}$ and $\mathbf{x}_1 \cap \mathbf{x}_2 \neq \varnothing$.
8. append $\mathbf{x}_1$ and $\mathbf{x}_2$ to the end of list $\mathcal{L}$
9. **goto** 1

On return, the list $\mathcal{M}$ – possibly empty – is the union of boxes covering the set $M_<(E)$. Of course, the set $\mathcal{M}$ almost surely (see step 6 of the algorithm) overestimates the sought set $M_<(E)$. Since the bisection procedure is carried out in such a manner that the intersection of two daughter boxes is always nonempty, then it is possible to find out whether the set $\mathcal{M}$ is connected. If it is not – then we can discover all its connected components.

It is clear that the above algorithm may be easily adjusted to produce the cover of $M_>(E)$. If needed, it can even deliver both covers, $\mathcal{M}_<$ and $\mathcal{M}_>$ in a single pass. For this to work, the step 6 should always update (append) both lists $\mathcal{M}_<(E)$ and $\mathcal{M}_>(E)$.

2 General remarks on optimization

In what follows we assume that *optimization* = *minimization*. This is very common situation in physics, when we are looking for the lowest energy state of the system under investigation. The economists prefer to *maximize* their profit instead, but they can also benefit from our considerations if they only choose to either:

- minimize the loses, or
- replace profit with minus profit in their calculations.

Not so long ago I have presented an especially simple interval-oriented algorithm discovering the type of each critical point [4]. While that method appeared not very

effective in practice[3], and surely needs obvious improvements, it is still an example of the algorithm designed from the very beginning as an interval-oriented tool never returning false results.

In distinction to the previous approach, which scales linearly with the number of critical points and with the dimensionality of the problem under investigation, the current proposal is more complex but, hopefully, equally easy to understand. Let us recall the definition of the minimum of a real-valued function f of many variables:

> *Definition.* The point $x^\star \in \mathbb{R}^n$ is the (proper) minimizer of a function f when
>
> $$\exists_{\varepsilon>0} \quad \forall_{\substack{x\in\mathcal{D}(f)\\ x\neq x^\star}} \quad \|x - x^\star\| < \varepsilon \implies f(x^\star) < f(x) \tag{5}$$
>
> Here "$\|\cdot\|$" denotes any norm on $\mathbb{R}^n$. In simple words: moving away from the point $x^\star$, but within the limited range ε, always leads to the increase of the function value compared to $f(x^\star)$.
> To get the definition of a maximizer, one has to invert the direction of the rightmost inequality in the definition (5) above. Now moving away from $\mathbf{x}^\star$ always results in the decrease of the function f.
> When $x^\star$ is a saddle point of f, then we can always find lower as well as higher values of $f(x)$ in the ε-neighborhood of $x^\star$.

Note 1. Seeing the symbol ε, many of us tend to think of it as of the rather small number, perhaps of order of machine accuracy or so. This is not the case here. It is easy to see that the safe value of ε may be as high as the distance to the other nearest critical point. This is indeed very interesting observation, since its immediate consequence is: **no two extrema of the same kind** (i.e. two minima or two maxima) **can be nearest neighbors**. This means that knowing the character of a given critical point we are able to say something about the type of its nearest neighbor. For example, if we are currently at the maximum of f (no matter whether local or global) then the nearest critical point has to be either a saddle point or a minimum (again: local or global), but certainly not another maximum. Similarly, the nearest neighbor of the minimum is either a saddle point or a maximum. All this can be seen in Fig. 1. Unfortunately, the nearest neighbor of a saddle point can be of arbitrary type, including another saddle point (at least I am unable to prove otherwise).

Note 2. The definition 5 says nothing about the properties of the function f. We will assume, for simplicity, the continuity of f but not necessarily its smoothness (differentiability). Yet, even the continuity is not required and the proposed search space exploring may be successful anyway, even for some discontinuous objectives. Strictly speaking, it may be senseless to speak about the saddle points when the function is not smooth. One should rather speak about the **barriers**, but we will retain the name *saddle point* while keeping in mind this note. The discussion concerning the neighboring critical points remains valid for non-smooth functions as well.

[3] In crude tests tests only 2—14% of cases were correctly recognized and the rest remained undecided. As expected, there were no cases of miss-classification.

3 Exploring the search space

Suppose that our objective function f is continuous and bounded on the domain of interest and that all its values belong to the interval $[E_{\min},\ E_{\max}]$. We will investigate now the behavior of the lower level set for various values of E, when E increases. Reminder: n is the number of variables, that is the dimensionality of the search space.

It is obvious that $M_<(E < E_{\min}) = \varnothing$ and $M_<(E > E_{\max}) = \mathcal{D}$ (the domain of f). It is also clear that once the increasing E exceeds $E_{\min}$, then $M_<(E)$ immediately ceases to be empty. If f has exactly one global minimum, then $M_<(E \gtrsim E_{\min})$[4] is the simply connected set, containing the global minimizer $\mathbf{x}^\star$. When f has more than one global minimum, then $M_<$ is a set of few disjoint but connected components, each one of them being also simply connected. The kind of connectedness (simple or multiple), in contrast to the connectedness itself, or a lack of it, is of no interest to us, so I will not discuss it further. It is worth noticing that for $E \gtrsim E_{\min}$ the number of connected components is exactly equal to the number of identical isolated minima of f. Note however, that this is not so, when we investigate the cover set $\mathcal{M}_<$ instead of $M_<$. Due to the possible overestimates, in such circumstances we can only say that each connected component of $\mathcal{M}_<$ contains *at least one* minimizer of f.

While the energy E continues to increase, the lower level set evolves accordingly. The most important feature affected by this evolution is the number of connected components of $M_<$. It increases every time when energy exceeds the consecutive local minimum and decreases whenever a barrier is encountered. Note that for $n = 1$ a barrier is equivalent to the maximum, while in higher dimension it is rather a saddle point.

Perhaps surprisingly, the number of connected components of $M_<(E)$ remains unchanged at local, as well as at global maximum (again, the case $n = 1$ is an exception). Perhaps even more surprisingly not every one saddle point influences the number of connected components of $M_<$. The very existence of "active" and "inactive" saddles can be linked to the real, physically important phenomena such as phase transitions and chemical reactions pathways. Enough to say that "inactive" saddles (barriers) do not affect the stability of the system wandering in the neighborhood of the local or global energy minimum.

Very similar analysis can be carried out for the case of decreasing energy and the upper level set evolution. To save space, it is left to the reader.

One more remark is in order here. The "inactive" barrier, going unnoticed during the evolution of lower level set *will show up* during upper level set evolution. In conclusion, it is best to count the connected components of both level sets simultaneously.

4 Advantages, disadvantages, and complexity of the method

Generalized level sets seem surprisingly powerful and promising tool for exhaustive exploration of the search space (or of the energy landscape, as it is often called). As such, they are applicable not only to the optimization but to a much wider spectrum of various problems. It can be said, with no exaggeration, that this tool substitutes to the high degree the more familiar classical methods of functional analysis. Finding extrema, whether in the interior of the function's domain or on its boundary, is done in exactly the

[4]The relation symbol "$\gtrsim$" should be read as "slightly higher."

same manner, without resorting to highly specialized (and thus rather sophisticated and conceptually complex) standard methods of global optimization. Moreover, it is a deterministic procedure. Smooth and non-smooth objectives are handled in the same way, no derivatives or Hessian matrices are ever needed.

Due to the hardly predictable and possibly very large number of boxes (worst case estimation goes as $\sim 2^n$), which have to be stored during calculations, the method is practically limited to the relatively low number of variables. On the other hand, the exponential explosion of the number of boxes is *not* very likely. This is because the so called clustering of small boxes mostly occurs in the strict neighborhoods of the critical points. Away from such regions, the objective function is monotonous in many (but not necessarily all!) directions simultaneously and, thanks to this property, the quality of its interval extension is good. This means the overestimates are small or even none. In result, the numerous small boxes will only appear on boundaries of both level sets. Their number may be controlled indirectly by defining properly when the box has to be considered as *small*.

Small boxes and accuracy. It is an illusion that the smaller the *small* boxes, the better is the localization of optimum. The size of a *small* box is only relevant when one might expect the existence of two or more optima of very similar objective values and located very close to each other. In such a case they can be erroneously interpreted as covering a single optimum, if the *small* box happens to be oversized. On the other hand it may be true that the locations of optima may be found with unsatisfactory precision. There are two remedies for that:

- try to improve the precision on energy scale first. This is not very costly, since the bisection in one dimension works really fast, and is only needed in some energy intervals;
- investigate each connected component of $M_<(E)$ (and, maybe, $M_>(E)$) separately, in the ε-neighborhood of the current critical point, *not* in the whole domain $\mathcal{D}$. Such a region is guaranteed to contain exactly one optimum. Moreover, since $\mathbf{x}^\star$ is located in the interior of such a region and not on its boundary, then other, usually more efficient, methods of optimization may/should be applied.

As a norm in interval calculus, the results depend heavily on the quality of the interval extension of the real-valued objective function. It is therefore strongly recommended to follow the algorithm discovering the cover sets strictly in the order outlined here. By doing so, we at least minimize the number of boxes analyzed and deposited on lists $\mathcal{M}_<$ and $\mathcal{M}_>$.

Counting the connected components is easy: the complexity of this task is at most $\mathcal{O}(N^2)$, where N is the number of boxes – elements of the list $\mathcal{M}$.

5 Summary

Here I present the original, highly reliable and interval-calculus-based procedure for finding optima of real-valued functions of many variables. It is very different from any other, numerical or symbolic known algorithm able to deal with such a type of problems. Of course, not everything presented in this paper is my own invention. I was inspired by several articles not necessarily directly related to the current topic.

The roots of this kind of "interval thinking" seem to be due to Sergei Shary, 2001 [5]. He was probably the first one to utilize the lower level set (without naming it explicitly) and to propose the repeated bisection of the energy scale. The spirit of his approach seems to steadily proliferate ever since, even among scientists never exposed to interval analysis. Other authors, including – of course – interval community professionals, are joining quickly. I appreciated the importance of the notion of connectedness after reading the recent papers by Luc Jaulin and others [6, 7]. The mystery why there is a saddle point in the configuration space at the phase transition, but not necessary conversely, was the true driving force in my investigations. It turns out, that set-valued computations are of great help when studying the topological properties of the energy landscapes. As of today, the interval methods seem the only mature and practical tool to handle such computations. As far as I am aware, nobody has put yet the pieces of the puzzle together in the way shown here.

Probably the time for methods of this new type and for their immediate application had just come.

Last but not least, it is a pleasure to mention one more wonderful source of my inspiration, well applicable to the non-smooth energy landscapes, namely the paper [8] written by the great physicists of XIX century, James Clerk Maxwell.

Bibliography

[1] For a concise and nice introduction into interval calculations and methods visit the website http://www.cs.utep.edu/interval-comp/ and references therein.

[2] Brian Hayes, *A Lucid Interval*, American Scientist **91** (6), 484–488, 2003

[3] Jorge J. More, Todd S. Munson, *Computing mountain passes and transition states*, Math. Program. B **100**:151–182, 2004

[4] Marek W. Gutowski. *Classifying extrema using intervals*, arXiv:cs/0609082

[5] Sergey P. Shary, *A Surprising Approach in Interval Global Optimization*, Reliable Computing **7**:497–505, 2001

[6] Nicolas Delanoue, Luc Jaulin, Bertrand Cottenceau, *Using interval arithmetic to prove that a set is path-connected*, Theor. Comp. Sci. **351**:119, 2006

[7] M. Lhommeau, L. Jaulin, L. Hardouin, *Inner and outer approximation of capture basin using interval analysis*, A slide show presented at: *International Conference on Informatics in Control, Automation and Robotics*, 2007

[8] Maxwell, James Clerk, *On Hills and Dales*, Phil. Mag. **40** (269), 421–427, 1870