Why Locating Local Optima Is Sometimes More Complicated Than Locating Global Ones

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

In most applications, practitioners are interested in locating global optima. In such applications, local optima that result from some optimization algorithms are an unnecessary side effect. In other words, in such applications, locating global optima is a much more computationally complex problem than locating local optima. In several practical applications, however, local optima themselves are of interest. Somewhat surprisingly, it turned out that in many such applications, locating all local optima is a much more computationally complex problem than locating all global optima. In this paper, we provide a theoretical explanation for this surprising empirical phenomenon.

1 Formulation of the Problem

A usual understanding is that global optimization is harder. There are many optimization techniques, starting with the simple gradient descent. A usual problem with these techniques is that when they converge, they often lead to a *local* optimum, not to a global one. It takes a special effort to come up with a global optimum instead of a local one.

From this viewpoint, it looks like global optimization is more difficult that locating local optima; see, e.g., [1, 2, 3, 5, 6].

While locating a local optimum may be easier, locating all local optima is difficult. What is indeed relatively easy is locating *a* local optimum. In some practical situations, however, we are actually interested in *all* local optima (see, e.g., [7]); for example:

- in spectral analysis, chemical species are identified by local maxima of the spectrum;
- in radioastronomy, radiosources and their components are identified as local maxima of the brightness distribution; see, e.g., [8];

- elementary particles are identified by locating local maxima of the dependence of scattering intensity on the energy.

It turns out that empirically, the computation problem of finding all local optima is much more computationally complicated than the problem of finding all global optima; see, e.g., [4].

Problem – and what we do in this paper. While empirically, computing local optima is often more complex than computing global ones, there has been, to the best of our knowledge, no convincing theoretical explanation for this complexity.

The main goal of this paper is to provide such a theoretical explanation.

2 Local Optima Are Often More Complex to Locate Than Global Optima: A Possible Theoretical Explanation

Approximating the objective function: a frequent way to solve optimization problems. Often, the computational complexity of an optimization problem is due to the complexity of the objective function. Thus, a reasonable idea is:

- to approximate the original objective function $f(x)$ by a close simpler one $f_\varepsilon(x)$,
- solve the corresponding optimization problem for this simpler objective function $f_\varepsilon(x)$, and
- to use the resulting solution x_ε as a first approximation to the solution of the original optimization problem.

This idea indeed helps in solving global optimization problems, see, e.g., [1, 2, 3, 5, 6].

What we do in this paper. What we will prove is that this simplifying idea cannot be used for locating local optima. This is our first theoretical explanation of why locating local optima is often more computationally complicated than locating global optima.

Definitions and the main result. Let us first explain why the above idea is helpful for locating global maxima: namely, that the above idea helps us dismiss some locations as definitely not containing locations of global optima:

Proposition 1. *Let $f(x)$ and $f_\varepsilon(x)$ be two functions which are ε-close, i.e., for which $|f(x)-f_\varepsilon(x)| \le \varepsilon$ for all x, and let x_ε be a location of the global maximum of the function $f_\varepsilon(x)$, i.e., $f_\varepsilon(x_\varepsilon) = \max\limits_x f_\varepsilon(x)$. Then, for each location $x_{\max}$ of the global maximum of the function $f(x)$, we have $f_\varepsilon(x_{\max}) \ge f_\varepsilon(x_\varepsilon) - 2\varepsilon$.*

Proof. From the fact that $x_{\max}$ is a location of the global maximum of the function $f(x)$, we conclude, in particular, that $f(x_{\max}) \geq f(x_\varepsilon)$. Here,

$$|f(x) - f_\varepsilon(x)| \leq \varepsilon$$

for all x, in particular, $f_\varepsilon(x_{\max}) \geq f(x_{\max}) - \varepsilon$ and $f(x_\varepsilon) \geq f_\varepsilon(x_\varepsilon) - \varepsilon$. Thus,

$$f_\varepsilon(x_{\max}) \geq f(x_{\max}) - \varepsilon \geq f(x_\varepsilon) - \varepsilon \geq (f_\varepsilon(x_\varepsilon) - \varepsilon) - \varepsilon = f_\varepsilon(x_\varepsilon) - 2\varepsilon.$$

The proposition is proven.

Proposition 2. *For every $\varepsilon > 0$, for every continuous function $f_\varepsilon(x)$, and for every point x_0, there exists a function $f(x)$ which is ε-close to $f_\varepsilon(x)$ and which attains a local maximum at the point x_0.*

Discussion. Thus, even if we know everything about the approximating function, we cannot dismiss any point x as a possible location of a local maximum – in other words, the above idea indeed does not work for locating local optima.

Proof. Since the function $f_\varepsilon(x)$ is continuous, there exists a $\delta > 0$ for which $d(x_0, x) \leq \delta$ implies that $|f_\varepsilon(x) - f_\varepsilon(x_0)| \leq \dfrac{\varepsilon}{2}$.

Let us define an auxiliary function $g(x)$ which is equal to:

- $g(x) = f_\varepsilon(x_0)$ when $d(x, x_0) \geq \dfrac{\delta}{2}$ and
- $g(x) = f_\varepsilon(x_0) + \dfrac{\varepsilon}{2} - d(x, x_0) \cdot \dfrac{\varepsilon}{\delta}$ for all other x.

One can easily see that this function is continuous, and that it has a local maximum (actually, even global maximum) for $x = x_0$.

For values x for which $d(x, x_0) \leq d$, the largest possible difference

$$|g(x) - f_\varepsilon(x_0)|$$

between $g(x)$ and $f_\varepsilon(x_0)$ is attained in the second case at the point x_0, when the distance $d(x, x_0) = 0$. In this case, the difference is equal to $|g(x_0) - f_\varepsilon(x_0)| = \dfrac{\varepsilon}{2}$. Thus, for all x, we have $|f_\varepsilon(x_0) - g(x)| \leq \dfrac{\varepsilon}{2}$. So, for all these x, we have

$$|f_\varepsilon(x) - g(x)| \leq |f_\varepsilon(x) - f_\varepsilon(x_0)| + |f_\varepsilon(x_0) - g(x)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence, when x is δ-close to x_0, the values $g(x)$ and $f_\varepsilon(x)$ are ε-close.

Let us now consider the second auxiliary function $w(x)$, which is equal to:

- $w(x) = 1$ when $d(x, x_0) \leq \dfrac{\delta}{2}$;
- $w(x) = 1 - \dfrac{d(x, x_0)}{\delta/2}$ when $\dfrac{\delta}{2} \leq d(x, x_0) \leq \delta$; and

- $w(x) = 0$ when $d(x, x_0) \geq \delta$.

One can check that this function $w(x)$ is also continuous, and its values are always between 0 and 1.

Thus, the convex combination $f(x) \stackrel{\text{def}}{=} w(x) \cdot g(x) + (1 - w(x)) \cdot f_\varepsilon(x)$ is continuous and ε-close to the original function $f_\varepsilon(x)$. For points x for which $d(x, x_0) \leq \dfrac{\delta}{2}$, we have $f(x) = g(x)$, and thus, the function $f(x)$ indeed attains a local maximum for $x = x_0$. The proposition is proven.

Additional theoretical explanation. An additional theoretical explanation for the empirical computational complexity of locating local optima is that this problem also has a higher logical complexity, i.e., needs more quantifiers to describe.

Indeed, the fact that a function $f(x)$ attains its global maximum at a point x_0 is naturally described by a one-quantifier formula $\forall x\,(f(x) \leq f(x_0))$. However, to describe the fact that there is a local maximum at the point x_0, we need two quantifiers: $\exists \delta\, \forall x\,(d(x, x_0) \leq \delta \rightarrow f(x) \leq f(x_0))$.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE- 0926721.

The authors are thankful to Walter Murray to all the participants of the International Conference on Frontiers in Global Optimization (Santorini, Greece, June 8–12, 2003) for valuable discussions.

References

[1] E. M. T. Hendrix and B. G.-Tóth, *Introduction to Nonlinear and Global Optimization*, Springer Verlag, New York, 2010.

[2] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*, Kluwer Academic Publishers, Dordrecht, 2000.

[3] M. Locatelli and F. Schoen, *Global Optimization: Theory, Algorithms, and Applications*, SIAM Publishers, Philadelphia, Pennsylvania, 2013.

[4] W. Murray, In: *Proceedings of the International Conference on Frontiers in Global Optimization*, Santorini, Greece, June 8–12, 2003.

[5] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Verlag, New York, 2006.

[6] P. M. Pardalos and H. E. Romeijn, *Handbook of Global Optimization: Volume 2. Nonconvex Optimization and Its Applications*, Kluwer Academic Publishers, Dordrecht, 2002.

[7] K. Villaverde and V. Kreinovich, "A linear-time algorithm that locates local extrema of a function of one variable from interval measurement results," *Interval Computations*, 1993, No. 4, pp. 176–194.

[8] G. L. Verschuur and K. I. Kellermann, *Galactic and Extra-Galactic Radio Astronomy*, Springer Verlag, Berlin, Heidelberg, New York, 1974.