Computability of Approximate Optima of Nonconvex Functions

K. Lakshmanan
Department of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi, 221005 India.
Email: lakshmanank.cse@iitbhu.ac.in

Abstract

It is known that finding approximate optima of non-convex functions is intractable. We give a simple proof to show that this problem is not even computable.

Keywords: Computablity of Approximate Optima, Non-convex functions

1 Introduction and Preliminaries

We consider the problem of finding the global minima of a non-convex continuous function $f : C \to \mathbb{R}$, where $C \subset \mathbb{R}^d$ is a closed, compact subset. Global minima is the point $x^* \in C$ which satisfies the following property: $f(x^*) \leq f(x)$ for all $x \in C$. The function f attains this minimum at least once by extreme value theorem. Our goal is to find one such point. This problem is well-studied with many books written on the subject, see for example [3].

We note that we consider an oracle setting, where the function values are given by an oracle. This is different from the computablity of optima computable real functions studied for example in [6]. It is easy to see that a simple grid search will output a sequence of points converging to the global optima. And for a Lipschitz continuous function, it requires an exponential number of oracle calls [5] if the Lipschitz constant is known.

Let us define $S = \{x || f(x) - f(x^*)| \leq \varepsilon\}$ and term the members of S as ε-optima. In optimization literature [**?**, **?**], it is known that finding approximations to optima is not tractable for non-convex functions. For non-convex functions, ε-stationary point which is weaker than ε-optima is also known to be not tractable [**?**]. We show more in this paper, that this set S is not computable. This is much stronger than saying they it is intractable. We assume for formality finite-precision numbers. This assumption of finite-precision numbers is useful to model real-world systems and we show is not restrictive.

1.1 Finite Precision Reals

We now briefly explain what we mean by this. Consider any real number $x \in \mathbb{R}$. Let r_0 be the largest integer such that $r_0 \leq x$. Having chosen $r_0, r_1, \ldots, r_{k-1}$ choose largest positive integer r_k such that

$$r_0 + \frac{r_1}{10} + \frac{r_2}{10^2} + \ldots \frac{r_k}{10^k} \leq x.$$

This is the decimal expansion of the number. We can check that this expansion is unique. We define precision length to be the number k. Now for finite precision representaion of a real we need to specify this precision length k. And we say for any real $x \in \mathbb{R}$, the numbers $r_0, r_1, \ldots, r_k$ is its finite precision representation. Note that $r_i, 0 \leq i \leq k$ can be zero. For a point $x \in \mathbb{R}^d$, given a precision length k we can have decimal expansions for all it's co-ordinates. Note that though we give binary representations to the Turing machine, for simplicity we assume precisions denote the decimal precisions.

Remark 1.1 *Suppose $r_1, \ldots, r_k$ is the finite precision representation with length k of some real x. Let $\bar{x}$ be the number with decimal expansion $r_1, \ldots, r_k$ as x and $r_l = 9$ for $l \geq k+1$. And let $\underline{x}$ be the number with decimal expansion $r_1, \ldots, r_k$ as x and $r_l = 0$ for $l \geq k+1$. And the length of this interval $[\underline{x}, \bar{x}]$ is $\varepsilon = 10^{-k}$. We then say with precision length k we can represent consecutive numbers with gap greater than or equal to ε.*

1.2 The Problem

We assume there is an oracle for our continuous function f. This oracle gives the value $f(x)$ up to any finite-precision for an given finite-precision x. The Turing-machine has access to this function oracle. We give also give a value $\varepsilon > 0$ as input to the Turing machine. Our main problem is to write any point x_o of the finite precision length such that $|f(x_o) - f(x*)| < \varepsilon$ i.e., it should find $\varepsilon-$ approximation of the global optima. We show that this problem is not computable.

Let us assume we have a three-tape Turing machine, one is used for calculations, second is for the giving the finite precision real and the precision length required to the function oracle and third one has the value returned from the oracle [4]. Note that the third tape can also store the previous values. That is suppose we start with x_0 and find $x_1, \ldots, x_k$ this tape can store all these and also the corresponding function values obtained from the function oracle $f(x_0), \ldots, f(x_k)$ for finding x_{k+1}. We now give definition of the standard Turing machine here:

Definition 1.2 *Turing machine has a three infinite tapes divided into cells, a reading head which scans one cell of the tape at a time, and a finite set of internal states $Q = \{q_0, q_1, \ldots, q_n\}, n \geq 1$. Each cell is either blank or has symbol 1 written on it. In a single step the machine may simultaneously (1) change the from one state to another; (2) change the scanned symbol s to another symbol $s' \in S = \{1, B\}$;*

(3) move the reading head one cell to the right (R) or left (L). This operation of machine is controlled by a partial map $\Gamma : Q \times S^3 \to Q \times (S \times \{R, L\})^3$.

Remark 1.3 *The map Γ viewed as a finite set of quintuples is called a Turing program. The interpretation is that if $(q, s_1, s_2, s_3, q', s'_1, X_1, s'_2, X_2, s'_3, X_3) \in \Gamma$, in state q, scanning symbols s_1, s_2, s_3 changes state to q' and in the tape i input symbol to s'_i and moves to scan one square to the right if $X_i = R$ (or left if $X_i = L$.) in the tape i.*

We consider this problem in the paper.

Problem 1.4 *Given a continuous, nonconvex function f, is there a Turing machine with access to the function oracle which can find a $\varepsilon-$ approximation to the global optima of the function f ?*

2 Main Theorem

Given the objective function f, let the set of global minima be denoted by G^f. Now consider ε-approximation to the global minima.

Lemma 2.1 *For all $\varepsilon > 0$ there exists a point x^*_n of finite precision length n such that $|f(x^*) - f(x^*_n)| < \varepsilon$.*

Proof

Let $\delta > 0$ be such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$. Such an $\delta > 0$ exists for all $\varepsilon > 0$ because the function f is continuous. Let n be the precision length required to represent numbers with gap $\varepsilon/10$ between consecutive numbers (Remark 1.1). Then we see for the global minima x^* (like for all other points) it's finite precision representation x^*_n with precision length n is such that $|x^*_n - x^*| < \varepsilon$.

Definition 2.2 *Let $G^f_{\varepsilon,k}$ be the set of points with given finite precision length $k \geq 1$ where the function value is $\varepsilon > 0$ close to the global minima. And G^f_ε be the union of all such sets.*

We consider only finite-precision numbers. As there are only finite number of points with precision length k, the set $G^f_{\varepsilon,k}$ is finite. Since we would like an algorithm to computably converge to a single point, for simplicity we assume the global optima is unique i.e., G^f is a singleton. This is not uncommon in optimization literature as strict convexity gives unique local (global) minima and is assumed for objective functions. Now we state the main theorem.

Lemma 2.3 *There is no algorithm to check if a point x_k is a $\varepsilon-$ approximation to the global optima.*

Proof

We consider an equivalent problem. We define

$$h^{\varepsilon}_{x_k}(x) := \max\{0, f(x_k) - f(x) - \varepsilon\}.$$

Since our objective function f is continuous, $h^{\varepsilon}_{x_k}(\cdot)$ is also continuous. This function is identically zero if and only if $|f(x_k) - f(x)| < \varepsilon$, for all x. This happens only if x_k is a $\varepsilon-$ approximation to the optimum (See Figure 1). Note that x_k and x are represented with some finite precision.

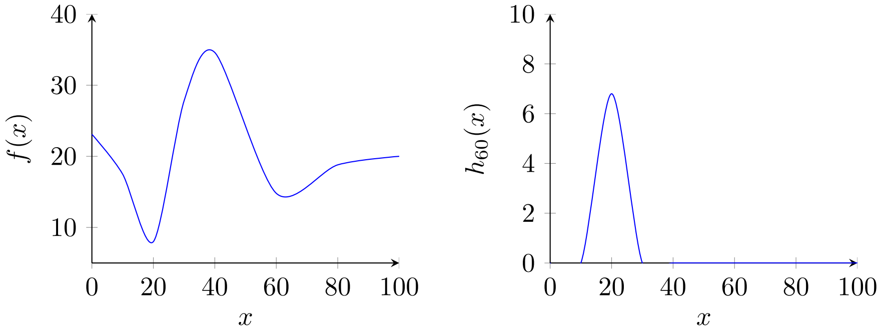

Figure 1: The figure on the right shows a sample objective function. The figure of the left is the function $\max\{0, f(60) - f(x)\}$. This function is not identically zero as $x = 60$ is not the global minima of $f(x)$.

Thus the set G^f_{ε} of all points with finite precision that are ε close to the global minima, is precisely the set of all points x_k where the function $h^{\varepsilon}_{x_k}(\cdot)$ is identically zero. That is,

$$x_k \in G^f_{\varepsilon} \Leftrightarrow h^{\varepsilon}_{x_k}(x) = 0 \text{ for all } x \in C$$

But this cannot be checked for a particular x_k unless it is checked for all x of any finite precision length. As there are infinitely many such points, there is no Turing machine which can compute (halt) if a function is zero at infinitely many points. So given any point x_k, it is not possible to say if it is an $\varepsilon-$ approximation to the global optima or not.

Theorem 2.4 *We assume the objective function we wish to minimize is known by its oracle. There is no algorithm which can compute the ε-approximate optima of a continuous, non-convex objective function on a compact domain.*

Proof

Let x'_k be any point of some finite precision length n_k such that $|f(x^*) - f(x'_k)| < \varepsilon$. Such a point exists by Lemma 2.1 i.e., the set G^f_{ε,n_k} is non-empty for $\varepsilon > 0$. Suppose that we have an algorithm to find a $\varepsilon/2-$ approximation point x'_k.

Now for any point $x_k \in C$ we can say it is ε close to optimum if $|f(x_k) - f(x'_k)| < \varepsilon/2$ else it is not. Thus we have an algorithm to check if a point is an $\varepsilon-$

approximation to the global optima or not. This is a contradiction to Lemma 2.3. Thus for a $\varepsilon > 0$, there is no algorithm to find a $\varepsilon-$ approximation to the global optima.

Corollary 2.5 *The finite-precision assumption is not restrictive. If there is an algorithm to find a general real number which is $\varepsilon-$ approximation, we can take the first k digits which gives $\varepsilon-$ approximation to get a finite precision approximation.*

Corollary 2.6 *The problem of checking whether local minima z is global is not computable as this also involves checking whether $h_z^\varepsilon(\cdot)$ is identically zero.*

Remark 2.7 *Even in presence of higher order oracles, i.,e oracles which give derivatives of the function, the equivalent problem of checking if $h_z^\varepsilon(\cdot)$ is identically zero remains. Hence global optima even in presence of these higher-order function oracle is not computable.*

Remark 2.8 *As we mentioned before, our result is for algorithms having access to the function oracle. This is different from the setting of computable function and reals studied in computable analysis. [6]*

Remark 2.9 *As we can find an ball of some radius ω where the continuous function $h_z^\varepsilon(\cdot)$ is non-zero around a local optima. The same proof does not hold for converging to local optima as we can check if $h_z^\varepsilon(\cdot)$ is identically zero in steps of size less than ω.*

3 Global Optima Property

In this section, we see a simple property a function satisfies if the global optima is computable. For this let us first define the set G^f of global optima of the function $f : C \to \mathbb{R}$, $C \subset \mathbb{R}^d$ as

$$G^f := \{x \mid \text{ for all } y,\, f(x) \leq f(y), x \in C\}.$$

This set G^f is a subset of C. As we are interested in finding only one optima, we call G^f computable if atleast one of it's member is.

Definition 3.1 *A function f satisfies global property if there is a first order (3-ary) predicate $P_\zeta(y,x)$ and a $\zeta \in \mathbb{R}$ such that $P_\zeta(y,x)$ is True for all y,x in the domain C of the function f.*

We say a property $P_\zeta(y,x)$ can be computed if ζ in the definition can be computed.

Definition 3.2 *A set of functions $\mathscr{F}$ satisfies a global property if there is a first order predicate $P_\zeta(y,x)$ satisfied by all the functions $\mathscr{F}$.*

Remark 3.3 *If the global minima is computable for a function f, we can easily define the ζ to be the norm of a member of G^f which is computable, say x^* and the global property to be*

$$P_\zeta(y,x) \text{ is True for all } x, y \text{ if there exists a } x \text{ s.t. } \parallel x \parallel = \zeta.$$

It is clear that if global optima is computable then this $P_{\parallel x^* \parallel}(y,x)$ is satisfied by the function f and $\parallel x^* \parallel$ can be computed.

Remark 3.4 *Lipschitz continuity is another example of such a global property. Let $P_L(y,x)$ be*

$$|f(x) - f(y)| \leq L \parallel x - y \parallel, \text{ for all } x, y \in C.$$

And here the number ζ is the Lipschitz constant L. Let the set of functions on some compact domain C satisfying the Lipschitz property be denoted by $\mathscr{L}$. It is known that if the Lipschitz constant or an upper bound to it is known then the global optima for this class of functions $\mathscr{L}$ is computable. (For example refer to Theorem 1.1.2 of [5]). Another example of a global property is bounded derivatives. If a bound on the gradient is known then the global minima can be computed.

4 Conclusion

We have proved that there is no algorithm which finds a $\varepsilon-$ approximation to the optima of nonconvex function f. This result holds even if the function has higher-order derivatives.

References

[1] C.M. Bishop., Pattern Recognition and Machine Learning, Springer (2006).

[2] T.F. Edgar and D.M. Himmelblau., Optimization of Chemical Processes, Mc-Graw Hill (1988).

[3] R. Horst and H. Tuy., Global Optimization: Deterministic Approaches, Springer-Verlag (1996).

[4] R.I. Soare., Turing Computability: Theory and Applications, Springer-Verlag (2016).

[5] Y. Nesterov., Introductory Lectures on Convex Optimization A Basic Course, Kluwer Academic Publishers (2003).

[6] M. Pour-El and J. Richards., Computability in analysis and physics, Springer, Heidelberg (1989).