DIMACS Technical Report 97-23
June 1997

Extension of Piyavskii's Algorithm to Continuous Global Optimization

by

Robert J. Vanderbei[1,2]
Statistics and Operations Research
Princeton University
Princeton, NJ 08544
E-mail:rvdb@princeton.edu

[1]Permanent Member
[2]Research supported by the NSF through grant CCR-9403789 and by AFOSR through grant F49620-95-1-0351.

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs, Bellcore, and Bell Labs.
DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

Abstract

We use the simple, but little-known, result that a uniformly continuous function on a convex set is ϵ-Lipschitz (as defined below) to extend Piyavskii's algorithm for Lipschitz global optimization to the larger domain of continuous (not-necessarily-Lipschitz) global optimization.

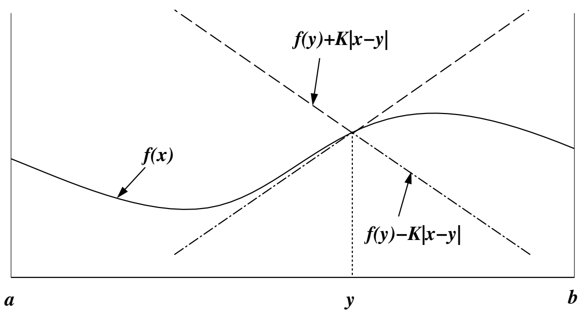

FIGURE 1. Upper and lower bounding function derived from a single observation at y.

1. Introduction

The problem of optimizing a Lipschitz continuous function (with known Lipschitz constant) over a compact set in $\mathbb{R}^n$ is an important global optimization problem since it embodies seemingly minimal assumptions and yet it yields to effective algorithms. In this paper, we give a new characterization of continuous functions on compact, convex domains as being ϵ-Lipschitz (to be defined shortly). Using this new characterization, we are able to extend most, if not all, of the theory of optimization of Lipschitz continuous functions to the more general class of continuous functions (that are not necessarily Lipschitz).

To focus on the main idea, we assume for most of this paper that $n = 1$ and that the domain is a compact interval $[a, b] \subset \mathbb{R}$. We leave to others the interesting task of applying the main idea to the many and varied algorithms that already exist for Lipschitz global optimization. These algorithms are surveyed in [HJL92a] and [Pin96].

The problem then is, given a continuous function f on $[a, b]$, find x^* that attains the global maximum:

$$\max_{x \in [a,b]} f(x).$$

If f is Lipschitz with a known Lipschitz constant (or overestimate thereof), we call the problem a *Lipschitz global optimization* problem.

2. Lipschitz Global Optimization

We begin with a brief description of the main estimate that underlies Lipschitz global optimization. So in this section, we suppose that f is Lipschitz with parameter K. Then, for every $y \in [a, b]$, we have the inequality:

$$|f(x) - f(y)| \leq K|x - y|, \qquad \text{for all } x \in [a, b].$$

Thinking of y as a point at which we've sampled the function and x as variable, the Lipschitz inequality gives us bounds (both upper and lower) on the values of the function f at other points:

$$f(y) - K|x - y| \leq f(x) \leq f(y) + K|x - y|$$

(see Figure 1). Hence, even after sampling the objective function just once, we have a (crude) bound on how far the optimal solution is from the observed sample. Further sampling allows us to improve the bound. For example, if we observe f at

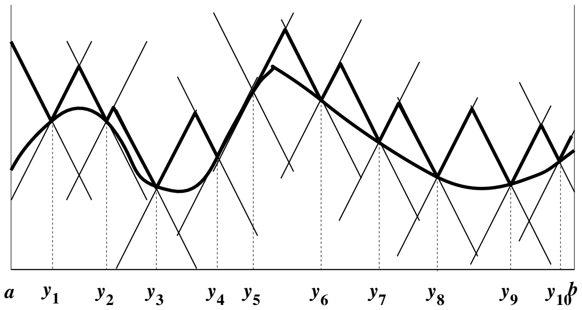

FIGURE 2. Lipschitz continuous objective function and upper bounding function obtained from observations at $y_1, y_2, \ldots, y_{10}$.

the left and right end-points of the feasible interval, $y = a$ and $y = b$, we get the following two upper-bounding functions:

$$f(x) \leq f(a) + K(x-a)$$

and

$$f(x) \leq f(b) - K(x-b).$$

Denoting by z the value of x at which these two bounding functions cross, it is easy to see that

$$z = \frac{1}{2}(a+b) + \frac{1}{2K}(f(b) - f(a)).$$

The common value of the two bounds at this point gives an upper bound on the function f over the entire interval:

$$\max_{x \in [a,b]} f(x) \leq \frac{1}{2}K(b-a) + \frac{1}{2}(f(a) + f(b)).$$

From this bound, we derive an estimate for the closeness of the maximum of $f(a)$ and $f(b)$ to optimality:

$$\begin{aligned} \max_{x \in [a,b]} f(x) - \max\{f(a), f(b)\} &\leq \frac{1}{2}K(b-a) - \frac{1}{2}|f(b) - f(a)| \\ &\leq \frac{1}{2}K(b-a). \end{aligned}$$

Of course, both K and $b-a$ might be quite large. But, by taking further samples, one can in effect replace $b-a$ by the length of successively smaller subintervals (see Figure 2). In this manner, one can develop efficient algorithms for finding a solution that is within a prespecified tolerance of an optimal solution in just a finite number of iterations.

The papers [HJL91], [HJL92a], and [HJL92b] give an extensive survey of algorithms for Lipschitz global optimization all of which are based on this idea. Of course, the methods generally assume *a priori* knowledge of the parameter K. This can be problematic. But, if the objective function is differentiable, then any bound on the magnitude of its derivative can be used for K.

3. A Characterization of Uniform Continuity

The following result will allow us to extend the idea behind Lipschitz global optimization to continuous global optimization:

Theorem 1. *A real-valued function f defined on a convex domain $D \subset \mathbb{R}^n$ is uniformly continuous if and only if, for every $\epsilon > 0$, there exists a $K < \infty$ such that $|f(x) - f(y)| \leq K\|x-y\| + \epsilon$ for all $x, y \in D$.*

This theorem was proved in the unpublished technical report [Van91]. For completeness, we repeat the proof here.

Proof. Suppose that f is uniformly continuous on D and fix $\epsilon > 0$. Then, there exists a $\delta > 0$ such that $|f(z) - f(z')| < \epsilon$ whenever $\|z - z'\| < \delta$. Fix x and y in D and let

$$z_k = y + k\frac{\delta}{2}\frac{x-y}{\|x-y\|} \qquad \text{for } k = 0, 1, 2, \ldots, N$$

where

$$N = \left\lfloor \frac{\|x-y\|}{\delta/2} \right\rfloor,$$

and $\lfloor \cdot \rfloor$ denotes the greatest-integer function. From the convexity of D, we see that each z_k belongs to D. Also,

$$z_0 = y,$$

$$\|z_k - z_{k-1}\| = \delta/2,$$

and

$$\|x - z_N\| < \delta/2.$$

Hence,

$$\begin{aligned} |f(x) - f(y)| &\leq \sum_{k=1}^{N} |f(z_k) - f(z_{k-1})| + |f(x) - f(z_N)| \\ &< (N+1)\epsilon \\ &\leq \frac{2\epsilon}{\delta}\|x-y\| + \epsilon. \end{aligned}$$

Picking $K = 2\epsilon/\delta$ establishes the "only if" direction.

For the "if" part, suppose that f satisfies the condition given in the theorem. Fix $\epsilon > 0$ and choose K so that

$$|f(x) - f(y)| \leq K\|x-y\| + \epsilon/2$$

for all $x, y \in D$. Put $\delta = \epsilon/2K$. If $\|x - y\| < \delta$, then we see that

$$|f(x) - f(y)| < \epsilon.$$

Hence, f is uniformly continuous. □

Since a continuous function on a compact set is uniformly continuous, we can apply Theorem 1 to our optimization problem. Indeed, assuming that a fixed pair (ϵ, K) is known, one can proceed along the same lines as before and derive the following bounds on the values of f based on a single observation at a point x:

$$f(y) - K|x-y| - \epsilon \leq f(x) \leq f(y) + K|x-y| + \epsilon$$

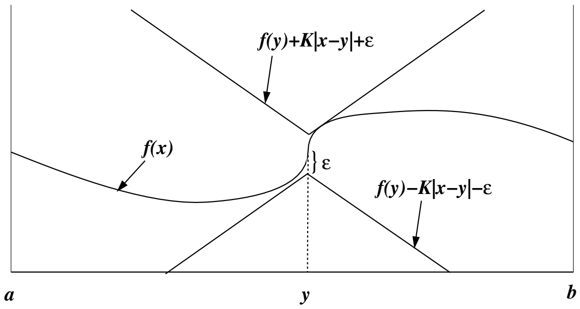

FIGURE 3. Upper and lower bounding function derived from a single observation at y

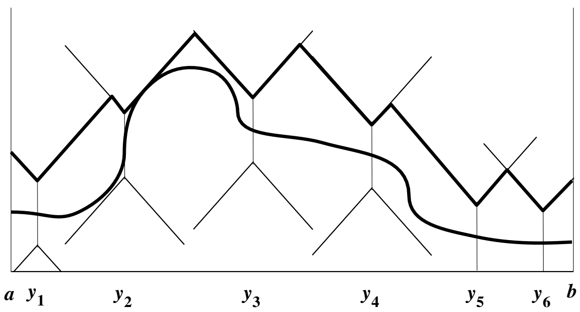

FIGURE 4. Continuous, non-Lipschitz, objective function and upper bounding function obtained from observations at $y_1, y_2, \dots, y_6$.

(see Figure 3). And as before, by judicious choice of subsequent sample points, one can sequentially refine this estimate (see Figure 4). Hence, it is possible to extend essentially every algorithm for Lipschitz global optimization to continuous global optimization. We carry out this program for one such algorithm in the next section.

4. Extension of Piyavskii's Algorithm

The first algorithm for Lipschitz global optimization was published in [Piy67] (see also [Piy72]) and was independently rediscovered by [Shu72] and [Tim77]. Recently, [GJE97] gave an extension of Piyavskii's algorithm to Hölder continuous functions. In this section, we present a different (in fact, simpler) extension of Piyavskii's algorithm to an even larger problem domain, namely, the domain of continuous optimization (there are, of course, continuous functions that are not Hölder continuous, e.g., $1/\log(|x|)$ in a neighborhood of 0).

We begin by introducing some notation. We shall let $\mathcal{Y}$ denote the set of points at which f has been sampled at a particular stage of the algorithm. The first two sample points will be the endpoints of the interval. Hence, we always assume that a and b belong to $\mathcal{Y}$. Next, we shall let $\mathcal{P}$ denote the set of open intervals the union of which is the set of points not yet sampled. That is,

$$\cup_{(y_l, y_r) \in \mathcal{P}} (y_l, y_r) = [a, b] - \mathcal{Y}.$$

At each stage, we shall denote by y^* the best point of those sampled so far. That is,

$$y^* = \text{argmax}\{f(y) : y \in \mathcal{Y}\}.$$

We shall refer to y^* as the *incumbent solution.* Given an interval, say (y_l, y_r), let $u(y_l, y_r)$ denote the upper bound on f obtained by sampling f at the two endpoints of the interval and using the fact that f is (ϵ, K)-Lipschitz. That is,

$$u(y_l, y_r) = \frac{1}{2}K(y_r - y_l) + \frac{1}{2}(f(y_l) + f(y_r)) + \epsilon.$$

Finally, let $z(y_l, y_r)$ denote the point at which the upper bounding function attains this bound:

$$z(y_l, y_r) = \frac{1}{2}(y_r + y_l) + \frac{1}{2K}(f(y_l) - f(y_r)).$$

With these notations, Piyavskii's algorithm is easy to state. First, one must pick two parameters $0 \leq \epsilon < \epsilon'$ and then a value for K in accordance with Theorem 1. The algorithm starts with

$$\begin{aligned} \mathcal{Y} &= \{a, b\} \\ \mathcal{P} &= \{(a, b)\} \\ y^* &= \begin{cases} a & \text{if} f(a) > f(b) \\ b & \text{else.} \end{cases} \end{aligned}$$

Then it loops through the following sequence of steps:

1. Pick $(y_l, y_r) = \text{argmax}\{u(y_l, y_r) : (y_l, y_r) \in \mathcal{P}\}$.
2. If $u(y_l, y_r) - f(y^*) < \epsilon'$, then stop, the incumbent solution y^* is ϵ'-optimal.
3. Put $y = z(y_l, y_r)$.
4. Add y to $\mathcal{Y}$.
5. Remove (y_l, y_r) from $\mathcal{P}$ and add (y_l, y) and (y, y_r) to $\mathcal{P}$.
6. If $f(y) > f(y^*)$, then reset the incumbent to y. That is, $y^* \leftarrow y$.

If f is Lipschitz with parameter K, then one can choose $\epsilon = 0$. The algorithm then reduces exactly to Piyavskii's algorithm for Lipschitz global optimization.

The analysis of the extended Piyavskii algorithm is essentially the same as the usual analysis of Piyavskii's algorithm in the Lipschitz case. We refer the reader to [HJL91] for details for that case.

Of course, to implement Piyavskii's algorithm it must be possible to find an (ϵ, K)-pair. Finding such a pair is very much analogous to finding the K for a Lipschitz continuous function. Indeed, for functions that are differentiable except at a few singular points, finding an (ϵ, K)-pair amounts to analyzing the function at the singular points. This is the topic of the next section.

5. Estimating K

The subject of Lipschitz global optimization is reasonable because an estimate for the Lipschitz constant can often be found. For example, if f is differentiable, then $||f'||_\infty$ can be used as the Lipschitz constant. (Here, the prime denotes differentiation and the norm is the uniform norm.) Since there are specific formulas for computing derivatives (i.e., linearity of differentiation, the product rule, the chain rule, and formulas for the derivatives of specific elementary functions) and bounding them, it is then easy to derive an (upper) estimate for K. In this section, we describe how to carry out the analogous program for continuous functions. Throughout we assume that all functions have compact domains so that no distinction need be made between continuous and uniformly continuous functions.

The assertion that "for every $\epsilon > 0$ there exists a $K < \infty$" in Theorem 1 is equivalent to the assertion that there exists a real-valued function $K(\epsilon)$ defined on the set of positive reals. Of course, such a function is not unique. The set of such functions is closed under pointwise minimization and hence there exists a smallest such function. Denote this smallest function by κ. The following statements are easy to verify:

Theorem 2. (a)

$$\kappa(\epsilon) = \sup_{x,y\in \mathrm{dom}(f)} \frac{|f(x) - f(y)| - \epsilon}{|x-y|}$$

(b) *κ is a decreasing function and therefore its domain of definition can be extended to include 0:*

$$\kappa(0) = \lim_{\epsilon\to 0} \kappa(\epsilon).$$

(c) *If f is Lipschitz with parameter L, then $\kappa(0) \leq L$.*
(d) *If f is continuously differentiable, then $\kappa(0) = \|f'\|_\infty$.*

To emphasize the dependence of κ on the function f, we shall sometimes write κ^f. The following theorem provides results analogous to the linearity of differentiation and analogous to the product and chain rules. Some of the formulas involve *infimal convolution* operator $\oplus$, which is defined by

$$\kappa_1 \oplus \kappa_2(\epsilon) = \inf_{0\leq\delta\leq\epsilon} \left(\kappa_1(\delta) + \kappa_2(\epsilon-\delta)\right).$$

Of course, in practice one does not need to compute the infimal convolution but rather one simply picks some δ between 0 and ϵ to get a suitable upper bound.

Theorem 3. *Let f and g be continuous functions and let α be a real number. The following computational identities hold:*

(a) $\kappa^{\alpha f}(\epsilon) = |\alpha|\kappa^f(\epsilon/|\alpha|)$.
(b) $\kappa^{f+g} \leq \kappa^f \oplus \kappa^g$.
(c) $\kappa^{fg} \leq \kappa^{\|g\|_\infty f} \oplus \kappa^{\|f\|_\infty g}$.
(d) $\kappa^{f\circ g}(\epsilon) \leq \kappa^f(\epsilon)\kappa^g(0)$.

Proof. (a) Let f be a continuous function and let α be a real number. Then,

$$\begin{aligned}\kappa^{\alpha f}(\epsilon) &= \sup_{x,y} \frac{|\alpha f(x) - \alpha f(y)| - \epsilon}{|x-y|} \\ &= |\alpha| \sup_{x,y} \frac{|f(x) - f(y)| - \epsilon/|\alpha|}{|x-y|} \\ &= |\alpha|\kappa^f(\epsilon/|\alpha|).\end{aligned}$$

(b) Let f and g be continuous functions defined on the same domain. Then, for any $0 \leq \delta \leq \epsilon$, we have

$$\begin{aligned}
\kappa^{f+g}(\epsilon) &= \sup_{x,y} \frac{|f(x)+g(x)-f(y)-g(y)|-\epsilon}{|x-y|} \\
&\leq \sup_{x,y} \frac{|f(x)-f(y)|+|g(x)-g(y)|-\epsilon}{|x-y|} \\
&\leq \sup_{x,y} \frac{|f(x)-f(y)|-\delta}{|x-y|} + \sup_{x,y} \frac{|g(x)-g(y)|-(\epsilon-\delta)}{|x-y|} \\
&= \kappa^f(\delta) + \kappa^f(\epsilon-\delta).
\end{aligned}$$

Now taking the infimum over all δ, we get

$$\kappa^{f+g}(\epsilon) \leq \inf_{0\leq\delta\leq\epsilon} \left(\kappa^f(\delta)+\kappa^g(\epsilon-\delta)\right) = \kappa^f \oplus \kappa^g(\epsilon).$$

(c) Let f and g be continuous functions defined on the same domain. Then

$$\begin{aligned}
\kappa^{fg}(\epsilon) &= \sup_{x,y} \frac{|f(x)g(x)-f(y)g(y)|-\epsilon}{|x-y|} \\
&\leq \sup_{x,y} \frac{|f(x)-f(y)||g(x)|+|g(x)-g(y)||f(x)|-\epsilon}{|x-y|} \\
&\leq \sup_{x,y} \frac{|f(x)-f(y)|\|g\|_\infty-\delta}{|x-y|} + \sup_{x,y} \frac{|g(x)-g(y)|\|f\|_\infty-(\epsilon-\delta)}{|x-y|} \\
&= \kappa^{\|g\|_\infty f}(\delta) + \kappa^{\|f\|_\infty g}(\epsilon-\delta).
\end{aligned}$$

Now taking the infimum over all δ, we get

$$\kappa^{fg}(\epsilon) \leq \inf_{0\leq\delta\leq\epsilon} \left(\kappa^{\|g\|_\infty f}(\delta)+\kappa^{\|f\|_\infty g}(\epsilon-\delta)\right) = \kappa^{\|g\|_\infty f} \oplus \kappa^{\|f\|_\infty g}(\epsilon).$$

(d) Let f and g be continuous functions such that the range of g is contained in the domain of f. Then

$$\begin{aligned}
\kappa^{f\circ g}(\epsilon) &= \sup_{x,y} \frac{|f(g(x))-f(g(y))|-\epsilon}{|x-y|} \\
&= \sup_{x,y} \left(\frac{|f(g(x))-f(g(y))|-\epsilon}{|g(x)-g(y)|} \cdot \frac{|g(x)-g(y)|}{|x-y|}\right) \\
&\leq \sup_{\xi,\eta} \left(\frac{|f(\xi)-f(\eta)|-\epsilon}{|\xi-\eta|}\right) \sup_{x,y} \left(\frac{|g(x)-g(y)|}{|x-y|}\right) \\
&= \kappa^f(\epsilon)\kappa^g(0).
\end{aligned}$$

□

Finally, it is easy to compute κ for most elementary functions. For example, if

$$f(x) = x^\alpha, \qquad x \in [0,b],$$

where $0 < \alpha < 1$, then an easy calculation shows that

$$\kappa(\epsilon) = \alpha \left(\frac{1-\alpha}{\epsilon}\right)^{\frac{1-\alpha}{\alpha}}.$$

Also, if

$$f(x) = 1/\log x, \qquad x \in [0,a],$$

where a is a small positive number, then

$$\begin{aligned}\kappa(\epsilon) &= \left(\frac{2\epsilon}{1+\sqrt{1-4\epsilon}}\right)^2 \exp\left(\frac{1+\sqrt{1-4\epsilon}}{2\epsilon}\right) \\ &\approx \epsilon^2 e^{1/\epsilon}.\end{aligned}$$

6. Extension to Higher Dimensions

To extend Piyavskii's algorithm to higher dimensions, the main difficulty is to be able to evaluate the location and value of the maximum of a lower envelope of functions whose graphs are lower envelopes of cones and whose vertices are at the corners of a multidimensional rectangle. This maximization is difficult if the cones correspond to the Euclidean norm; i.e., if the functions are of the form:

$$\phi(x) = f(y) + K\|x-y\|_2.$$

But, we are free to choose any norm. The above optimization is much easier if the norm is the l^1 norm instead of the l^2 norm. And since in finite dimensions all norms are equivalent, the choice of norm has no effect on the class of functions covered by the theory except that the value of K will vary. Furthermore, in low dimensions even K won't change very much. Hence, it is natural to construct higher dimensional variants of the algorithm using the l^1 norm. Again, we leave the details to the interested reader.

References

[GJE97] E. Gourdin, B. Jaumard, and R. Ellaia. Global optimization of Hölder functions. *J. of Global Optimization*, 1997. To appear.

[HJL91] P. Hansen, B. Jaumard, and S.H. Lu. On the number of iterations of Piyavskii's global optimization algorithm. *Mathematics of Operations Research*, 16:334–350, 1991.

[HJL92a] P. Hansen, B. Jaumard, and S.H. Lu. Global optimization of univariate Lipschitz functions: I. Survey and properties. *Mathematical Programming*, 55:251–272, 1992.

[HJL92b] P. Hansen, B. Jaumard, and S.H. Lu. Global optimization of univariate Lipschitz functions: II. New algorithms and computational comparison. *Mathematical Programming*, 55:273–292, 1992.

[Pin96] J.D. Pintér. *Global Optimization in Action – Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications.* Kluwer, Dordrecht, 1996.

[Piy67] S.A. Piyavskii. An algorithm for finding the absolute minimum of a function. *Theory of Optimal Solutions*, 2:13–24, 1967. IK Akad. Nauk USSR, Kiev.

[Piy72] S.A. Piyavskii. An algorithm for finding the absolute extremum of a function. *USSR Comput. Math. Math. Phys.*, 12:57–67, 1972.

[Shu72] B.O. Shubert. A sequential method seeking the global maximum of a function. *SIAM J. Numer. Anal.*, 9:379–388, 1972.

[Tim77] L.N. Timonov. An algorithm for search of a global extremum. *Engrg. Cybernetics*, 15:38–44, 1977.

[Van91] R.J. Vanderbei. Uniform continuity is almost Lipschitz continuity. Technical Report SOR-91-11, Statistics and Operations Research Series, Princeton University, 1991.