Journal of Global Optimization **8:** 91–103, 1996.

Estimation of the Lipschitz Constant of a Function

G. R. WOOD
Department of Mathematics and Computing, Central Queensland University, Rockhampton, Queensland 4702, Australia

and

B. P. ZHANG
Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand

(Received: 1 June 1994; accepted: 10 April 1995)

Abstract. A number of global optimisation algorithms rely on the value of the Lipschitz constant of the objective function. In this paper we present a stochastic method for estimating the Lipschitz constant. We show that the largest slope in a fixed size sample of slopes has an approximate Reverse Weibull distribution. Such a distribution is fitted to the largest slopes and the location parameter used as an estimator of the Lipschitz constant. Numerical results are presented.

Key words: Global optimisation, Lipschitz constant, Reverse Weibull distribution, Gnedenko condition, asymptotic distribution

1. Introduction

Many global optimisation algorithms are based on the Lipschitz continuity of the objective function, g. For such algorithms it is critical to have an estimate of the Lipschitz constant. It is also necessary to have a good estimate, since the better the estimate the faster such algorithms converge, as described in [6]. Finding the Lipschitz constant of a function is itself a global optimisation problem, since the Lipschitz constant is the supremum of the magnitudes of directional derivatives of the function.

Existing methods dealing with the Lipschitz constant estimation problem in the literature fall into two categories. In the first the analytical form of the objective function and its derivatives are known explicitly, while in the second the form is unknown and only the function value can be evaluated. We term these objective functions *white box* and *black box* functions respectively.

For the white box problem, Shubert [13] gave a univariate example of Lipschitz constant estimation using the upper bound of the derivative. Mladineo [9] discussed the two dimensional case and chose the upper bound of $\sqrt{(\partial g/\partial x)^2 + (\partial g/\partial y)^2}$ as the estimate. Range inclusion techniques of interval analysis due to Moore [10] were used by Gourdin, *et al.* [5] to estimate the Lipschitz constant for the problem of maximum likelihood estimation of the three-parameter Weibull distribution. It is evident that any appropriate global optimisation method may be applied to find the Lipschitz constant of a white box objective function.

On the other hand, for the black box problem, we have to find an upper bound for the magnitude of the gradient of the function using only the available function evaluations. Strongin proposed a method for univariate functions in [14]. After k evaluations, the ordered evaluation points $x_1 < x_2 < \ldots < x_k$ and corresponding function values $g(x_1), g(x_2), ..., g(x_k)$ are available and an under-estimation of the Lipschitz constant is given by $\hat{m} = \max_i\{|g(x_i) - g(x_{i-1})|/(x_i - x_{i-1})\}$. The Strongin estimate is then obtained by multiplying $\hat{m}$ by a factor $r > 1$. There is no guarantee, however, that the estimate $r\hat{m}$ is greater than or equal to the true Lipschitz constant. Hansen *et al.* [7] showed that no matter how large the factor r is chosen, the Strongin estimator is an under-estimation of the true Lipschitz constant for a class of constructed Lipschitz functions, and hence Strongin's companion algorithm may terminate at a local optimum. In [2] de Haan proposed a method for estimating the minimum of a function using order statistics, and discussed necessary conditions on the objective function. While the method is similar in philosophy to our approach, our method requires only objective function evaluations, and not those of a derivative, to estimate the Lipschitz constant.

Our method builds on the ideas of Strongin and de Haan and addresses the Lipschitz constant estimation problem for univariate functions alone. The development here can informally be described as follows. Recall that our aim is to find the supremum of all slopes $|g(x) - g(y)|/|x - y|$ for distinct points x and y in the domain of g. If we sample X and Y uniformly from the domain, then the random variable $|g(X) - g(Y)|/|X - Y|$ itself has cumulative distribution function F, which we term the *slope distribution* of g. The upper bound of its support is the Lipschitz constant we need. Unfortunately we do not know the form of F for an arbitrary objective function g.

Suppose now that we draw a random sample of say n absolute slopes, and consider the distribution of the largest. Provided that F satisfies the Gnedenko condition, given in the next section, then the distribution of the largest absolute slope is known to be approximately Reverse Weibull. Its location parameter (the upper bound of the support) will estimate our Lipschitz constant.

We formalise these ideas in Section 2 and illustrate the method with numerical results in Section 3. In Section 4 we show that for a large class of functions the slope distribution does satisfy the Gnedenko condition. In Section 5 we discuss directions for future research.

2. The Method

For $[a, b]$ an interval in the reals, $\mathbf{R}$, define $L(M)$ to be all functions $g : [a, b] \to \mathbf{R}$ such that $|g(x) - g(y)| \leq M|x - y|$ for all $a \leq x, y \leq b$. These are the Lipschitz continuous functions from $[a, b]$ to $\mathbf{R}$, with Lipschitz constant M.

We begin by illustrating the method with a simple example. In Figure 1 we show the function $g(x) = x - x^3/2$ on $[-1, 1]$. The Lipschitz constant is $M = 1$ since $g'(0) = \max_{x \in [-1,1]} g'(x) = 1$. We choose x and y uniformly and independently in $[-1, 1]$. In the example $x = -0.78$ and $y = 0.84$. The absolute value of the slope estimate is $s = |g(x) - g(y)|/|x - y| = 0.67$.

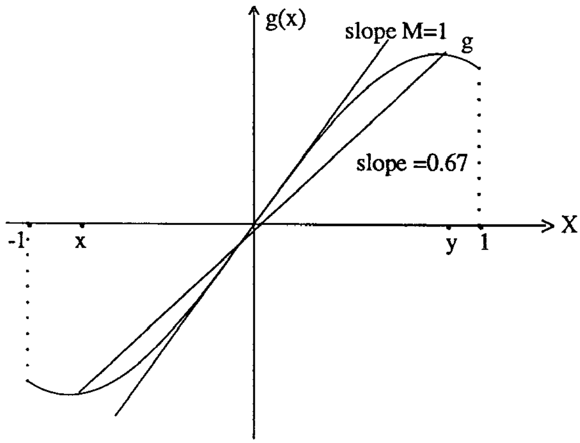

Fig. 1. The objective function g and a sampled slope of 0.67.

Suppose now that we repeat this procedure many times. Then the cumulative distribution of such slope absolute values converges to the cumulative slope distribution F, shown in Figure 2. Note that the associated probability density function will have support with upper limit M.

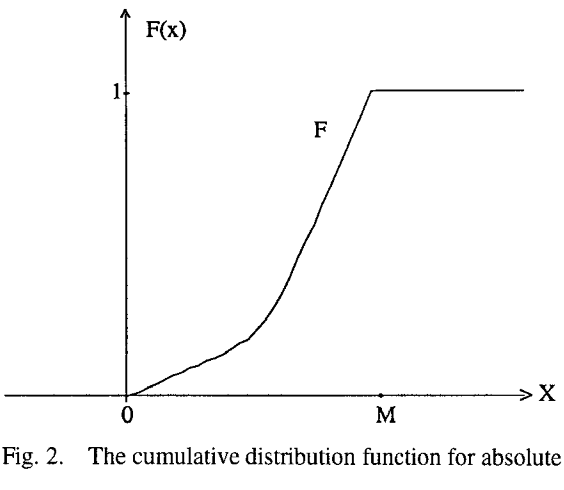

Fig. 2. The cumulative distribution function for absolute values of slopes.

Now take a random sample of size $n = 5$ from this distribution. Let l be the largest of these absolute slopes. Then the cdf of l will be $F^5(x)$. This maximum slope cumulative distribution function is shown in Figure 3. Note that the probability density function of l will have the same support upper bound M.

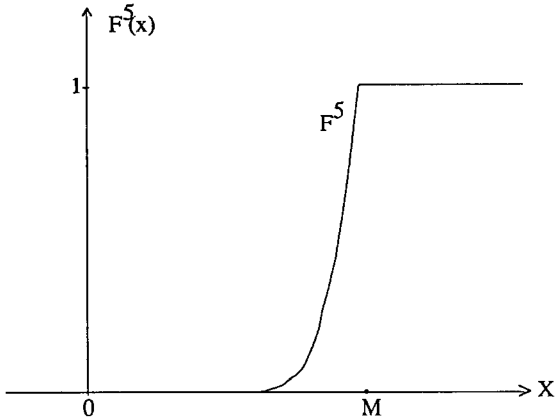

Fig. 3. The cumulative distribution function for the largest of a sample of five absolute slopes.

In general, what is the distributional form of F^n? If we know the distribution towards which F^n tends as n increases, then the estimate of the location parameter M can be used as an estimate of the Lipschitz constant of g.

The theory of extreme value distributions tells us that there are just three distributional types for extreme values. Since the support of F is bounded above, standard theory yields that F^n is approximately Reverse Weibull, provided that the original distribution F satisfies the Gnedenko condition

$$\lim_{\epsilon\to 0+} \frac{1 - F(M - c\epsilon)}{1 - F(M - \epsilon)} = c^k$$

for any $c > 0$ and some constant $k > 0$. This result first appeared in [4].

The cumulative distribution function of the three-parameter Reverse Weibull distribution is

$$H(l) = \begin{cases} \exp\{(-(u - l)^w)/v\} & \text{if} \quad l \leq u \\ 1 & \text{if} \quad l > u, \end{cases}$$

where $u \in \mathbf{R}$ is the location parameter, $v > 0$ is the scale parameter and $w > 0$ is the shape parameter. A precise description of the manner in which F^n converges to a Reverse Weibull distribution is given in Section 4. We fit a Reverse Weibull distribution to a sample of largest absolute slopes and use the estimate of the location parameter as an estimate of M.

We now formally present this stochastic procedure to estimate the Lipschitz constant M of g. Our method assumes only that the function values can be evaluated. Larger values of the absolute slope will be found for (x, y) pairs chosen near the diagonal of $[a, b] \times [a, b]$. For this reason we set up a sampling scheme in Step 1 which allows pairs to be chosen according to a uniform distribution on a strip around the diagonal.

Step 1: Sample the slopes *Given $\delta > 0$, choose pairs (x_i, y_i) uniformly on $\{(x,y) \in [a,b] \times [a,b] : |x-y| \le \delta\}$ and evaluate*

$$s_i = \frac{|g(x_i) - g(y_i)|}{|x_i - y_i|} \quad \textit{for } i = 1, \ldots, n.$$

Step 2: Calculate the maximum slope *Let*

$$l = \max\{s_1, \ldots, s_n\}.$$

Steps 1 and 2 are performed m times, giving $l_1, \ldots, l_m$.

Step 3: Fit the Reverse Weibull distribution *Fit a three-parameter Reverse Weibull distribution to $l_1, \ldots, l_m$.*

Output: *Our estimate $\hat{M}$, of M, is the location parameter of the fitted Reverse Weibull distribution.*

We remark that we have converted the Lipschitz constant estimation problem into a routine curve fitting problem.

3. Numerical Results

Does the method give good results? We investigate this question now. We begin by observing that if a random variable has a Reverse Weibull distribution, then its negative has a Weibull distribution. Thus we can employ standard Weibull fitting methods.

Methods for finding maximum likelihood estimates for the three-parameter Weibull distribution are discussed in detail in the surveys of Zanakis and Kyparisis [16] and Panchang and Gupta [11]. We use a combination of profile likelihood and the method of moments to fit the Weibull distribution. For fixed u, we can straightforwardly use first and second moments to find v and w. We then select the (u, v, w) combination which maximises the likelihood of the observed maximum slopes. We remark that Gourdin *et al.* in [5] proposed a new global algorithm to solve this very problem. We have used Gourdin's method whenever suitable to check our results.

We describe results on three test functions. The first test function is the simple function $x - x^3/3$ on $[-1, 1]$ with Lipschitz constant $M = 1$. The other two test functions are drawn from [15, p. 177]: $\sin x + \sin(2x/3)$ on $[3.1, 20.4]$ with $M = 1.67$ and $-\sum_{i=1}^{5} \sin((i+1)x + i)$ on $[-10, 10]$ with $M = 67$. Tables I, II and III give the estimates of M for various choices of n and m with the sampling performed according to a uniform distribution on $[a,b] \times [a,b]$. Table IV gives estimates of M for these three functions using $m = 100$, various choices of n and with $\delta = 0.05$. In the next section we show that the slope distribution of these test functions satisfies the Gnedenko condition. All of the estimates are given in the form $a \pm b$, where a is the mean and b the standard deviation over ten runs for the given n and m.

Note that the estimates improve as n and m increase. Note also that use of a small δ improves these results, especially when the domain interval is large. This is due to the fact that when the interval is large, sampling from $[a,b] \times [a,b]$ rarely produces a pair (x,y) in the region where the maximum absolute slope is achieved.

TABLE I. Lipschitz constant estimators for $x - x^3/3$ on $[-1,1]$, using uniform sampling on $[-1,1] \times [-1,1]$. Here $M = 1$.

	$m = 25$	$m = 50$	$m = 75$	$m = 100$
$n = 3$	0.9990±0.0074	1.0020±0.0042	1.0000±0.0000	1.0000±0.0000
$n = 5$	1.0000±0.0067	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000

TABLE II. Lipschitz constant estimators for $\sin x + \sin(2x/3)$ on $[3.1, 20.4]$, using uniform sampling on $[3.1, 20.4] \times [3.1, 20.4]$. Here $M = 1.67$.

	$m = 25$	$m = 50$	$m = 75$	$m = 100$
$n = 3$	2.4120±0.7605	2.5840±0.6702	2.8590±0.4459	2.0090±0.6839
$n = 5$	2.4060±0.7327	2.4760±0.6712	2.6170±0.5781	2.0170±0.6784
$n = 7$	1.9520±0.5713	1.9920±0.5289	1.8270±0.2957	1.7140±0.2049
$n = 9$	1.5840±0.1327	1.7220±0.1078	1.7120±0.0863	1.7190±0.0926
$n = 11$	1.7000±0.1441	1.6750±0.1918	1.6720±0.0379	1.6740±0.0398

TABLE III. Lipschitz constant estimators for $-\sum_{i=1}^{5} \sin((i+1)x + i)$ on $[-10, 10]$, using uniform sampling on $[-10, 10] \times [-10, 10]$. Here $M = 67$.

	$m = 25$	$m = 50$	$m = 75$	$m = 100$
$n = 3$	47.3820±24.5935	54.3096±16.0915	61.0462±10.3427	60.1370±9.9099
$n = 5$	59.4610±13.7087	60.0020±12.0155	58.3790±10.7152	59.7570±9.1609
$n = 7$	61.5820±6.2873	55.4950±5.1661	62.3840±8.0365	67.7140±6.5446
$n = 9$	62.2960±10.6238	60.2250±9.8699	66.4400±7.6948	69.4510±6.3622
$n = 11$	60.1370±9.9099	63.6500±8.6893	68.7500±6.3424	72.0440±5.0883

The following example allows us to compare our numerical results with those of Strongin's method. The test function is a modification of that introduced in [7] by Hansen *et al.* Let

TABLE IV. Lipschitz constant estimators for the three test functions using $\delta = 0.05$.

Function		$x - x^3/3$	$\sin x + \sin(2x/3)$	$-\sum_{i=1}^{5} \sin((i+1)x + i)$
Interval		$[-1, 1]$	$[3.1, 20.4]$	$[-10, 10]$
$n = 3$	$m = 100$	1.0000±0.0000	1.7040±0.0227	73.3870±1.8872
$n = 5$	$m = 100$	1.0000±0.0000	1.6790±0.0074	68.4040±0.0975
$n = 7$	$m = 100$	1.0000±0.0000	1.6750±0.0085	68.4250±0.0474
$n = 9$	$m = 100$	1.0000±0.0000	1.6720±0.0042	68.4080±0.0282

$$g(x) = \begin{cases} \max\{2\sin(\frac{\beta\pi x}{2}), x\} & \text{if} \quad 0 \le x < 2/\beta \\ x & \text{if} \quad 2/\beta \le x \le 1 \end{cases}$$

where $\beta = 2(\frac{2r}{r+1})^{\Lambda}$, $\Lambda = \text{ceil}\{\frac{\ln r}{\ln 2r - \ln(r+1)}\} - 2$ and r is the multiplier of Strongin's algorithm. Here ceil(x) is the smallest integer greater than or equal to x. Figure 4 shows g for the case where $r = 2$.

It can be proved that for any mutiplier $r > 1$, Strongin's estimate of the Lipschitz constant is always r, but the true Lipschitz constant M is $\pi\beta$, which equals $2\pi(\frac{2r}{r+1})$, if $1 < r \le 4$, as $\Lambda = 1$. Thus the Strongin estimate is an under-estimate and hence his algorithm converges to a non-global local optimum. Table V compares Lipschitz constant estimates using the Strongin method (S) and the Reverse Weibull method (RW) for two such Hansen test functions on $[0, 1]$.

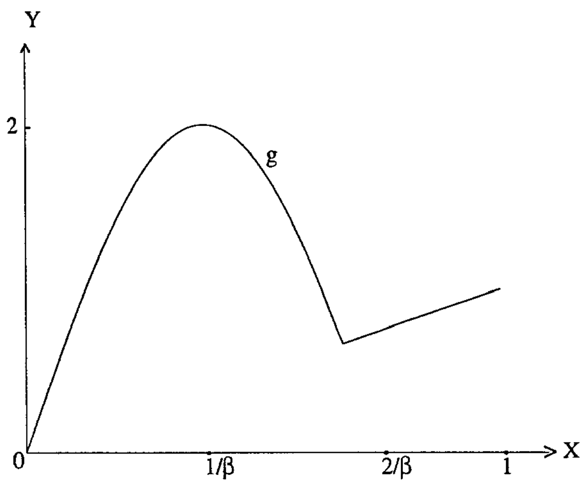

Fig. 4. The Hansen test function with $r = 2$.

TABLE V. Lipschitz constant estimation comparison: the Strongin estimate and the Reverse Weibull estimates for a large δ value and a small δ value.

Factor r	True M	S estimate	n	m	RW estimate $\delta > \|b-a\|$	RW estimate $\delta = 0.05$
			3	100	10.1394 ± 1.6007	8.3052 ± 0.0253
$r=2$	$\pi\beta$		5	100	8.5384 ± 0.2736	8.3792 ± 0.0042
	$= 8.3776$	2	7	100	8.4334 ± 0.1758	8.3752 ± 0.0042
			9	100	8.3644 ± 0.0585	8.3742 ± 0.0095
			20	100	8.3802 ± 0.0127	8.3772 ± 0.0000
			3	100	13.1292 ± 1.7332	9.3182 ± 0.0348
$r=3$	$\pi\beta$		5	100	9.8362 ± 0.5191	9.4192 ± 0.0042
	$= 9.4248$	3	7	100	9.6892 ± 0.2656	9.4222 ± 0.0053
			9	100	9.3772 ± 0.1124	9.4202 ± 0.0095
			20	100	9.4249 ± 0.0495	9.4262 ± 0.0057

4. The Gnedenko Condition

In order that the Reverse Weibull distribution approximate the distribution of the maximum absolute slope, a condition must be satisfied by the original cumulative distribution function F. The following proposition presents this result.

PROPOSITION 1. *Let $M = \sup\{s : F(s) < 1\}$ be finite and L be the largest value in a random sample of size n drawn from F. Then there are sequences a_n and $b_n > 0$ such that $F^n(a_n + b_n l)$ converges pointwise to the standard Reverse Weibull distibution $H(l;k)$ if and only if for any $c > 0$,*

$$\lim_{\epsilon \to 0+} \frac{1 - F(M - c\epsilon)}{1 - F(M - \epsilon)} = c^k,$$

the Gnedenko condition. Here $H(l;k)$ is the Reverse Weibull distribution with $u = 0$, $v = 1$ and $w = k$.

The proof of this proposition can be found in [3, pp.53–57 and pp. 87–91].

The above result tells us that an affine transformation $a_n + b_n l$ of the maximum absolute slope l has an approximate Reverse Weibull distribution. It is easy to prove that if the distribution of a random variable ξ is Reverse Weibull then the distribution of $a + b\xi$, for $b > 0$, is also Reverse Weibull. It follows that the maximum absolute slope itself has an approximate Reverse Weibull distribution. Formally, $F^n(l)$ is approximately $H\left((l - a_n)/b_n; k\right)$.

This result we used in Section 2. It remains to be shown that the Gnedenko condition does hold for the slope distribution F of a wide class of objective functions, g. This is the content of our main theorem, which we now present.

THEOREM 2. *Suppose that* $g \in C^{2p+2}[a,b]$ *for some natural number* p, *that there exists a unique* $z \in (a,b)$ *such that*

$$|g'(z)| = M = \max_{x \in (a,b)} |g'(x)| > 0,$$

and that $g^{(i)}(z) = 0$ *for* $i = 2, 3, \ldots, 2p$ *with* $g^{(2p+1)}(z) \neq 0$. *Given* $\delta > 0$, *let* (x,y) *be a point chosen uniformly on* $\{(x,y) \in [a,b] \times [a,b] : |x-y| < \delta\}$,

$$s(x,y) = \begin{cases} |g(x) - g(y)|/|x-y| & \text{if } x \neq y \\ |g'(x)| & \text{if } x = y \end{cases}$$

and F *denote the cdf of* s. *Then the Gnedenko condition holds for* F *with* $k = 1/p$.

In order to prove Theorem 2, we need the following preliminaries and lemma. Let $D = [a,b] \times [a,b]$. For fixed $\epsilon > 0$ and $c > 0$ let $D_\epsilon = \{(x,y) \in D : x \neq y \text{ and } |g(x) - g(y)|/|x-y| \geq M - \epsilon\} \cup \{(x,x) \in D : g'(x) \geq M - \epsilon\}$. Typical D_ϵ and $D_{c\epsilon}$ are illustrated in Figure 5, with $c > 1$.

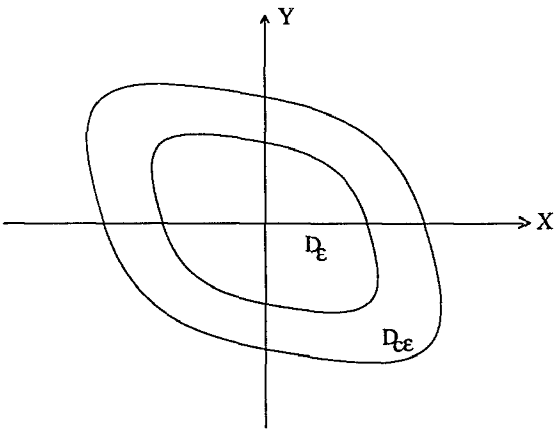

Fig. 5. Level sets D_ϵ and $D_{c\epsilon}$ for the slope function s of an objective function g.

Archetti *et al.* in [1] proposed a framework for global optimisation using random sampling of the objective function. They applied the theory of extreme value distributions to global maximisation problems, using the measure of the level set of the objective function to investigate the Gnedenko condition. They obtained a condition on the objective function which ensures that the measure of the level set possesses a certain limit property. This in turn ensures that the distribution of objective function values under random sampling satisfies the Gnedenko condition.

Our method is similar to the procedure of Archetti *et al.*, but here the function to be maximised is the absolute slope function s of the original objective function g, not g itself. Our aim has been to find a general condition on g which ensures that the slope distribution satisfies the Gnedenko condition. The next lemma, expressed here in terms of the absolute slope function s given in Theorem 2, is due to Archetti *et al.* [1,§2 and Theorems 6 and 7]. It presents conditions on the absolute slope function s which ensure that the cumulative distribution F for the absolute value of slopes satisfies the Gnedenko condition. We are then able to prove Theorem 2 by showing that the conditions of that theorem ensure that those of Lemma 3 hold.

LEMMA 3. *(1) Suppose that* $s(x,y) = M + Q_{2p}((x,y)-(x^*,y^*)) + R(||(x,y)-(x^*,y^*)||^{2p})$, *where* (x^*,y^*) *is an interior point of* D *at which* $s(x,y)$ *attains its unique global maximum* M, $Q_{2p}((x,y)-(x^*,y^*))$ *is a negative (for* $(x,y) \neq (x^*,y^*)$*) homogeneous polynomial of degree* $2p$, *for* p *a natural number and* $R(\epsilon)/\epsilon \to 0$ *as* $\epsilon \to 0$. *Then* $\lim_{\epsilon\to 0^+} \mu(D_\epsilon)/\epsilon^\alpha$ *exists and is finite and positive with* $\alpha = 1/p$. *Here* μ *denotes Lebesgue measure on* $\mathbf{R}^2$.

(2) If $\mu(D_\epsilon)$ *is such that*

$$\lim_{\epsilon\to 0+} \mu(D_\epsilon)/\epsilon^\alpha > 0$$

for some $\alpha > 0$, *then the Gnedenko condition holds for* $F(t) = \mu\{(x,y) \in D : s(x,y) \leq t\}/\mu(D)$, *with* $k = \alpha$.

The key to proving Lemma 3(1) is the construction of level sets D_ϵ^1 and D_ϵ^2 such that $D_\epsilon^1 \subset D_\epsilon \subset D_\epsilon^2$. It is then shown that both $\lim_{\epsilon\to 0^+} \mu(D_\epsilon^1)/\epsilon^\alpha$ and $\lim_{\epsilon\to 0^+} \mu(D_\epsilon^2)/\epsilon^\alpha$ exist and are equal. We use Lemma 3 now to prove Theorem 2.

Proof of Theorem 2. Without loss of generality we can assume that $z = g(z) = 0$ and $g'(z) = M$. (This follows since we can let $h(x) = g(x-z) - g(z)$ and prove the result for $\pm h$ as needed.) We can then write $g(x)$ as

$$g(x) = Mx + a_{2p+1}x^{2p+1} + o(x^{2p+1})$$

by assumption, where $a_{2p+1} = g^{(2p+1)}(0)/(2p+1)!$. Since $g'(0)$ is assumed to be the unique maximum of g', we must have $a_{2p+1} < 0$. Also $\mathrm{diam}(D_\epsilon) \to 0$ as $\epsilon \to 0$ else the uniqueness assumption is contradicted.

Since $g'(0) = M > 0$, g is increasing near 0, so we have

$$\frac{|g(x)-g(y)|}{|x-y|} = \frac{g(x)-g(y)}{x-y}$$

for any pair $x \neq y$ sufficiently close to 0. We prove that $(g(x)-g(y))/(x-y)$ is a function which satisfies the condition of Lemma 3(1).

Since $\mathrm{diam}(D_\epsilon) \to 0$ as $\epsilon \to 0$, for ϵ sufficiently small both D_ϵ and $D_{c\epsilon}$ are contained in D. Consider a point (x,y) in D_ϵ. Then the rearranged Taylor expansion of $g(x)$ about y yields

$$(g(x)-g(y))/(x-y)$$

$$= g'(y) + \frac{g''(y)}{2!}(x-y) + \cdots + \frac{g^{(2p+1)}(y)}{(2p+1)!}(x-y)^{2p} +$$
$$\frac{g^{(2p+2)}(\zeta_0)}{(2p+2)!}(x-y)^{2p+1}$$

where ζ_0 is between x and y. Writing the Taylor expansion of $g(y)$, $g'(y), \ldots$, $g^{(2p+1)}(y)$ about 0, we have

$$\begin{aligned}
&(g(x)-g(y))/(x-y)\\
&= M + \frac{g^{(2p+1)}(0)}{(2p)!}y^{2p} + \frac{g^{(2p+2)}(\zeta_1)}{(2p+1)!}y^{2p+1}\\
&+\frac{1}{2!}[\frac{g^{(2p+1)}(0)}{(2p-1)!}y^{2p-1} + \frac{g^{(2p+2)}(\zeta_2)}{(2p+1)!}y^{2p}](x-y) + \cdots\\
&+\frac{1}{(2p+1)!}[\frac{g^{(2p+1)}(0)}{0!} + \frac{g^{(2p+2)}(\zeta_{2p+1})}{(2p+1)!}y](x-y)^{2p} +\\
&\frac{g^{(2p+2)}(\zeta_0)}{(2p+2)!}(x-y)^{2p+1}\\
&= M + \frac{g^{(2p+1)}(0)}{(2p+1)!}\sum_{k=0}^{2p}\frac{(2p+1)!}{(2p-k)!(k+1)!}y^{2p-k}(x-y)^k +\\
&R(\zeta_0, \zeta_1, \ldots, \zeta_{2p+1}, x, y)
\end{aligned}$$

where $R(\zeta_0, \zeta_1, \ldots, \zeta_{2p+1}, x, y)$ equals

$$\frac{g^{(2p+2)}(\zeta_1)}{(2p+1)!}y^{2p+1} + \cdots + \frac{g^{(2p+2)}(\zeta_{2p+1})}{(2p+1)!}y(x-y)^{2p}$$
$$+\frac{g^{(2p+2)}(\zeta_0)}{(2p+2)!}(x-y)^{2p+1}$$

and $\zeta_1, \ldots, \zeta_{2p+1}$ lie between 0 and y. Now

$$\begin{aligned}
\sum_{k=0}^{2p}\frac{(2p+1)!}{(2p-k)!(k+1)!}y^{2p-k}(x-y)^k &= \sum_{k=0}^{2p}C_{k+1}^{2p+1}y^{2p-k}\sum_{i=0}^{k}C_i^k x^i y^{k-i}(-1)^{k-i}\\
&= \sum_{i=0}^{2p}x^i y^{2p-i}\sum_{k=i}^{2p}C_{k+1}^{2p+1}C_i^k(-1)^{k-i}
\end{aligned}$$

By using the generalised Vandermonde convolution formula of recurrence [11, p. 8] it can be shown that $\sum_{k=i}^{2p} C_{k+1}^{2p+1}C_i^k(-1)^{k-i} = 1$, for $i = 0, 1, \ldots, 2p$. Therefore

$$(g(x)-g(y))/(x-y)$$

$$= M + \frac{g^{(2p+1)}(0)}{(2p+1)!} \sum_{i=0}^{2p} x^i y^{2p-i} + R(\zeta_0, \zeta_1, ..., \zeta_{2p+1}, x, y)$$

It is now readily checked that $s(x, y)$ is a function of the form described in Lemma 3(1). Lemma 3(1) thus shows that $\lim_{\epsilon \to 0^+} \mu(D_\epsilon)/\epsilon^\alpha$ exists and is finite and positive with $\alpha = 1/p$. This limiting result rests on the fact that we sample uniformly from a δ-strip around the diagonal of $[a, b] \times [a, b]$, and that s assumes its maximum value on the diagonal. Lemma 3(2) then gives that the Gnedenko condition holds for $F(t) = \mu\{(x, y) \in D : s(x, y) \leq t\}/\mu(D)$ with $k = 1/p$. In a neighbourhood of M this coincides with the cdf of s, so Theorem 2 follows.

The requirement that the first non-zero derivative at z should occur for an odd order ensures that there is no contradiction to the assumption that $|g'(z)| = M = \max_{x \in (a,b)} |g'(x)|$. It can be confirmed that the three test functions discussed in Section 3 satisfy the conditions of Theorem 2. We remark that the proof for an objective function g containing only linear and cubic terms is much easier, as in this case the level sets $D_{c\epsilon}$ and D_ϵ are elliptical.

The following propositions, based on Theorem 2, show that the assumption that g' assumes its maximum absolute value at a unique point or an interior point can be removed. Both propositions can be proved in a straightforward fashion.

PROPOSITION 4. *Let $g \in C^{2p+2}(a, b)$ with finitely many points $z_1, z_2, \ldots, z_m \in (a, b)$ such that $|g'(z_i)| = sup_{x \in (a,b)} |g'(x)| = M > 0$ and $g''(z_i) = ... = g^{2p}(z_i) = 0$ but $g^{(2p+1)}(z_i) \neq 0$, for $i = 1, 2, ..., m$. Given $\delta > 0$, let (x, y) be chosen uniformly on $\{(x, y) \in [a, b] \times [a, b] : |x - y| < \delta\}$, $s = |g(x) - g(y)|/|x - y|$ and F denote the cumulative distribution function of s. Then the Gnedenko condition holds for F.*

PROPOSITION 5. *Suppose that $g \in C^{2p+2}(a, b)$ with $\lim_{\epsilon \to 0^+} (g(b) - g(b - \epsilon))/\epsilon = M > 0$, $g''(b) = g'''(b) = ... = g^{(2p+1)}(b) = 0$, $g^{(2p+1)}(b) < 0$, where $g^{(k)}(b)$ denotes the left kth order derivative, and $|g'(x)| < M$ on (a, b). Given $\delta > 0$, let (x, y) be chosen uniformly on $\{(x, y) \in [a, b] \times [a, b] : |x - y| < \delta\}$, $s = |g(x) - g(y)|/|x - y|$ and F denote the cumulative distribution function of s. Then the Gnedenko condition holds for F.*

5. Summary and Future Directions

In this paper we have presented a stochastic method for estimating the Lipschitz constant of a function of a single variable. The method is clearly successful, but computationally intensive. Three directions for further research suggest themselves. Firstly, how does the idea used here fare for functions of more than one variable? The authors have some partial results which will be developed and presented in another paper. Secondly, how should the slope sample size n and the size of the samples of maximum absolute slope m be chosen? Evidently the answer depends on the objective function g, its domain and the value of δ. Thirdly, it remains to interlock the Lipschitz constant estimator with Lipschitz based optimisation algorithms.

Acknowledgements

The authors heartily thank a referee for the suggestion to modify the sampling rule to that described in this paper.

References

1. Archetti, F., Betrò, B. and Steffè, S. (1976), "A theoretical framework for global optimisation via random sampling", in *Studies on Stochastic Optimization* (Quaderno dei gruppi di recerca mathematica del C.N.R.), University of Pisa.
2. de Haan, L. (1981), "Estimation of the minimum of a function using order statistics", *Journal of the American Statistical Association* **76**, 467–469.
3. Galambos, J. (1987), *The Asymptotic Theory of Extreme Order Statistics*, second edition, Krieger, Florida.
4. Gnedenko, V.B. (1943), "Sur la distribution limite du terme maximum d'une série aléatoire", *Annals of Mathematics* **44**, 423–453.
5. Gourdin, E., Hansen, P. and Jaumard, B. (1994), "Finding maximum likelihood estimators for the three-parameter Weibull distribution", *J. Global Optimization* **5**, 373–397.
6. Hansen, P., Jaumard, B. and Lu, S.H. (1991), "On the number of iterations of Piyavskii's global optimization algorithm", *Mathematics of Operations Research* **16**, 334–350.
7. Hansen, P., Jaumard, B. and Lu, S.H. (1992), "On the use of estimates of the Lipschitz constant in global optimisation", *Journal of Optimization Theory and Applications* **75**, 195–200.
8. Horst, R. and Pardalos, P.M. (editors) (1995), *Handbook of Global Optimization*, Kluwer Academic Publishers.
9. Mladineo, R.H. (1986), "An algorithm for finding the global maximum of a multimodal, multivariate function", *Mathematical Programming* **34**, 188–200.
10. Moore, R.E. (1966), *Interval Analysis*, Prentice-Hall, Englewood Cliffs, N.J.
11. Panchang, V.G. and Gupta, R.C. (1989), "On the determination of three-parameter Weibull MLE's, *Communication in Statistics B: Simulation and Computation* **18**, 1037–1057.
12. Riordan, J. (1968), *Combinatorial Identities*, Wiley, New York.
13. Shubert, B.O. (1972), "A sequential method seeking the global maximum of a function", *SIAM Journal on Numerical Analysis* **9**, 379–388.
14. Strongin, R.G. (1973), "On the convergence of an algorithm for finding a global extremum", *Engineering Cybernetics* **11**, 549–555.
15. Törn, A. and Žilinskas, A. (1989), "Global optimization", *Lecture Notes in Computer Science 350*, Springer-Verlag, Berlin.
16. Zanakis, S.H. and Kyparisis, J. (1986), "A review of maximum likelihood estimation methods for the three-parameter Weibull distribution", *Journal of Statistical Computation and Simulation* **25**, 53-73.