The Annals of Applied Probability
1997, Vol. 7, No. 3, 711–730

AVERAGE PERFORMANCE OF A CLASS OF ADAPTIVE ALGORITHMS FOR GLOBAL OPTIMIZATION

BY JAMES M. CALVIN[1]

New Jersey Institute of Technology

We describe a class of adaptive algorithms for approximating the global minimum of a continuous function on the unit interval. The limiting distribution of the error is derived under the assumption of Wiener measure on the objective functions. For any $\delta > 0$, we construct an algorithm which has error converging to zero at rate $n^{-(1-\delta)}$ in the number of function evaluations n. This convergence rate contrasts with the $n^{-1/2}$ rate of previously studied nonadaptive methods.

1. Introduction. The purpose of this paper is to characterize the average performance of a class of adaptive global minimization algorithms under the Brownian motion model for the objective function. The object of a global minimization method is to approximate the global minimum f^* of a function f, and sometimes also a location t^* where the minimum is attained. We take f to be a continuous function defined on the unit interval, and adopt the framework that the approximation is based on observation of the function value at sequentially selected points in the unit interval. That is, the searcher chooses points $t_1, t_2, \ldots, \in [0,1]$ and forms an approximation (t_n^*, f_n^*) to (t^*, f^*) based on $\{(t_i, f(t_i)): i = 1, 2, \ldots, n\}$. An adaptive algorithm chooses each new point t_{n+1} as a function of $\{(t_i, f(t_i)): i = 1, 2, \ldots, n\}$, while a nonadaptive algorithm chooses each point independently of the function values. We allow the possibility that the algorithm uses auxiliary randomness in the choice of observation sites.

We consider a class of adaptive algorithms that use only limited past information. For any $\delta > 0$, we construct an algorithm for which the error converges to 0 at rate $n^{-(1-\delta)}$, in contrast to the $n^{-1/2}$ rate characteristic of nonadaptive algorithms. We also identify the limiting distribution of the normalized error. The improved efficiency relative to nonadaptive algorithms comes from using information from past observations to concentrate the search in decreasing subregions of the minimizer.

Several methods have been used to compare the performance of different global optimization algorithms. In this paper we will be concerned with the average performance criterion. The idea is to regard f as the sample path of a stochastic process and then classify algorithms based on the average error in their approximations. This method has been used to study the average

Received March 1995; revised March 1997.
[1]Research supported in part by NSF Grants DDM-9010770 and DMI-9500173.
AMS 1991 *subject classifications.* Primary 60J65, 68Q25; secondary 11Y16, 73K40.
Key words and phrases. Brownian motion, global optimization, average complexity.

performance of nonadaptive algorithms in the case where f is taken to be a sample path of a Brownian motion process. Ritter (1990) showed that for the best non-adaptive method, the average error decreases at rate $n^{-1/2}$ in the number of observations n. Calvin (1995) compared the average error for deterministic uniformly spaced observations with the expected error associated with random uniform sampling. Al-Mharmah and Calvin (1996) show that the optimal nonadaptive sampling density for approximating the error for Brownian motion is a beta distribution. Asmussen, Glynn and Pitman (1995) describe the limiting distribution of the normalized error for the deterministic uniform grid. Calvin and Glynn (1996) extend many of these results to a more general class of diffusions.

Several authors, including Kushner (1964), Žilinskas (1976), and Archetti and Betrò (1979), have constructed adaptive optimization algorithms based on the Brownian motion model for the objective function. Most of the algorithms constructed are significantly more complex than the algorithms considered in this paper, and their performance is consequently harder to characterize.

In Section 2 we establish notation and describe the basic approach, as well as summarize some relevant facts about nonadaptive algorithms. In Section 3 we derive the limiting joint distribution of the error in the function value and function arguments under uniform sampling. These results are used to establish the convergence characteristics of the class of adaptive algorithms in Section 4. The results of some numerical experiments are presented in Section 5.

2. Notation and the basic idea. Let $(B(t)\colon 0 \le t \le 1)$ be a standard Brownian motion defined on a probability space $(\Omega_1, \mathscr{F}_1, P_1)$, and let $\{U_1, U_2, \ldots\}$ be a sequence of independent, uniform $(0, 1)$ random variables defined on a probability space $(\Omega_2, \mathscr{F}_2, P_2)$. Set $(\Omega, \mathscr{F}, P) = (\Omega_1 \times \Omega_2, \mathscr{F}_1 \times \mathscr{F}_2, P_1 \times P_2)$. Let B^* denote the global minimum of the Brownian motion, and t^* the (first) location where B^* is attained. The minimizer is almost surely unique, so the issue of which local minimizer to assign to t^* is not important for our results. Denote by u_n^* the value U_i $(1 \le i \le n)$ such that $B(U_i) \le B(U_j)$, $1 \le j \le n$, and U_i is the smallest value with this property. Let Δ_n denote the difference between the smallest value seen in the first n observations and the global minimum and let Γ_n denote the difference between the global minimizer and the minimizer of the first n observations. The formal definitions are as follows (see Figure 1):

$$B^* = \min_{0 \le t \le 1} B(t), \tag{2.1}$$

$$t^* = \inf\{t\colon B(t) = B^*\}, \tag{2.2}$$

$$u_n^* = \min\{U_i\colon B(U_i) \le B(U_j),\ 1 \le j \le n\}, \tag{2.3}$$

$$\Delta_n = \Delta_n(\omega_1, \omega_2) = \min_{1 \le i \le n} B(\omega_1, U_i(\omega_2)) - B^*(\omega_1), \tag{2.4}$$

$$\Gamma_n = \Gamma_n(\omega_1, \omega_2) = u_n^*(\omega_1, \omega_2) - t^*(\omega_1). \tag{2.5}$$

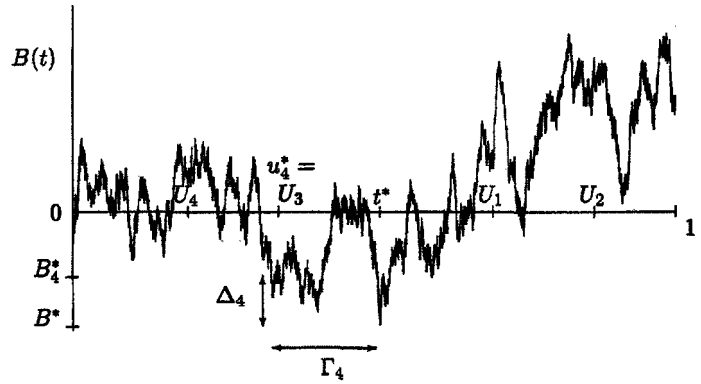

FIG. 1. *Brownian path and error variables.*

A simple nonadaptive global minimization algorithm consists of choosing the observation sites independently, uniformly over $[0, 1]$, and taking the smallest observed value as the approximation to the global minimum; Δ_n is the corresponding approximation error. The following result, proved in Calvin and Glynn (1997), characterizes the average performance of this nonadaptive algorithm. (This result will also emerge as a corollary of Theorem 3.1 below.)

THEOREM 2.1. *For* $y > 0$,

$$P\left(\sqrt{n}\,\Delta_n \le y\right) \to \tanh^2\left(\sqrt{2}\,y\right) \tag{2.6}$$

as $n \to \infty$.

Note that there are two sources of uncertainty, the random function and the random observation points, and the probability in Theorem 2.1 is the product of the two probabilities; that is, the error is averaged over all sample paths. If the sample path is fixed, then the error fails to converge in distribution for almost all sample paths [see Calvin and Glynn (1997) for details]. One can therefore think of Theorem 2.1 as giving an approximation to the error when averaged over many independent optimizations, but it says nothing about what happens when a fixed path is optimized.

In order to improve on the performance of the basic nonadaptive scheme just described, it is necessary to concentrate search effort near the minimizer. This will be accomplished by algorithms constructed according to the following general framework. On the $(n + 1)$st iteration, with probability $1/2$ we choose the observation site uniformly over the entire unit interval, and with probability $1/2$ we choose the site uniformly over a small subinterval centered at $\hat{t}_n^*$, where $\hat{t}_n^*$ is the location of the smallest observed value over those points chosen uniformly over the entire interval. We emphasize that $\hat{t}_n^*$ is *not* the location of the smallest value observed in the first n observations; rather it is the location of the smallest of those (on average $n/2$) observations chosen uniformly over the unit interval (i.e., we exclude the outcomes of the

local searches). The width of the interval of the local search decreases over time, so that the local searches become more concentrated as the search progresses.

Formally, let $\{\beta_i: i \geq 1\}$ be a sequence of independent Bernoulli (1/2) random variables defined on $(\Omega_2, \mathcal{F}_2, P_2)$, independent of the $\{U_i\}$. Let $\{a_n\}$ be a decreasing (deterministic) sequence of positive numbers. The algorithms have the following form:

Set $t_1 = \hat{t}_1^* = t_1^* \leftarrow U_1$, $\hat{B}_1^* = B_1^* \leftarrow B(t_1)$;
For $k = 1, 2, \ldots, n-1$,
  If $\beta_{k+1} = 0$, {perform global search}
    Set $t_{k+1} \leftarrow U_{k+1}$;
    If $B(t_{k+1}) < B_k^*$, then set $t_{k+1}^* \leftarrow t_{k+1}$, $B_{k+1}^* \leftarrow B(t_{k+1})$;
    If $B(t_{k+1}) < \hat{B}_k^*$, then set $\hat{t}_{k+1}^* \leftarrow t_{k+1}$, $\hat{B}_{k+1}^* \leftarrow B(t_{k+1})$;
  Else if $\beta_{k+1} = 1$, {perform local search}
    Set $t_{k+1} \leftarrow \hat{t}_k^* + a_{k+1}(U_{k+1} - \frac{1}{2})$;
    If $B(t_{k+1}) < B_k^*$, then set $t_{k+1}^* \leftarrow t_{k+1}$, $B_{k+1}^* \leftarrow B(t_{k+1})$.

In the second to last step, it is possible that for small k, t_{k+1} will be set to a value outside the unit interval. In this case, it is understood that a new realization U_k is chosen and the step repeated. This will not be important for our limit results, so we prefer not to complicate the algorithm description.

After the last step, B_n^* is our approximation to B^* and t_n^* is our approximation to t^*. We will be interested in the quality of the approximations produced by the algorithm as the number of steps $n \to \infty$.

It is easy to see that this algorithm is consistent (for any choice of sequence $\{a_n\}$) in the sense that the error converges to zero P_2-a.s. for any Brownian path. The only information from the past maintained by the algorithm consists of $\hat{t}_n^*$, $\hat{B}_n^*$, t_n^*, and B_n^*. The sequence $\{\hat{t}_n^*; n \geq 1\}$ is a Markov chain with values in $[0, 1]$, though note that $\{t_n; n \geq 1\}$ is *not* a Markov chain. A common definition of Markov algorithm is that the $(n+1)$st point has a distribution that depends on the nth point and the function value at the nth point, and not on previous observations; simulated annealing is an example of this class of algorithm. [See Zhigljavsky (1991) for this definition and several examples.] Our algorithm therefore does not fit this definition of Markovian algorithm, since in addition to $(t_n, f(t_n))$, the distribution of the next point will also depend on $\hat{t}_n^*$.

To complete the description of the algorithm, it remains to determine a choice of the sequence $\{a_n\}$. If a_n goes to 0 too fast relative to the speed at which $\hat{t}_n^* - t^*$ goes to 0, then the local search will tend to concentrate in subregions away from t^*. On the other hand, if a_n goes to 0 too slowly, then the performance gain relative to uniform sampling will be small. In particular, if $a_n \downarrow a > 0$, then the asymptotic performance will be of the same order as for uniform sampling. To determine an appropriate rate, it is necessary to know the rate at which $\hat{t}^* - t^*$ converges to 0, which will be determined in the next section.

3. Joint distribution of error variables. We begin by analyzing the joint distribution of the error variables (Δ_n, Γ_n) under uniform sampling. As shown in Theorem 1, $\sqrt{n}\,\Delta_n$ converges in distribution as $n \to \infty$. In this section we will show (in Theorem 3.2) that $n(u_n^* - t^*)$ converges in distribution, thus giving the convergence rate needed to determine the $\{a_n\}$ sequence for our adaptive algorithm. In fact, we will derive the limiting joint distribution of $(\sqrt{n}\,(B_n^* - B^*),\ n(u_n^* - t^*))$ as $n \to \infty$. As noted in the introduction, the error in approximating the location of the minimizer is also important for the global minimization method, so the results of this section are of independent interest.

We begin by describing some random variables that will be needed in the description of the limiting distribution of $(\sqrt{n}\,(B_n^* - B^*),\ n(u_n^* - t^*))$.

Let R_1 and R_2 be two independent three-dimensional Bessel processes, and define a "two-sided Bessel process" R by

$$R(t) = \begin{cases} R_1(t), & \text{if } t \geq 0, \\ R_2(-t), & \text{if } t < 0. \end{cases} \tag{3.1}$$

(The three-dimensional Bessel process is the diffusion that is identical in law to the modulus of a three-dimensional Brownian motion.) Let $\{\tau_i\}$ be an enumeration of the points of a Poisson point process on the line with unit intensity, independent of R, and set

$$\Delta = \inf_i R(\tau_i), \qquad \Gamma = \inf\{|\tau_i|\colon R(\tau_i) = \Delta\}.$$

THEOREM 3.1. *The joint Laplace transform of* (Δ, Γ) *is given by*

$$\begin{aligned} &\int_{[0,\infty)}\int_{[0,\infty)} e^{-\alpha t - \beta y} P(\Gamma \in dt,\ \Delta \in dy) \\ &\quad = 2\int_0^\infty \frac{\sqrt{1+\alpha}\exp\left(-\beta_x/\sqrt{2}\right)}{\sqrt{1+\alpha}\cosh\left(x\sqrt{1+\alpha}\right) + \sqrt{\alpha}\sinh\left(x\sqrt{1+\alpha}\right)} \frac{\sinh(x)}{\cosh^2(x)}\,dx. \end{aligned} \tag{3.2}$$

PROOF. Since R and the point process are symmetric about 0, we first derive the joint distribution of the error variables to the right of 0, say (Δ^R, Γ^R).

Fix $y > 0$, and let $L_y = \sup\{t\colon R_1(t) = y\}$ be the last time that the Bessel process R_1 hits the level y. Then [see Revuz and Yor (1991), page 294],

$$\{R_1(L_y - t)\colon 0 \leq t \leq L_y\} =_{\mathcal{D}} \{X(t)\colon 0 \leq t \leq T_0\}, \tag{3.3}$$

where X is a Brownian motion starting at y and T_0 is the hitting time of 0. Therefore, given that $\Delta^R = y$, the conditional distribution of Γ^R is that of the hitting time to 0 of a Brownian motion starting at y, conditioned not to be at a level below y at the points of a Poisson process. To determine this distribution it is convenient to consider the Brownian motion killed at unit

rate below level y; that is, consider the diffusion Y with generator A, where for $f \in C^2$,

$$Af(x) = \tfrac{1}{2}f''(x) - I_{\{x \le y\}}f(x).$$

Let P'_y denote the distribution of the killed Brownian motion Y starting at y (with E'_y the corresponding expectation operator), and let ζ denote the "death time" of Y. Denote by P_y the distribution of the Brownian motion X starting at y. Then if $\{\tau_i\}$ are the points of the Poisson point process,

$$P_y\Big(\min_{0 \le \tau_i \le T_0} X(\tau_i) > y\Big) = P'_y(T_0 < \zeta).$$

For each $\alpha > 0$, there exists a unique (up to multiplication by a positive constant) strictly positive decreasing solution f_α to

$$Af_\alpha(x) = \alpha f_\alpha(x), \tag{3.4}$$

such that $\lim_{x \to \infty} f_\alpha(x) = 0$ [see Itô and McKean (1965), Section 2.6]. That is, f_α satisfies

$$\tfrac{1}{2}(f'_\alpha(d) - f'_\alpha(c)) = \alpha\int_{[c,d)} f_\alpha(x)\,dx + \int_{[c,d)} I_{\{x \le y\}}f_\alpha(x)\,dx, \qquad c < d. \tag{3.5}$$

Then for $x > 0$,

$$E'_x\exp(-\alpha T_0) = \frac{f_\alpha(x)}{f_\alpha(0)} \tag{3.6}$$

[see Itô and McKean (1965), Section 4.6]. Given that $\Delta^R = y$, the conditional distribution of Γ^R is the P'_y distribution of T_0 given that 0 is reached before the process is killed. To determine this, we use (3.6) to first find the Laplace transform of the defective random variable T_0 under P'_y.

For fixed $\alpha > 0$, (3.5) has a decreasing solution given by

$$f_\alpha(x) = \begin{cases} \delta\cosh\big(x\sqrt{2(1+\alpha)}\big) + \beta\sinh\big(x\sqrt{2(1+\alpha)}\big), & \text{if } x \le y, \\ \gamma\cosh\big(x\sqrt{2\alpha}\big) + \xi\sinh\big(x\sqrt{2\alpha}\big), & \text{if } x > y, \end{cases}$$

for constants $\delta, \beta, \gamma, \xi$. Taking the solution with $f_\alpha(0) = 1$ implies $\delta = 1$. The requirement that $\lim_{x \to \infty} f_\alpha(x) = 0$ implies that $\gamma = -\xi$. In order for f_α and f'_α to be continuous at y [as required for a solution of (3.5)], we need that

$$\beta = \frac{-\sqrt{2(1+\alpha)}\sinh\big(y\sqrt{2(1+\alpha)}\big) - \sqrt{2\alpha}\cosh\big(y\sqrt{2(1+\alpha)}\big)}{\sqrt{2(1+\alpha)}\cosh\big(y\sqrt{2(1+\alpha)}\big) + \sqrt{2\alpha}\sinh\big(y\sqrt{2(1+\alpha)}\big)}$$

and

$$\gamma = \frac{\sqrt{2(1+\alpha)}\exp\big(y\sqrt{2\alpha}\big)}{\sqrt{2(1+\alpha)}\cosh\big(y\sqrt{2(1+\alpha)}\big) + \sqrt{2\alpha}\sinh\big(y\sqrt{2(1+\alpha)}\big)}.$$

Therefore,

$$E'_y \exp(-\alpha T_0) = \frac{f_\alpha(y)}{f_\alpha(0)} = \gamma \exp(-y\sqrt{2\alpha})$$

$$= \frac{\sqrt{2(1+\alpha)}}{\sqrt{2(1+\alpha)}\cosh\left(y\sqrt{2(1+\alpha)}\right) + \sqrt{2\alpha}\sinh\left(y\sqrt{2(1+\alpha)}\right)}.$$

Consequently,

$$\int_{[0,\infty)} e^{-\alpha t} P'_y(T_0 \in dt)$$

$$= \frac{\sqrt{2(1+\alpha)}}{\sqrt{2(1+\alpha)}\cosh\left(y\sqrt{2(1+\alpha)}\right) + \sqrt{2\alpha}\sinh\left(y\sqrt{2(1+\alpha)}\right)},$$

and

$$P'_y(T_0 < \infty) = \frac{1}{\cosh(y\sqrt{2})}. \tag{3.7}$$

Therefore,

$$\int_{[0,\infty)} e^{-\alpha t} P'_y(T_0 \in dt | T_0 < \infty)$$

$$= \frac{\sqrt{1+\alpha}\cosh(y\sqrt{2})}{\sqrt{1+\alpha}\cosh\left(y\sqrt{2(1+\alpha)}\right) + \sqrt{\alpha}\sinh\left(y\sqrt{2(1+\alpha)}\right)}. \tag{3.8}$$

By (3.3),

$$P(\Gamma^R \in dt | \Delta^R = y) = P'_y(T_0 \in dt | T_0 < \infty).$$

Therefore,

$$\begin{aligned} P(\Gamma \in dt,\ \Delta \in dy) &= P(\Gamma \in dt | \Delta = y) P(\Delta \in dy) \\ &= P(\Gamma^R \in dt | \Delta^R = y) P(\Delta \in dy), \end{aligned} \tag{3.9}$$

where the last equality follows from symmetry around 0 and conditioning on whether the minimum occurs to the left or right of 0. To obtain the marginal distribution of Δ, note that by (3.7) and the fact that Δ is the minimum of two independent copies of Δ^R,

$$P(\Delta > y) = P(\Delta^R > y)^2 = P_y(T_0 < \infty)^2 = \frac{1}{\cosh^2(y\sqrt{2})},$$

and therefore

$$P(\Delta \in dy) = \frac{2\sqrt{2}\sinh(y\sqrt{2})}{\cosh^3(y\sqrt{2})}\, dy. \tag{3.10}$$

Combining (3.10) with (3.9) gives

$$P(\Gamma \in dt,\ \Delta \in dy) = P'_y(T_0 \in dt | T_0 < \infty) \frac{2\sqrt{2}\sinh(y\sqrt{2})}{\cosh^3(y\sqrt{2})}\, dy.$$

Using (3.8) then gives

$$\begin{aligned}
&\int_{[0,\infty)}\int_{[0,\infty)} e^{-\alpha t-\beta y}P(\Gamma \in dt,\ \Delta \in dy)\\
&\quad = \int_{t=0}^{\infty}\int_{y=0}^{\infty} e^{-\alpha t-\beta y}P_y'(T_0 \in dt | T_0 < \infty)\frac{2\sqrt{2}\sinh(y\sqrt{2})}{\cosh^3(y\sqrt{2})}\,dy\,dt\\
&\quad = \int_{y=0}^{\infty}\frac{e^{-\beta y}\sqrt{1+\alpha}\cosh(y\sqrt{2})}{\sqrt{1+\alpha}\cosh\bigl(y\sqrt{2(1+\alpha)}\bigr)+\sqrt{\alpha}\sinh\bigl(y\sqrt{2(1+\alpha)}\bigr)}\\
&\qquad \times\frac{2\sqrt{2}\sinh(y\sqrt{2})}{\cosh^3(y\sqrt{2})}\,dy\\
&\quad = 2\int_{x=0}^{\infty}\frac{\sqrt{1+\alpha}\exp(-\beta_x/\sqrt{2})}{\sqrt{1+\alpha}\cosh(x\sqrt{1+\alpha})+\sqrt{\alpha}\sinh(x\sqrt{1+\alpha})}\frac{\sinh(x)}{\cosh^2(x)}\,dx,
\end{aligned}$$

which completes the proof. □

The Laplace transform of the distance to the minimizer is given by

$$\begin{aligned}
&\int_{[0,\infty)} e^{-\alpha t}P(\Gamma \in dt)\\
&\quad = \int_{y=0}^{\infty}\frac{2\sqrt{2}\sqrt{1+\alpha}\cosh(y\sqrt{2})}{\sqrt{1+\alpha}\cosh\bigl(y\sqrt{2(1+\alpha)}\bigr)+\sqrt{\alpha}\sinh\bigl(y\sqrt{2(1+\alpha)}\bigr)}\\
&\qquad \times\frac{\sinh(y\sqrt{2})}{\cosh^3(y\sqrt{2})}\,dy\\
&\quad = 2\int_{x=0}^{\infty}\frac{\sqrt{1+\alpha}}{\sqrt{1+\alpha}\cosh(x\sqrt{1+\alpha})+\sqrt{\alpha}\sinh(x\sqrt{1+\alpha})}\frac{\sinh(x)}{\cosh^2(x)}\,dx.
\end{aligned}$$

While the Laplace transform is awkward to invert, it serves to establish that $E(\Gamma) = \infty$.

For the remainder of this paper, we turn our attention to more general algorithms than independent uniform sampling, and we will adjust our notation slightly. For the rest of this section, $\{t_i\}$ is a sequence of observation sites generated by an arbitrary algorithm. We will still use Δ_n, Γ_n to denote the error variables, as before, but now they represent the errors based on the sequence $\{t_i\}$ and not the particular sequence $\{U_i\}$. Specifically,

$$\Delta_n = \Delta_n(\omega_1, \omega_2) = \min_{1\le i\le n} B(\omega_1, t_i(\omega_2)) - B^*(\omega_1), \tag{3.11}$$

$$\Gamma_n = \Gamma_n(\omega_1, \omega_2) = t_n^*(\omega_1, \omega_2) - t^*(\omega_1). \tag{3.12}$$

The next result shows that if the points can be normalized so as to converge to a Poisson point process, then the normalized error converges in

distribution to (Δ, Γ). Taking $c_n = n$ and $t_i = U_i$ gives the result promised in the first paragraph of this section for the limiting distribution of $(\sqrt{n}(B_n^* - B^*), n(u_n^* - t^*))$ as $n \to \infty$.

THEOREM 3.2. *Suppose that $\{c_n\}$ is a sequence of positive real numbers, $c_n \uparrow \infty$, and define a sequence of point processes on the line by*

$$N_n(A) = \sum_{i=1}^{n} I_{\{c_n(t_i - t^*) \in A\}}$$

for Borel sets A. Suppose that $N_n \to_{\mathscr{D}} N$ as $n \to \infty$, where N is a Poisson point process on the line with unit intensity, independent of B. That is, the law of N_n converges weakly to the law of N_n in $M_p(\mathbb{R})$, the space of point measures on $\mathbb{R}$. Then

$$\left(\sqrt{c_n}(B_n^* - B^*), c_n(t_n^* - t^*)\right) = \left(\sqrt{c_n}\Delta_n, c_n\Gamma_n\right) \to_{\mathscr{D}} (\Delta, \Gamma)$$

as $n \to \infty$, where the distribution of (Δ, Γ) is given in Theorem 3.1.

PROOF. The idea of the proof is contained in Asmussen, Glynn and Pitman (1995). Fix $T > 0$ and let

$$\Delta_n^T = \inf\{B(t_i) - B^*: 1 \le i \le n, |t_i - t^*| \le Tc_n^{-1}\},$$
$$\tilde{\Delta}_n^T = \inf\{B(t_i) - B^*: 1 \le i \le n, |t_i - t^*| > Tc_n^{-1}\},$$
$$\Gamma_n^T = \inf\{|t_i - t^*|: B(t_i) - B^* = \Delta_n^T, |t_i - t^*| \le Tc_n^{-1}\}.$$

Let

$$Z_n(t) = \begin{cases} \sqrt{c_n}(B(t^* + t/c_n) - B^*), & \text{if } 0 \le t \le T, \\ \sqrt{c_n}(B(t^* - t/c_n) - B^*), & \text{if } -T \le t < 0. \end{cases} \tag{3.13}$$

Conditional on $B^* = b$, $t^* = s$, $B(1) = y - b$, Z_n converges in distribution to R (restricted to $[-T, T]$) in $C[-T, T]$; this is Lemma 2 in Asmussen, Glynn and Pitman (1995) and Brownian scaling. Define the map $\Psi: C[-T, T] \times M_p([-T, T]) \to \mathbb{R}^2$ by

$$\Psi(Z, M) = \left(\min_{i:\, |t_i| \le T} Z(t_i), \min\{|t_i|: Z(t_i) \le Z(t_j), |t_j| \le T\}\right), \tag{3.14}$$

where the t_i's are the support of M. This map is continuous except possibly at (Z, M) for which $M(\{-T, T\}) > 0$, and $N(\{-T, T\}) = 0$ almost surely. Therefore, by the continuous mapping lemma, since $(Z_n, N_n) \to_{\mathscr{D}} (R, N)$, we conclude that

$$\Psi(Z_n, N_n) = \left(\sqrt{c_n}\Delta_n^T, c_n\Gamma_n^T\right) \to_{\mathscr{D}} \Psi(R, N) = (\Delta^T, \Gamma^T) \tag{3.15}$$

as $n \to \infty$, where

$$\Delta^T = \inf_{|\tau_i| \le T} R(\tau_i), \qquad \Gamma^T = \inf\{|\tau_i|: R(\tau_i) = \Delta^T\},$$

and R and $\{\tau_i\}$ are defined before Theorem 3.1. For any fixed $a < \infty$, Lemma 4 of Asmussen, Glynn and Pitman (1995) gives that

$$\lim_{T\to\infty} \limsup_{n\to\infty} P\left(\sqrt{c_n}\,\tilde{\Delta}_n^T \le a\right) = 0. \tag{3.16}$$

Therefore, for any $a, b > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} \left| P\left(\sqrt{c_n}\,\Delta_n \le a,\ c_n\Gamma_n \le b\right) - P(\Delta \le a,\ \Gamma \le b)\right| \\
&\quad \le \limsup_{n\to\infty} \left| P\left(\sqrt{c_n}\,\Delta_n \le a,\ c_n\Gamma_n \le b\right) - P\left(\sqrt{c_n}\,\Delta_n^T \le a,\ c_n\Gamma_n^T \le b\right)\right| \\
&\qquad + \limsup_{n\to\infty} \left| P\left(\sqrt{c_n}\,\Delta_n^T \le a,\ c_n\Gamma_n^T \le b\right) - P(\Delta^T \le a,\ \Gamma^T \le b)\right| \\
&\qquad + \left|P(\Delta^T \le a,\ \Gamma^T \le b) - P(\Delta \le a,\ \Gamma \le b)\right| \\
&\quad \triangleq L_1(T) + L_2(T) + L_3(T).
\end{aligned}$$

By (3.16), $L_1(T) \to 0$ as $T \uparrow \infty$, since

$$L_1(T) \le \limsup_{n\to\infty} P\left(\sqrt{c_n}\,\tilde{\Delta}_n^T \le a\right).$$

By (3.15), $L_2(T) = 0$. Finally, as $T \uparrow \infty$, $(\Delta^T, \Gamma^T) \to_{\mathscr{D}} (\Delta, \Gamma)$, and so $L_3(T) \to 0$ as $T \uparrow \infty$, and the theorem is proved. □

As noted at the end of Section 2, to analyze the adaptive algorithms we are interested in the convergence rate of $\hat{t}_n^* - t^*$ to 0. Theorem 3.2 establishes (taking $c_n = n$) that $n(U_n^* - t^*) \to_{\mathscr{D}} \Gamma$ as $n \to \infty$. Therefore, if $M_n \sim$ Binomial$(n, 1/2)$ independent of the $\{U_i\}$ and t^*, then $M_n(U_{M_n}^* - t^*) \to_{\mathscr{D}} \Gamma$. Since n/M_n converges to 2 in probability, we have that $n(\hat{t}_n^* - t^*)$ converges in distribution as $n \to \infty$, and if $\delta > 0$, then $n^{1-\delta}(\hat{t}_n^* - t^*) \to_{\mathscr{D}} 0$.

4. Limiting behavior of algorithms. In this section we will analyze the limiting distribution of the normalized error under the adaptive algorithm described in Section 2. Throughout this section, $\delta \in (0, 1)$ will be fixed. We will use the sequence $a_n = [2(2-\delta)]^{-1} n^{-(1-\delta)}$ in the definition of the algorithm. Since $n(\hat{t}_n^* - t^*)$ converges in distribution, this choice of a_n ensures that the distance between $\hat{t}_n^*$ (the center of the local search) and t^* will be asymptotically negligible compared to the scope of the local search.

Letting $\Theta_n = a_n(U_n - 1/2) = [2(2-\delta)]^{-1} n^{-(1-\delta)}(U_n - 1/2)$, the algorithm described in Section 2 takes the following specific form:

Set $t_1 = \hat{t}_1^* = t_1^* \leftarrow U_1$, $\hat{B}_1^* = B_1^* \leftarrow B(t_1)$;
For $k = 1, 2, \ldots, n-1$,
  If $\beta_{k+1} = 0$, {perform global search}
    Set $t_{k+1} \leftarrow U_{k+1}$;
    If $B(t_{k+1}) < B_k^*$, then set $t_{k+1}^* \leftarrow t_{k+1}$, $B_{k+1}^* \leftarrow B(t_{k+1})$;
    If $B(t_{k+1}) < \hat{B}_k^*$, then set $\hat{t}_{k+1}^* \leftarrow t_{k+1}$, $\hat{B}_{k+1}^* \leftarrow B(t_{k+1})$;

Else if $\beta_{k+1} = 1$, {perform local search}
Set $t_{k+1} \leftarrow \hat{t}_k^* + \Theta_{k+1}$;
If $B(t_{k+1}) < B_k^*$, then set $t_{k+1}^* \leftarrow t_{k+1}$, $B_{k+1}^* \leftarrow B(t_{k+1})$.

For the rest of this section, $\{t_i\}$ will denote the observations produced by this algorithm with fixed δ, and Δ_n and Γ_n will denote the corresponding error variables.

We will make use of the following result from Kallenberg (1976), specialized to our setting.

THEOREM 4.1. *Let N_n $(n \geq 1)$ and N be point processes on $\mathbb{R}^k$. Let $\mathscr{A}$ denote the class of rectangles $(s_1, t_1] \times (s_2, t_2] \times \cdots \times (s_k, t_k] \subset \mathbb{R}^k$. If for any j and $A, A_1, A_2, \ldots, A_j \in \mathscr{A}$,*

$$EN_n(A) \to EN(A)$$

and

$$P\left(N_n\left(\bigcup_{i=1}^{j} A_i\right) = 0\right) \to P\left(N\left(\bigcup_{i=1}^{j} A_i\right) = 0\right)$$

as $n \to \infty$, then $N_n \to_{\mathscr{D}} N$.

We will use Theorem 3.2 with $c_n = n^{2-\delta}$ to establish the limiting error distribution.

THEOREM 4.2. *For $n \geq 1$ and Borel sets A, let*

$$N_n(A) = \sum_{k=1}^{n} I_{\{n^{2-\delta}(t_k - t^*) \in A\}}.$$

Then $N_n \to_{\mathscr{D}} N$, where N is a Poisson point process on the line with unit intensity, independent of B.

Here is the main idea of the proof. The points chosen uniformly over the entire unit interval are in the limit negligible under the $n^{2-\delta}$ scaling, so only the local search sites play a role. If the local search intervals were centered exactly at t^*, then it would be easy to show that the point processes converge to the Poisson limit; however, the local search intervals are offset by the error random variables $\varepsilon_n \triangleq \hat{t}_n^* - t^*$. The problem then is to show that the ε_n are small enough not to figure in the limit.

PROOF. We are to prove that

$$(N_n, B) \to_{\mathscr{D}} (N, B), \tag{4.1}$$

with N independent of B. Let

$$C = \{B(s_1) \in C_1, B(s_2) \in C_2, \ldots, B(s_m) \in C_m\}$$

for $m \geq 1$, $0 \leq s_1 < s_2 < \cdots < s_m \leq 1$ and Borel sets C_i. Define a point process M on $\mathbb{R}^k$ by

$$M(C_1, C_2, \ldots, C_k) = I_{\{B(t_1) \in C_1, \ldots, B(t_k) \in C_k\}}. \tag{4.2}$$

We will show that $(N_n, M) \to_{\mathscr{D}} (N, M)$, where N and M are independent, from which (4.1) will follow. We will appeal to Theorem 4.1 after showing that

$$E(N_n((s,t])M(C_1, \ldots, C_k)) \to E(N((s,t]))EM(C_1, \ldots, C_k) \tag{4.3}$$

and that

$$\begin{aligned} &P(N_n(A) = 0, M(C_1, \ldots, C_k) = 0) \\ &\quad \to P(N(A) = 0)P(M(C_1, \ldots, C_k) = 0), \end{aligned} \tag{4.4}$$

where A is a union of disjoint intervals.

We begin by establishing (4.3). Setting $\lambda = 4(2 - \delta)$ (so that $\lambda k^{1-\delta}\Theta_k \sim U(-1, 1)$),

$$\begin{aligned} E(N_n(s,t)I_C) &= \sum_{k=1}^{n} P\big(n^{2-\delta}(t_k - t^*) \in (s,t], C\big) \\ &= \sum_{k=1}^{n} P\big(\big(\beta_k = 0,\ n^{2-\delta}(U_k - t^*) \in (s,t], C\big) \\ &\qquad \cup \big(\beta_k = 1,\ n^{2-\delta}\big(\hat{t}_k^* + \Theta_k - t^*\big) \in (s,t], C\big)\big) \\ &= \sum_{k=1}^{n} P\big(\beta_k = 0,\ n^{2-\delta}(U_k - t^*) \in (s,t], C\big) \\ &\quad + \sum_{k=1}^{n} P\big(\beta_k = 1,\ n^{2-\delta}\big(\hat{t}_k^* + \Theta_k - t^*\big) \in (s,t], C\big). \end{aligned}$$

The first sum in the last expression is negligible, since

$$\begin{aligned} &\sum_{k=1}^{n} P\big(\beta_k = 0,\ n^{2-\delta}(U_k - t^*) \in (s,t], C\big) \\ &\qquad \leq P(C) \sum_{k=1}^{n} \tfrac{1}{2}|t - s|n^{-2+\delta} \to 0, \end{aligned}$$

as $n \to \infty$, and so

$$\begin{aligned} &E(N_n(s,t)I_C) \\ &\quad = o(1) + \sum_{k=1}^{n} \frac{1}{2} P\left((k/n)^{1-\delta}\lambda\frac{s}{n} < \lambda k^{1-\delta}\varepsilon_k + \lambda k^{1-\delta}\Theta_k \leq (k/n)^{1-\delta}\lambda\frac{t}{n}, C\right) \end{aligned}$$

as $n \to \infty$. We will show that the last sum has the same limit, as $n \to \infty$, as

$$\sum_{k=1}^{n} \frac{1}{2} P\left((k/n)^{1-\delta}\lambda\frac{s}{n} < \lambda k^{1-\delta}\Theta_k \leq (k/n)^{1-\delta}\lambda\frac{t}{n}, C\right).$$

To simplify notation, let

$$b_{k,n} = (k/n)^{1-\delta}\frac{\lambda}{n}, \qquad W_k = \lambda k^{1-\delta}\varepsilon_k, \qquad V_k = \lambda k^{1-\delta}\Theta_k.$$

Note that $V_k \sim U(-1,1)$, and recall from the discussion following Theorem 3.2 that $W_k \to_{\mathscr{D}} 0$. We need to establish that

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left[P(sb_{k,n} < W_k + V_k \le tb_{k,n}, C) - P(sb_{k,n} < V_k \le tb_{k,n}, C)\right] = 0.$$

To see this, observe that

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left[P(sb_{k,n} < W_k + V_k \le tb_{k,n}, C) - P(sb_{k,n} < V_k \le tb_{k,n}, C)\right]$$

$$= \lim_{n\to\infty} \sum_{k=1}^{n} \int_{x=-\infty}^{\infty} \left[P(sb_{k,n} - x < V_k \le tb_{k,n} - x, C | W_k = x)\right.$$

$$\left. - P(sb_{k,n} < V_k \le tb_{k,n}, C | W_k = x)\right] P(W_k \in dx).$$

The random variables W_k and V_k are independent, and

$$P(sb_{k,n} - x < V_k \le tb_{k,n} - x, C | W_k = x) = P(sb_{k,n} < V_k \le tb_{k,n}, C | W_k = x)$$

unless $x < -1 + tb_{k,n}$ or $x > 1 + sb_{k,n}$, since $V_k \sim U(-1,1)$. Also

$$|P(sb_{k,n} - x < V_k \le tb_{k,n} - x, C | W_k = x)$$

$$- P(sb_{k,n} < V_k \le tb_{k,n}, C | W_k = x)| \le (t-s) b_{k,n}$$

for all x. If $n > 2\lambda \max\{|s|, |t|\}$, then $|tb_{k,n}| < 1/2$ and $|sb_{k,n}| < 1/2$ for all $k \le n$, and therefore,

$$\left| \sum_{k=1}^{n} \left[P(sb_{k,n} < W_k + V_k \le tb_{k,n}, C) - P(sb_{k,n} < V_k \le tb_{k,n}, C)\right] \right|$$

$$\le (t-s) \sum_{k=1}^{n} b_{k,n} P(C, |W_k| > 1/2)$$

$$\le (t-s) \frac{\lambda}{n} \sum_{k=1}^{n} \left(\frac{k}{n}\right)^{1-\delta} P(|W_k| > 1/2)$$

$$\le (t-s) \frac{\lambda}{n} \sum_{k=1}^{n} P(|W_k| > 1/2),$$

which converges to 0 because $W_k \to_{\mathscr{D}} 0$. Therefore, since $\lambda k^{1-\delta}\Theta_k \sim U(-1,1)$ (independent of B),

$$\lim_{n\to\infty} E(N_n(s,t) I_C)$$

$$= \lim_{n\to\infty} \frac{1}{2} \sum_{k=1}^{n} P\left((k/n)^{1-\delta}\lambda \frac{s}{n} < \lambda k^{1-\delta}\Theta_k \le (k/n)^{1-\delta}\lambda \frac{t}{n}, C\right)$$

$$= \lim_{n\to\infty} \frac{1}{2} \sum_{k=1}^{n} \frac{1}{2} 4(2-\delta)(t-s) \frac{1}{n} \left(\frac{k}{n}\right)^{1-\delta} P(C)$$

$$= (2-\delta)(t-s) \int_{x=0}^{1} x^{1-\delta}\, dx\, P(C)$$

$$= (t-s) P(C).$$

We next establish (4.4) by showing that $P(N_n(A) = 0, C) \to P(N(A) = 0)P(C)$, where $A = \bigcup_{i=1}^{k} (s_i, t_i\,]$ is a finite union of disjoint intervals, and we write $|A| = \sum_{i=1}^{k}(t_i - s_i)$ and $Ab_{k,n} = \{xb_{k,n}\colon x \in A\}$. Decomposing the events according to $\beta_k = 0$ or 1,

$$
\begin{aligned}
&P(N_n(A) = 0, C)\\
&\quad = P\Bigg(\bigcap_{k=1}^{n}\Big(\big[\,\beta_k = 0,\ n^{2-\delta}(U_k - t^*) \notin A\big]\\
&\qquad\qquad \cup \big[\,\beta_k = 1,\ n^{2-\delta}(\epsilon_k + \Theta_k) \notin A\big]\Big) \cap C\Bigg)\\
&\quad = P\Bigg(\bigcap_{k=1}^{n}\Big(\big[\,\beta_k = 0,\ n^{2-\delta}(U_k - t^*) \notin A\big]\\
&\qquad\qquad \cup \big[\,\beta_k = 1,\ W_k + V_k \notin Ab_{k,n}\big]\Big) \cap C\Bigg)\\
&\quad = P\Bigg(\bigcap_{k=1}^{n}\Big(\big[(\,\beta_k = 0) \cup (\,\beta_k = 1,\ W_k + V_k \notin Ab_{k,n})\big]\\
&\qquad\qquad \cap \big[\big(n^{2-\delta}(U_k - t^*) \notin A\big) \cup (\,\beta_k = 1,\ W_k + V_k \notin Ab_{k,n})\big]\Big) \cap C\Bigg).
\end{aligned}
$$

Now,

$$
\begin{aligned}
&P\Bigg(\bigcap_{k=1}^{n}\big[\big(n^{2-\delta}(U_k - t^*) \notin A\big) \cup (\,\beta_k = 1,\ W_k + V_k \notin Ab_{k,n})\big]\Bigg)\\
&\qquad \geq P\Bigg(\bigcap_{k=1}^{n}\big(n^{2-\delta}(U_k - t^*) \notin A\big)\Bigg)\\
&\qquad = \prod_{k=1}^{n} P\big(\big(n^{2-\delta}(U_k - t^*) \notin A\big)\big)\\
&\qquad \geq \left(1 - \frac{|A|}{n^{2-\delta}}\right)^{n} \to 1,
\end{aligned}
$$

and so

$$
\begin{aligned}
&\lim_{n\to\infty} P(N_n(A) = 0,\ C)\\
&\qquad = \lim_{n\to\infty} P\Bigg(\bigcap_{k=1}^{n}\big[(\,\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})\big] \cap C\Bigg).
\end{aligned}
\tag{4.5}
$$

We will next show that the limit remains unchanged if we delete the W_k term from the expression in (4.5); that is,

$$
\begin{aligned}
&\lim_{n\to\infty} P(N_n(A) = 0,\ C)\\
&\qquad = \lim_{n\to\infty} P\Bigg(\bigcap_{k=1}^{n}\big[(\,\beta_k = 0) \cup (V_k \notin Ab_{k,n})\big] \cap C\Bigg).
\end{aligned}
\tag{4.6}
$$

First note that for n sufficiently large,

$$P((\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})|W_k) = P((\beta_k = 0) \cup (V_k \notin Ab_{k,n})|W_k)$$

for $|W_k| \leq 1/2$, say, and therefore for n sufficiently large,

$$\begin{aligned}
&P\left(C \cap \bigcap_{k=1}^{n} [(\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})]\right) \\
&\quad = EP\left(C \cap \bigcap_{k=1}^{n} [(\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})]|W_1, W_2, \ldots, W_n\right) \\
&\quad = E\Bigg[P(C|W_1, W_2, \ldots, W_n) \\
&\qquad \times P\left(\bigcap_{k=1}^{n} [(\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})]|W_1, W_2, \ldots, W_n\right)\Bigg] \\
&\quad = E\Bigg[P(C|W_1, W_2, \ldots, W_n) \\
&\qquad \times \prod_{k=1}^{n} P((\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})|W_1, W_2, \ldots, W_n)\Bigg] \\
&\quad = E\left[P(C|W_1, W_2, \ldots, W_n) \prod_{k=1}^{n} P((\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})|W_k)\right] \\
&\quad = E\Bigg[P(C|W_1, W_2, \ldots, W_n) \prod_{\substack{k=1 \\ |W_k| \leq 1/2}}^{n} P((\beta_k = 0) \cup (V_k \notin Ab_{k,n})|W_k) \\
&\qquad \times \prod_{\substack{k=1 \\ |W_k| > 1/2}}^{n} P((\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})|W_k)\Bigg] \\
&\quad = E\Bigg[P(C|W_1, W_2, \ldots, W_n) \prod_{\substack{k=1 \\ |W_k| \leq 1/2}}^{n} P((\beta_k = 0) \cup (V_k \notin Ab_{k,n})) \\
&\qquad \times \prod_{\substack{k=1 \\ |W_k| > 1/2}}^{n} P((\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})|W_k)\Bigg] \\
&\quad = E\Bigg[P(C|W_1, W_2, \ldots, W_n) \prod_{k=1}^{n} P((\beta_k = 0) \cup (V_k \notin Ab_{k,n})) \\
&\qquad \times \prod_{\substack{k=1 \\ |W_k| > 1/2}}^{n} \frac{P((\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})|W_k)}{P((\beta_k = 0) \cup (V_k \notin Ab_{k,n}))}\Bigg].
\end{aligned}$$

Similarly, we can write

$$P\left(C \cap \bigcap_{k=1}^{n} [(\beta_k = 0) \cup (V_k \notin Ab_{k,n})]\right)$$
$$= E\left[P(C|W_1, W_2, \ldots, W_n) \prod_{k=1}^{n} P((\beta_k = 0) \cup (V_k \notin Ab_{k,n}))\right].$$

Therefore,

$$0 \leq P\left(C \cap \bigcap_{k=1}^{n} [(\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})]\right)$$
$$- P\left(C \cap \bigcap_{k=1}^{n} [(\beta_k = 0) \cup (V_k \notin Ab_{k,n})]\right)$$
$$= E\left[P(C|W_1, W_2, \ldots, W_n) \prod_{k=1}^{n} P((\beta_k = 0) \cup (V_k \notin Ab_{k,n})) \right.$$
$$\left. \times \prod_{\substack{k=1 \\ |W_k| > 1/2}}^{n} \frac{P((\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})|W_k)}{P((\beta_k = 0) \cup (V_k \notin Ab_{k,n}))}\right]$$
$$- E\left[P(C|W_1, W_2, \ldots, W_n) \prod_{k=1}^{n} P((\beta_k = 0) \cup (V_k \notin Ab_{k,n}))\right]$$
$$= E\left[P(C|W_1, W_2, \ldots, W_n) \prod_{k=1}^{n} P((\beta_k = 0) \cup (V_k \notin Ab_{k,n})) \right.$$
$$\left. \times \left(\prod_{\substack{k=1 \\ |W_k| > 1/2}}^{n} \frac{P((\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})|W_k)}{P((\beta_k = 0) \cup (V_k \notin Ab_{k,n}))} - 1 \right)\right]$$
$$\leq E\left(\prod_{\substack{k=1 \\ |W_k| > 1/2}}^{n} \frac{P((\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})|W_k)}{P((\beta_k = 0) \cup (V_k \notin Ab_{k,n}))} - 1 \right).$$

In order to establish the validity of (4.6), it therefore suffices to show that

$$\lim_{n \to \infty} E\left(\prod_{\substack{k=1 \\ |W_k| > 1/2}}^{n} \frac{P((\beta_k = 0) \cup (W_k + V_k \notin Ab_{k,n})|W_k)}{P((\beta_k = 0) \cup (V_k \notin Ab_{k,n}))} \right) = 1. \tag{4.7}$$

Clearly the left-hand side is at least 1, and

$$E\left(\prod_{\substack{k=1\\|W_k|>1/2}}^{n}\frac{P((\beta_k=0)\cup(W_k+V_k\notin Ab_{k,n})|W_k)}{P((\beta_k=0)\cup(V_k\notin Ab_{k,n}))}\right)$$

$$=E\prod_{k=1}^{n}\left(I_{(|W_k|\le 1/2)}+I_{(|W_k|>1/2)}\frac{P((\beta_k=0)\cup(W_k+V_k\notin Ab_{k,n})|W_k)}{P((\beta_k=0)\cup(V_k\notin Ab_{k,n}))}\right)$$

$$\le E\prod_{k=1}^{n}\left(I_{(|W_k|\le 1/2)}+I_{(|W_k|>1/2)}\frac{1}{P((\beta_k=0)\cup(V_k\notin Ab_{k,n}))}\right)$$

$$\le E\prod_{k=1}^{n}\left(1+I(|W_k|>1/2)\left(\frac{1}{P((\beta_k=0)\cup(V_k\notin Ab_{k,n}))^2}-1\right)\right),$$

and the last product converges to 1 as $n\to\infty$ if

$$\lim_{n\to\infty}\sum_{k=1}^{n}P(|W_k|>1/2)\left(\frac{1}{P((\beta_k=0)\cup(V_k\notin Ab_{k,n}))^2}-1\right)=0. \tag{4.8}$$

For n large enough that $Ab_{k,n}\in[-1,1]$,

$$P((\beta_k=0)\cup(V_k\notin Ab_{k,n}))=1-2(2-\delta)\left(\frac{k}{n}\right)^{1-\delta}\frac{|A|}{n}, \tag{4.9}$$

and so

$$\sum_{k=1}^{n}P(|W_k|>1/2)\left(\frac{1}{P((\beta_k=0)\cup(V_k\notin Ab_{k,n}))^2}-1\right)$$

$$=\sum_{k=1}^{n}P(|W_k|>1/2)$$

$$\times\left(\frac{(4(2-\delta)(k/n)^{1-\delta}|A|/n)-\left[2(2-\delta)(k/n)^{1-\delta}|A|/n\right]^2}{\left[1-2(2-\delta)(k/n)^{1-\delta}|A|/n\right]^2}\right)$$

$$\le\sum_{k=1}^{n}P(|W_k|>1/2)\left(\frac{4(2-\delta)(k/n)^{1-\delta}|A|/n}{\left[1-2(2-\delta)(k/n)^{1-\delta}|A|/n\right]^2}\right).$$

If n is large enough that $2(2-\delta)(|A|/n)\le 1/2$, then the denominator in the

last expression is at least $(1/2)^2$, and so the last sum is at most

$$16(2-\delta)|A|\frac{1}{n}\sum_{k=1}^{n}P(|W_k|>1/2)\left(\frac{k}{n}\right)^{1-\delta}$$
$$\leq 16(2-\delta)|A|\frac{1}{n}\sum_{k=1}^{n}P(|W_k|>1/2),$$

which converges to 0 as $n\to\infty$ since $W_k\to_{\mathscr{D}}0$ implies that $P(|W_k|>1/2)\to 0$.

Therefore, from (4.6),

$$\begin{aligned}\lim_{n\to\infty}P(C,N_n(A)=0) &= \lim_{n\to\infty}P\left(C,\bigcap_{k=1}^{n}[\beta_k=0]\cup[\beta_k=1,V_k\notin Ab_{k,n}]\right)\\ &= \lim_{n\to\infty}\prod_{k=1}^{n}P(C,[\beta_k=0]\cup[\beta_k=1,V_k\notin Ab_{k,n}])\\ &= \lim_{n\to\infty}\prod_{k=1}^{n}\left(\frac{1}{2}+\frac{1}{2}\left(\frac{1}{2}\right)\left[2-\lambda(k/n)^{1-\delta}\frac{|A|}{n}\right]\right)P(C)\\ &= \lim_{n\to\infty}\prod_{k=1}^{n}\left(1-(2-\delta)\left(\frac{k}{n}\right)^{1-\delta}\frac{|A|}{n}\right)P(C)\\ &= \exp(-|A|)P(C).\end{aligned}$$

We have established (4.3) and (4.4), and so by Theorem 4.1, $(N_n,M)\to_{\mathscr{D}}(N,M)$, where N and M are independent. Therefore,

$$\begin{aligned}&P\big(N_n(A_1)=n_1,\ldots,N_n(A_j)=n_j,\ M(C_1,\ldots,C_k)=1\big)\\ &\quad=P\big(N_n(A_1)=n_1,\ldots,N_n(A_j)=n_j,\ B(t_1)\in C_1,\ldots,B(t_k)\in C_k\big)\\ &\quad\to P\big(N(A_1)=n_1,\ldots,N(A_j)=n_j\big)P\big(B(t_1)\in C_1,\ldots,B(t_k)\in C_k\big).\end{aligned}$$

This shows that $(N_n,B)\to_{\mathscr{D}}(N,B)$, with N,B independent, which completes the proof. □

Combining Proposition 4.2 and Theorem 3.2 gives our main result, establishing the limiting behavior of the class of adaptive algorithms.

THEOREM 4.3. *Under the adaptive algorithm for any* $0<\delta<1$,

$$\left(n^{1-\delta/2}\Delta_n,n^{2-\delta}\Gamma_n\right)\to_{\mathscr{D}}(\Delta,\Gamma)$$

as $n\to\infty$.

Note in particular the marginal limiting distribution of the function approximation error:

$$P\left(n^{1-\delta/2}\Delta_n\leq y\right)\to\tanh^2\left(y\sqrt{2}\right). \tag{4.10}$$

TABLE 1
Sample medians

$\delta \backslash n$	1000	2000	3000	4000
0.8	0.5396	0.5345	0.5349	0.5276
0.6	0.7020	0.6351	0.6730	0.6362
0.4	0.8972	0.8822	0.8126	0.7952
0.2	1.7200	1.6409	1.6081	1.5185

5. Numerical experiments. Numerical experiments were performed to determine how well the limit distribution given by Theorem 4.3 approximates the error Δ_n for moderate values of n. Experiments were performed for several values of δ and n; the results are reported in Tables 1 and 2. For each choice of δ and n, 1000 independent replications of the algorithm were simulated, and the empirical distribution function F_{1000} was calculated for the observed values of $n^{1-\delta/2}\Delta_n$. Two different quantities were calculated for the comparison; the Kolmogorov–Smirnov statistic

$$D = \sup_{y \geq 0} \left| F_{1000}(y) - \tanh^2\left(y\sqrt{2}\right) \right|,$$

and the average of the smallest and largest medians of F_{1000}. The median of the distribution of the limit random variable Δ is

$$2^{-1/2} \tanh(2^{-1/2}) \approx 0.623225.$$

The data are reported in Tables 1 and 2. Note that the approximation is much worse for small values of δ (corresponding to quicker localization of the search). To see why this should be the case, consider the ratio of the local search horizon (half-width of the local search interval size, which is $(2-\delta)^{-1}n^{-(1-\delta)}$) to the error $(\hat{t}_n^* - t^*)$. This ratio is asymptotically proportional to $(2-\delta)^{-1}n^{\delta}$. To achieve the same ratio achieved with $\delta = 0.8$ and n iterations, with $\delta = 0.2$, about $(3/2)^5 n^4$ iterations would be required.

6. Final remarks. It is not clear if Theorem 4.3 would remain valid if we set $\hat{t}_n^*$ to the location of the minimum of the first n observations in the definition of the algorithm. Such a change would result in an algorithm that would be difficult to analyze, since the distribution of the best observed

TABLE 2
Kolmogorov–Smirnov statistics

$\delta \backslash n$	1000	2000	3000	4000
0.8	0.1027	0.1055	0.1017	0.1032
0.6	0.1045	0.0445	0.0800	0.0586
0.4	0.2458	0.2453	0.1823	0.1887
0.2	0.5182	0.4825	0.4841	0.4567

location is quite complicated. In any case, the asymptotic performance of the algorithms described in this paper could not improve with such a change, since by their construction they would have the same asymptotic performance even if the local searches were centered exactly at the minimizer.

While the algorithms we have described are adaptive, they are "barely" adaptive in that only the center of the local searches are updated according to the past observed values. The width of the local searches, determined by the sequence $\{a_n\}$, is deterministic.

It would be interesting to determine the best possible convergence rate of a Markovian algorithm, and also the best possible convergence rate of *any* algorithm. Because of the great difference in computational and storage requirements between Markov and general adaptive algorithms, it is of interest to have an upper bound on how much can be gained by maintaining all past information.

REFERENCES

AL-MHARMAH, H. and CALVIN, J. (1996). Optimal random nonadaptive algorithm for global optimization of Brownian motion. *J. Global Optim.* **8** 81–90.

ARCHETTI, F. and BETRÒ, B. (1979). A probabilistic algorithm for global optimization. *Calcolo* **16** 335–343.

ASMUSSEN, S., GLYNN, P. and PITMAN, J. (1995). Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. Appl. Probab.* **5** 875–896.

CALVIN, J. (1995). Average performance of nonadaptive algorithms for global optimization of Brownian motion. *J. Math. Anal. Appl.* **191** 608–617.

CALVIN, J. and GLYNN, P. W. (1997). Average behavior of random search for the maximum. *J. Appl. Probab.* To appear.

CALVIN, J. and GLYNN, P. (1996). Complexity of nonadaptive optimization algorithms for a class of diffusions. *Comm. Statist. Stochastic Models* **12** 343–365.

ITÔ, K. and MCKEAN, H. (1965). *Diffusion Processes and Their Sample Paths.* Springer, Berlin.

KALLENBERG, O. (1976). *Random Measures.* Akademie, Berlin.

KUSHNER, H. (1964). A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering* **86** 97–106.

REVUZ, D. and YOR, M. (1991). *Continuous Martingales and Brownian Motion.* Springer, Berlin.

RITTER, K. (1990). Approximation and optimization on the Wiener space. *J. Complexity* **6** 337–364.

ZHIGLJAVSKY, A. (1991). *Theory of Global Random Search.* Kluwer, Dordrecht.

ŽILINSKAS, A. (1976). The method of one-dimensional multiextremal minimization. *Tekhn. Kibernet.* **76.4** 71–74.

DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE
NEW JERSEY INSTITUTE OF TECHNOLOGY
UNIVERSITY HEIGHTS
NEWARK, NEW JERSEY 07102-1982
E-MAIL: calvin@cis.njit.edu