JOURNAL OF COMPLEXITY **17**, 306–344 (2001)
doi:10.1006/jcom.2001.0574, available online at http://www.idealibrary.com on IDEAL®

A One-Dimensional Optimization Algorithm and Its Convergence Rate under the Wiener Measure[1]

James M. Calvin

Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey 07102
E-mail: calvin@cis.njit.edu

Received November 11, 1999

In this paper we describe an adaptive algorithm for approximating the global minimum of a continuous function on the unit interval, motivated by viewing the function as a sample path of a Wiener process. It operates by choosing the next observation point to maximize the probability that the objective function has a value at that point lower than an adaptively chosen threshold. The error converges to zero for any continuous function. Under the Wiener measure, the error converges to zero at rate $e^{-n\delta_n}$, where $\{\delta_n\}$ (a parameter of the algorithm) is a positive sequence converging to zero at an arbitrarily slow rate. © 2001 Academic Press

Key Words: Wiener measure; global optimization.

1. INTRODUCTION

A fundamental optimization problem is to approximate the minimum of a one-dimensional continuous function by evaluating the function at sequentially selected points. This problem has received considerable attention for several special classes of objective function. For example, if the function is known to be unimodal, then the Fibonacci search method converges and enjoys a certain optimality property (see [6]). Because of the power of the unimodality assumption, each function evaluation allows an increasing subset of the interval to be ignored.

In this paper we describe an optimization algorithm that approximates the minimum when the objective function is only assumed to be continuous. In contrast to the unimodal case, if the method is to converge for all continuous functions then no subinterval of the domain can ever be

[1] This work was supported in part by NSF Grant DMI-9696243.

0885-064X/01 $35.00

ruled out and the sequence of evaluation points must be dense. The algorithm is motivated by viewing the objective function as a sample path of a Wiener process. Each new point is chosen to maximize the probability that the new function value will fall below the minimum seen so far, minus a positive increment that depends on the past observations.

To be more precise, let $\{X(t): 0 \leqslant t \leqslant 1\}$ be a Wiener process. We are interested in approximating the global minimum of X by means of sequential observation at

$$t_1, t_2(X(t_1)), t_3(X(t_1), X(t_2)), \ldots.$$

Let $\{z_n\}$ be a sequence of positive numbers, where z_n may depend on the past history $\{t_i, X(t_i); i<n\}$, and let $M_n = \min_{1 \leqslant i \leqslant n} X(t_i)$ be the minimum observed value by time n. We study the optimization method based on the following procedure: Given $t_1, t_2, \ldots, t_n$ and $X(t_1)=x_1, X(t_2)=x_2, \ldots, X(t_n)=x_n$, choose the next point $t_{n+1} \in [0, 1]$ to maximize

$$P(X(t_{n+1}) < M_n - z_n \mid X(t_1)=x_1, \ldots, X(t_n)=x_n). \tag{1}$$

In order for this algorithm to choose distinct points it is necessary that $z_n > 0$, since if $z_n = 0$ the conditional probability in (1) approaches its upper bound of 1/2 as t_{n+1} approaches the point t_i where $X(t_i)$ is smallest.

This method is a variant of an algorithm described by Kushner [7], where the increments are a positive sequence independent of the function. The algorithm proposed by Kushner was called the P-algorithm by Žilinskas [12], who established a formal justification for the approach. The convergence properties of this algorithm for the special case where the increment is a positive constant (i.e., $z_1 = z_2 = \cdots = c > 0$) were studied in [3] both when applied to a fixed continuous function, and also when applied to a random path of a Brownian motion process. Let M denote the global minimum of X and let $\Delta_n = M_n - M$ denote the error after n observations. In [3] it is shown that there exists an increasing sequence of stopping times $\{n_k\}$ such that

$$P\left(\sqrt{n_k}\, \Delta_{n_k} \left(\int_{t=0}^{1} \left(1 + \frac{X(t)-M}{c}\right)^{-2} dt\right)^{-1/2} \leqslant y\right) \to F(y), \tag{2}$$

where F is the limiting distribution function of the normalized error with equi-spaced observations; that is,

$$P(\sqrt{n} \min_{0 \leqslant i \leqslant n} (X(i/n) - M) \leqslant y) \to F(y). \tag{3}$$

Notice that the normalizing sequence multiplying Δ_{nk} in (2) includes a random component containing an integral of a function of X. This random component can be made arbitrarily large by choosing c small. Therefore by decreasing c the P-algorithm can attain any desired asymptotic speedup factor over the equi-spaced grid, but in terms of the number of function evaluations the convergence rate remains $n^{-1/2}$.

Ritter [9] showed that the best nonadaptive algorithms have convergence rate $n^{-1/2}$; thus the asymptotic convergence rate in (2) is the same as for the best nonadaptive algorithms. Nonadaptive algorithms do not need to store the past history of observations, and there are simple algorithms that achieve the optimal nonadaptive $n^{-1/2}$ convergence rate (for example, bisect the largest subinterval). The P-algorithm, in contrast, stores the entire history of observations and has a computational cost that grows quadratically with the number of iterations.

A randomized algorithm was constructed in [2] with the property that for any $0<\delta<1$, a version can be constructed so that under the Wiener measure,

$$P(n^{1-\delta}\,\Delta_n \leqslant y) \to \tanh^2(\sqrt{2}\,y), \qquad y>0. \tag{4}$$

This algorithm maintains a memory of two past observations, and the computational cost grows linearly with the number of iterations. In light of this result, the convergence rate implied by (2) is quite slow. (Throughout this paper we focus on asymptotic convergence rates, and so these comments of course do not imply anything about the efficiency of the P-algorithm with constant c in the non-asymptotic sense.)

In order to improve on the $n^{-1/2}$ convergence rate, we explore the possibility of replacing the constant c with a positive sequence $\{z_n\}$ converging to 0. It turns out that it is possible to improve (dramatically) on the $n^{-1/2}$ convergence rate obtained with a constant c, but the choice of decreasing sequence involves a delicate balance. While the convergence rate improves with a smaller sequence up to a point, if the sequence converges to zero too quickly then the algorithm may not converge to the global minimum. The purpose of this paper is to construct an appropriate sequence and to prove a limit theorem for the resulting algorithm. Instead of a fixed deterministic sequence, we allow the sequence to depend on the previous observations.

An informal summary of the algorithm and main result follows. Let τ_n be the smallest gap between observation points after n observations, and let $\{\gamma_n\}$ be an increasing sequence of positive numbers such that $\gamma_n \to \infty$,

$$\frac{\gamma_n^2 \log(\gamma_n)}{n} \to 0, \qquad \text{and} \qquad \frac{\gamma_n}{\gamma_{n+1}} \to 1. \tag{5}$$

```
1.  Optimize(X, {γ_k}, n)
2.      M ← 0
3.      z ← γ_0
4.      for k ← 1 to n do
5.          t_k ← argmax P(X(t_k) < M − z_{k−1} | X(t_i) = x_i; i < k)
6.          x_k ← X(t_k)
7.          M ← min{x_k, M}
8.          τ ← min_{i,j≤k} |t_i − t_j|
9.          z ← min{γ_k √τ, z}
10.     return M
```

FIG. 1. Algorithm to minimize X with n observations and sequence $\{\gamma_n\}$.

The sequence $\{z_n\}$ is defined by $z_n = \min_{i \leqslant n} \{\sqrt{\tau_i}\, \gamma_i\}$. The algorithm is depicted in Fig. 1.

In the next section an explicit formula will be given for the maximization in line 5 of the algorithm. The work for the kth iteration of the **for** loop is $O(k)$, and so the computational cost is $O(n^2)$ and the storage is $O(n)$.

With this algorithm the error converges to zero for any continuous function. Furthermore, there exists a sequence of stopping times $\{n_k\}$ such that under the Wiener measure, there exist constants $0 < c_1 < c_2 < \infty$ such that for any $y \in (0, \infty)$,

$$P(\exp(c_1 n_k \gamma_{n_k}^{-2})\, \Delta_{n_k} > y) \to 0$$

and

$$P(\exp(c_2 n_k \gamma_{n_k}^{-2})\, \Delta_{n_k} < y) \to 0.$$

The precise statement is in Theorem 6.2 in the last section.

In Section 2 we introduce the notation and the basic calculations underlying the algorithm. In Section 3 we show that the algorithm converges for any continuous function. Sections 4 and 5 contain a collection of auxiliary results on the behavior of the algorithm that are used in Section 6 to bound the rate at which the error converges to zero.

2. DEFINITION OF THE ALGORITHM

The purpose of this section is to introduce the necessary notation and describe the basic calculations underlying the general form of the algorithm (that is, for any positive sequence $\{z_n\}$).

Let $\Omega = C([0, 1])$ be the space of continuous functions from $[0, 1]$ into $\mathbb{R}$. For $0 \leqslant t \leqslant 1$ and $\omega \in \Omega$ let $X(t, \omega) = \omega(t)$ denote the position of the path ω at time t. Let $\mathscr{F} = \sigma\{X(s): 0 \leqslant s \leqslant 1\}$ be the σ-field generated by the process, and denote by P the Wiener measure on $(\Omega, \mathscr{F})$. Under P, the coordinate process X is a standard Brownian motion starting from 0. Let $t^* = \inf\{t \in [0, 1] : X(t) = M\}$ denote the (first) global minimizer.

We will denote the set of observation points by $\{t_0 = 0, t_1 = 1, t_2, ...\}$. Since $X(0) = 0$, $t_0 = 0$ is not really an observation point, but it will be notationally convenient to treat it as one. Because $X(t)$ is normally distributed with mean 0 and variance t, $P(X(t) < -z_1)$ is maximized by $t = 1$, so that $t_1 = 1$.

In addition to the sequences of observation sites $\{t_0, t_1, t_2, ...\}$ and values $\{x_0, x_1, x_2, ...\}$, it will be necessary to refer to the ordered observations for each fixed n. Therefore, for $n \geqslant 2$ let

$$0 = t_0^n < t_1^n < t_2^n < \cdots < t_{n-1}^n < t_n^n = 1$$

be the ordered observations, so that $\{t_i^n : i \leqslant n\} = \{t_i : i \leqslant n\}$, and denote the corresponding observed function values by $\{x_0^n, x_1^n, x_2^n, ..., x_n^n\}$; i.e., $X(t_i^n) = x_i^n$, $i \leqslant n$.

We now summarize the basic calculations underlying the algorithm; these are the same whether z_n is adaptive or not, and the details are given in [7] or [10]. Conditional on $t_1, t_2, ..., t_n$ and $X(t_1) = x_1$, $X(t_2) = x_2$, ..., $X(t_n) = x_n$, for $t_{i-1}^n < t < t_i^n$,

$$X(t) \sim N\left(\frac{t - t_{i-1}^n}{t_i^n - t_{i-1}^n}\, x_i^n + \frac{t_i^n - t}{t_i^n - t_{i-1}^n}\, x_{i-1}^n, \frac{(t - t_{i-1}^n)(t_i^n - t)}{t_i^n - t_{i-1}^n}\right),$$

where $N(a, b)$ denotes a normal random variable with mean a and variance b. The value of $t \in (t_{i-1}^n, t_i^n)$ that maximizes

$$P(X(t) < M_n - z_n \mid X(t_1) = x_1, ..., X(t_n) = x_n) \tag{6}$$

is given by

$$t = t_{i-1}^n + (t_i^n - t_{i-1}^n)\, \frac{x_{i-1}^n - M_n + z_n}{x_{i-1}^n - M_n + z_n + x_i^n - M_n + z_n}, \tag{7}$$

with the conditional probability in (6) then given by

$$1 - \Phi\left(\frac{2\sqrt{(x_{i-1}^n - M_n + z_n)(x_i^n - M_n + z_n)}}{\sqrt{t_i^n - t_{i-1}^n}}\right), \tag{8}$$

where Φ is the normal cumulative distribution function. Therefore, maximizing the probability in (6) is equivalent to minimizing the argument of Φ in (8), or equivalently, maximizing the reciprocal of its square. Thus the algorithm chooses $i \leqslant n$ to maximize

$$\rho_i^n \triangleq \frac{t_i^n - t_{i-1}^n}{(x_i^n - M_n + z_n)(x_{i-1}^n - M_n + z_n)}, \tag{9}$$

and then chooses t_{n+1} according to (7). The quantity ρ_i^n is a measure of how much we expect to gain from placing the next observation (optimally) in the subinterval (t_{i-1}^n, t_i^n). Set $\rho^n \triangleq \max_{i \leqslant n} \rho_i^n$.

Let $\mathscr{F}_n = \sigma\{X(t_i^n), i \leqslant n\}$ be the sigma-field of subsets of $\mathscr{F}$ representing the information obtained by the searcher through the first n observations. Denote the linear interpolation between observed values by

$$L_n(s) = \frac{t_i^n - s}{t_i^n - t_{i-1}^n} X(t_{i-1}^n) + \frac{s - t_{i-1}^n}{t_i^n - t_{i-1}^n} X(t_i^n), \qquad t_{i-1}^n \leqslant s \leqslant t_i^n, \qquad 0 \leqslant i \leqslant n. \tag{10}$$

As noted above,

$$E(X(t) \mid \mathscr{F}_n) = L_n(t), \qquad 0 \leqslant t \leqslant 1.$$

In order to simplify many expressions in the paper, define the translated process

$$Y_n(t) = X(t) - M_n + z_n, \qquad 0 \leqslant t \leqslant 1.$$

$Y_n(t)$ is the height of $X(t)$ above the lower boundary $M_n - z_n$ that we are trying to undershoot.

We can express ρ_i^n as an integral according to the formula

$$\rho_i^n = \frac{t_i^n - t_{i-1}^n}{Y_n(t_i^n)\; Y_n(t_{i-1}^n)} = \int_{s = t_i^n - 1}^{t_i^n} (L_n(s) - M_n + z_n)^{-2}\, ds. \tag{11}$$

The triangular array $\{\rho_i^n : 0 \leqslant i \leqslant n, n \geqslant 2\}$ will play a central role in determining the performance of the algorithm.

In summary, using the notation introduced above: After having evaluated $X(t_i^n)$, $i \leqslant n$, the algorithm operates by calculating the ρ_i^n, $i \leqslant n$.

The new observation is made in the interval $(t^n_{i_n-1}, t^n_{i_n})$, where $\rho^n_{i_n} = \max_{i \leqslant n} \rho^n_i = \rho^n$, at the location

$$t_{n+1} = t^n_{i_n-1} + (t^n_{i_n} - t^n_{i_n-1}) \frac{Y_n(t^n_{i_n-1})}{Y_n(t^n_{i_n-1}) + Y_n(t^n_{i_n})}. \tag{12}$$

We will say that the algorithm "splits interval i at time $n+1$" when $t_{n+1} \in (t^n_{i-1}, t^n_i)$.

3. CONVERGENCE

We turn now to the question of convergence of the algorithm, specifically the question of whether $M_n \downarrow M$ for any continuous function. The answer to this question depends on the choice of sequence $\{z_n\}$. Convergence is easy to establish if $\{z_n\}$ is bounded below by a positive number (see [3]), but a much better rate of convergence is obtained by letting the sequence approach 0. If $z_n \to 0$ too quickly, it is possible for M_n to converge to a number strictly larger than the global minimum; the trick is to find the appropriate rate. We will be more specific later in this section (after Lemma 3.1), but for now assume only that $z_n > 0$ and $z_n \to 0$.

Three subintervals will figure prominently in the study of the convergence properties of the algorithm. The following definitions are illustrated in Fig. 2. Denote by t^n_L and t^n_R the closest observation points on the left and on the right of t^*, respectively (so $t^n_L \leqslant t^* < t^n_R$), and let

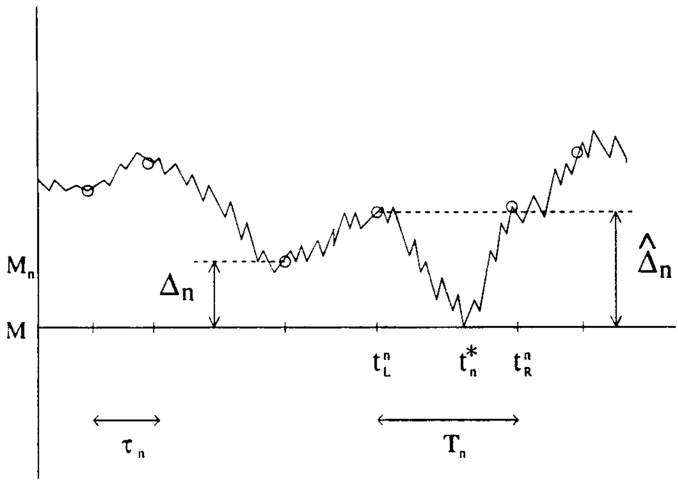

FIG. 2. Path near minimizer.

$T_n = t_R^n - t_L^n$ be the width of the subinterval straddling t^* after n observations. Let $[s_L^n, s_R^n]$ denote the subinterval to be split at time n, and denote its width by $S_n = s_R^n - s_L^n$. Therefore, after n observations, $t_{n+1} \in (s_L^n, s_R^n)$. Finally, let τ_n be the width of the smallest subinterval after n observations; that is,

$$\tau_n = \min_{1 \leqslant i \leqslant n} t_i^n - t_{i-1}^n.$$

Because of the way the algorithm chooses observation points, one would expect that for sufficiently regular functions the straddling interval length T_n would not be much larger than the smallest interval width τ_n (if they are not in fact equal). One of our main goals in the next section is to show that under the Wiener measure this is the case with probability approaching one, but in this section we are not concerned with probabilistic considerations.

We say that the algorithm converges if $M_n \downarrow M$ for any continuous function. We begin with a criterion for convergence in terms of the sequence $\{\rho^n\}$ defined at (9).

LEMMA 3.1. *The algorithm converges if and only if* $\liminf \rho^n = 0$.

Proof. Suppose that the error does not converge to 0; i.e., suppose that $M_n \downarrow \bar{M} > M$. Then $t_L^n \uparrow t_L < t^*$ and $t_R^n \downarrow t_R > t^*$ where $X(t_L) > M$ and $X(t_R) > M$, and

$$\liminf \rho^n \geqslant \liminf \frac{t_R^n - t_L^n}{(X(t_R^n) - M_n + z_n)(X(t_L^n) - M_n + z_n)}$$

$$= \frac{t_R - t_L}{(X(t_R) - \bar{M})(X(t_L) - \bar{M})} > 0.$$

For the other direction, suppose that $\liminf \rho^n > 0$. We will show that in this case there exists a continuous function for which M_n does not converge to M. Eventually, after an interval $[t_{i-1}^n, t_i^n]$ is such that

$$\rho_i^n = \frac{t_i^n - t_{i-1}^n}{(X(t_i^n) - M_n + z_n)(X(t_{i-1}^n) - M_n + z_n)}$$

$$< \frac{t_i^n - t_{i-1}^n}{(X(t_i^n) - M_n)(X(t_{i-1}^n) - M_n)} < \liminf \rho^n,$$

it will never subsequently be split. But then the algorithm will fail to converge for the continuous function X modified to have its global minimum in (t_{i-1}^n, t_i^n). ∎

Now we will be more specific about the sequence $\{z_n\}$. Let $\{\gamma_n\}$ be an increasing sequence of positive numbers satisfying (5), and set $z_n = \min_{i \leqslant n} \sqrt{\tau_i}\, \gamma_i$. This sequence satisfies the assumption made in the beginning of the section that $z_n \to 0$, since the smallest interval is at most of average length, so $\tau_n \leqslant 1/n$ and

$$z_n \leqslant \sqrt{\tau_n}\, \gamma_n \leqslant \frac{\gamma_n}{\sqrt{n}} \to 0$$

by (5).

Let m_k be the kth time that the sequence $\{z_n\}$ is about to decrease. That is, $z_{m_k+1} < z_{m_k}$ and $z_{j+1} = z_j$ for $j \notin \{m_k\}$. The (m_k+1)st observation creates a new smallest interval, since the $\{\gamma_k\}$ are increasing. Therefore, $\tau_{m_k+1} < \tau_{m_k}$, and also

$$z_{m_k+1} = \sqrt{\tau_{m_k+1}}\, \gamma_{m_k+1}.$$

Notice that (5) bounds how quickly γ_n can increase, but as long as $\gamma_n \uparrow \infty$ it can increase arbitrarily slowly.

For the remainder of the paper, by "the algorithm" we mean the rule given at (1) with the sequence $\{z_n\}$ defined above. This is a class of algorithms for different choices of $\{\gamma_n\}$ satisfying (5).

THEOREM 3.1. *The algorithm converges for any continuous function.*

Proof. We will show that $\rho^{m_k} \to 0$, which, in light of Lemma 3.1, will prove the theorem.

Recall that $[s_L^{m_k}, s_R^{m_k}]$ is the interval to be split at time m_k. We will use the notation $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. Using (7), the smallest new interval will be of length

$$\tau_{m_k+1} = (s_R^{m_k} - s_L^{m_k}) \frac{Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_R^{m_k})}{Y_{m_k}(s_L^{m_k}) + Y_{m_k}(s_R^{m_k})}. \tag{13}$$

Since we are assuming that a new smallest interval is formed, $\tau_{m_k+1} < \tau_{m_k}$ and so (13) implies that

$$\tau_{m_k} > (s_R^{m_k} - s_L^{m_k}) \frac{Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_R^{m_k})}{Y_{m_k}(s_L^{m_k}) + Y_{m_k}(s_R^{m_k})}. \tag{14}$$

Using (14) to obtain the first inequality below,

$$\begin{aligned}
\rho^{m_k} &= \frac{s_R^{m_k} - s_L^{m_k}}{Y_{m_k}(s_L^{m_k})\, Y_{m_k}(s_R^{m_k})} \\
&< \frac{\tau_{m_k}(Y_{m_k}(s_L^{m_k}) + Y_{m_k}(s_R^{m_k}))}{Y_{m_k}(s_L^{m_k})\, Y_{m_k}(s_R^{m_k})(Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_R^{m_k}))} \\
&= \tau_{m_k}\left(\frac{1}{Y_{m_k}(s_L^{m_k})\, Y_{m_k}(s_R^{m_k})} + \frac{1}{(Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_R^{m_k}))^2}\right) \\
&\leqslant \tau_{m_k}\left(\frac{1}{z_{m_k}^2} + \frac{1}{z_{m_k}^2}\right) \\
&= \frac{2}{\gamma_{m_k}^2},
\end{aligned}$$

which converges to 0. Therefore, $\rho^{m_k} \to 0$, as was to be shown. ∎

4. FURTHER CONVERGENCE PROPERTIES

Up to now the analysis has been non-probabilistic. In this section we begin the task of deriving probabilistic bounds on the error Δ_n. We will begin by analyzing the random variables $\hat{\Delta}_n = \min\{X(t_R^n), X(t_L^n)\} - M$, which are the smallest of the two straddling values (which may exceed the minimum value observed up to time n) minus the global minimum, thus $\Delta_n \leqslant \hat{\Delta}_n$ (see Fig. 2). In this section we will establish some basic properties of the algorithm that will be used later. These properties include the relationship among the interval lengths τ_{m_k}, T_{m_k}, and S_{m_k}.

For random variables $\{X_n\}$, X, we use the notation $X_n \xrightarrow{P} X$ to indicate that X_n converges in probability to X; that is, for any $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) \to 0.$$

The notation $X_n \xrightarrow{\mathscr{D}} X$ (or $X_n \xrightarrow{\mathscr{D}} \mu$) indicates that X_n converges in distribution to X (or to the probability measure μ); that is, the probability law of X_n converges weakly to the law of X (or to μ). By $X_n \xrightarrow{P} \infty$ or $X_n \xrightarrow{\mathscr{D}} \infty$ we mean that for any $A < \infty$, $P(X_n \leqslant A) \to 0$.

LEMMA 4.1. *As $k \to \infty$,*

$$\frac{S_{m_k}}{\tau_{m_k+1}} \xrightarrow{P} 2; \tag{15}$$

i.e., the intervals $[s_L^{m_k}, s_R^{m_k}]$ are eventually approximately bisected.

Proof. By (12),

$$\begin{aligned}
\tau_{m_k+1} &= S_{m_k} \frac{Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_R^{m_k})}{Y_{m_k}(s_L^{m_k}) + Y_{m_k}(s_R^{m_k})} \\
&= S_{m_k}\left(2 + \frac{|Y_{m_k}(s_R^{m_k}) - Y_{m_k}(s_L^{m_k})|}{\sqrt{S_{m_k}}} \frac{\sqrt{S_{m_k}}}{Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_R^{m_k})}\right)^{-1},
\end{aligned}$$

so that

$$\frac{S_{m_k}}{\tau_{m_k+1}} = 2 + \frac{|Y_{m_k}(s_R^{m_k}) - Y_{m_k}(s_L^{m_k})|}{\sqrt{S_{m_k}}} \frac{\sqrt{S_{m_k}}}{Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_R^{m_k})}. \tag{16}$$

By straightforward rearrangement,

$$\begin{aligned}
&\frac{S_{m_k}}{(Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_R^{m_k}))^2} \\
&\quad = \frac{S_{m_k}}{Y_{m_k}(s_L^{m_k})\, Y_{m_k}(s_R^{m_k})} \frac{Y_{m_k}(s_L^{m_k}) \vee Y_{m_k}(s_R^{m_k})}{Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_R^{m_k})} \\
&\quad = \rho^{m_k}\left(1 + \frac{|Y_{m_k}(s_R^{m_k}) - Y_{m_k}(s_L^{m_k})|}{\sqrt{S_{m_k}}} \frac{\sqrt{S_{m_k}}}{Y_{m_k}(s_R^{m_k}) \wedge Y_{m_k}(s_L^{m_k})}\right).
\end{aligned}$$

Solving this quadratic equation gives

$$\begin{aligned}
\frac{\sqrt{S_{m_k}}}{Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_L^{m_k})} &= \frac{1}{2}\rho^{m_k} \frac{|Y_{m_k}(s_R^{m_k}) - Y_{m_k}(s_L^{m_k})|}{\sqrt{S_{m_k}}} \\
&\quad + \left(\rho^{m_k} + \left(\frac{1}{2}\rho^{m_k} \frac{|Y_{m_k}(s_R^{m_k}) - Y_{m_k}(s_L^{m_k})|}{\sqrt{S_{m_k}}}\right)^2\right)^{1/2} \xrightarrow{P} 0
\end{aligned}$$

since $|Y_{m_k}(s_R^{m_k}) - Y_{m_k}(s_L^{m_k})|/\sqrt{S_{m_k}} = o_P(1)$ and $\rho^{m_k} \to 0$. It follows that

$$\frac{\sqrt{S_{m_k}}}{Y_{m_k}(s_L^{m_k}) \wedge Y_{m_k}(s_R^{m_k})} \xrightarrow{P} 0. \tag{17}$$

We conclude from (17) and (16) that

$$\frac{S_{m_k}}{\tau_{m_k+1}} \xrightarrow{P} 2. \quad \blacksquare$$

Let R and R' be independent 3-dimensional Bessel processes; that is, $\{R(t) : t \geqslant 0\}$ has the same distribution as

$$\{\sqrt{B_1(t)^2 + B_2(t)^2 + B_3(t)^2} : t \geqslant 0\}, \tag{18}$$

where the B_i are independent one-dimensional Wiener processes. We can think informally of R as a Wiener process killed on hitting 0, conditioned to never hit 0. Notice from (18) that R inherits the Brownian scaling properties; e.g., $R(t) \stackrel{\mathscr{D}}{=} \sqrt{t}\, R(1)$, where "$\stackrel{\mathscr{D}}{=}$" denotes equality in distribution.

The following theorem characterizes the "local" error; that is, the error based only on the nearest observation to the left and right of the global minimizer.

THEOREM 4.1. *Let U be uniformly distributed on* $(0, 1)$, *independent of* R, R'. *Then as* $n \to \infty$,

$$\left(\frac{X(t^n_L) - M}{\sqrt{t^n_R - t^n_L}}, \frac{X(t^n_R) - M}{\sqrt{t^n_R - t^n_L}}, \frac{t^* - t^n_L}{t^n_R - t^n_L}\right) \xrightarrow{\mathscr{D}} (R(U), R'(1-U), U).$$

Proof. Let

$$\rho^n_s = \frac{t^n_R - t^n_L}{Y_n(t^n_L)\, Y_n(t^n_R)} = \frac{t^n_R - t^n_L}{(X(t^n_L) - M_n + z_n)(X(t^n_R) - M_n + z_n)} \tag{19}$$

denote the ρ value for the subinterval straddling t^*. We know from Theorem 3.1 that $\rho^{m_k}_s \leqslant \rho^{m_k} \to 0$, but we have not determined how ρ^n_s behaves otherwise; we will show that $\rho^n_s \xrightarrow{P} 0$. Observe from (19) that ρ^n_s is decreasing in n except for the times when τ_n decreases; that is, when $n \in \{m_k\}$. At these times, the factor by which $\rho^{m_k}_s$ increases is at most

$$\frac{\rho^{m_k+1}_s}{\rho^{m_k}_s} \leqslant \left(\frac{z_{m_k}}{z_{m_k+1}}\right)^2 \leqslant \frac{\tau_{m_k}}{\tau_{m_k+1}}.$$

Since $S^{m_k}/\tau_{m_k+1} \xrightarrow{P} 2$ by Lemma 3.1 and $S^{m_k} \geqslant \tau_{m_k}$,

$$P(\tau_{m_k}/\tau_{m_k+1} > 2+\varepsilon) \leqslant P(S^{m_k}/\tau_{m_k+1} > 2+\varepsilon) \to 0$$

for any $\varepsilon > 0$. Therefore, if $m_j(n) \leqslant n < m_{j+1}(n)$, then

$$P(\rho_s^n > (2+\varepsilon)\, \rho^{m_j(n)}) \to 0$$

and so $\rho_s^n \xrightarrow{P} 0$ as $n \to \infty$, since $\rho^{m_j(n)} \to 0$.

For the remainder of this proof we focus on the subinterval straddling t^* and the times at which the next observation is made in (t_L^n, t_R^n). Instead of introducing new notation for such a sequence, let the index n indicate the nth time that (t_L^n, t_R^n) is split.

Let $P_{t,m,x}$ denote the conditional distribution of X given $(t^*, M, X(1)) = (t, m, x)$, and consider the Markov chain (under $P_{t,m,x}$ with respect to $\mathscr{F}_n$)

$$W'_n = (Y'_n, Z'_n, U'_n) = \left(\frac{X(t_L^n) - M}{\sqrt{t^* - t_L^n}}, \frac{X(t_R^n) - M}{\sqrt{t_R^n - t^*}}, \frac{t^* - t_L^n}{t_R^n - t_L^n} \right).$$

Under $P_{t,m,x}$, $\{X(t^* - s) - M\}$ and $\{X(t^* + s) - M\}$ are independent three-dimensional Bessel bridges ([5]), and so if we condition on $R(t) = -m$ and $R'(1-t) = x - m$, W' has the same distribution as

$$W_n = (Y_n, Z_n, U_n) = \left(\frac{R(t^* - t_L^n)}{\sqrt{t^* - t_L^n}}, \frac{R'(t_R^n - t^*)}{\sqrt{t_R^n - t^*}}, \frac{t^* - t_L^n}{t_R^n - t_L^n} \right).$$

The family of probability distributions induced by the $\{W_n\}$ is tight in the sense that for any $\varepsilon > 0$ there exists a compact set K_ε such that $\inf_n P(W_n \in K_\varepsilon) \geqslant 1 - \varepsilon$. Since U_n is bounded and Y_n and Z_n have the same distribution, it suffices to show that there exists $K_\varepsilon < \infty$ such that

$$P(Z_n \leqslant K_\varepsilon) \geqslant 1 - \varepsilon, \tag{20}$$

for which it suffices (using Markov's inequality) to show that

$$\sup_n E(Z_n^2) = \sup_n E\left(\frac{R'(t_R^n - t^*)^2}{t_R^n - t^*} \right) < \infty. \tag{21}$$

Using the representation (18) and the fact that t_R^n is independent of $\{R'(s) : 0 \leqslant s \leqslant t_R^{n-1}\}$, it can be shown that

$$E(Z_n^2) = E\left(\frac{t_R^n}{t_R^{n-1}}\, Z_{n-1}^2 + 3\,\frac{t_R^{n-1} - t_R^n}{t_R^{n-1}}\right),$$

and so

$$|E(Z_n^2) - 3| \leqslant |E(Z_{n-1}^2) - 3|.$$

Thus (21) is satisfied.

Let θ_n be the division point of the straddling interval on the nth time it is split. From (12),

$$\theta_n = \frac{X(t_L^n) - M_n + z_n}{X(t_L^n) - M_n + z_n + X(t_L^n) - M_n + z_n}. \tag{22}$$

Then we can write $\theta_n = (2 + \varepsilon_n)^{-1}$, where

$$\begin{aligned}
\varepsilon_n &= \frac{X(t_R^n) - X(t_L^n)}{X(t_L^n) - M_n + z_n} \\
&= \frac{X(t_R^n) - X(t_L^n)}{\sqrt{(X(t_L^n) - M_n + z_n)(X(t^n{}_R) - M_n + z_n)}} \left(\frac{X(t_R^n) - M_n + z_n}{X(t_L^n) - M_n + z_n}\right)^{1/2} \\
&= \sqrt{\rho_s^n}\, \frac{X(t_R^n) - X(t_L^n)}{\sqrt{t_R^n - t_L^n}} \left(\frac{X(t_R^n) - X(t_L^n)}{X(t_L^n) - M_n + z_n} + 1\right)^{1/2} \\
&= \sqrt{\rho_s^n}\, (Y_n \sqrt{U_n} - Z_n \sqrt{1 - U_n})(\varepsilon_n + 1)^{1/2}.
\end{aligned}$$

Since $\rho_s^n \xrightarrow{P} 0$ and, by (20), Y_n and Z_n are bounded in probability, this implies that $\varepsilon_n \xrightarrow{P} 0$, and so $\theta_n \xrightarrow{P} 1/2$. Furthermore, for any $m > 0$ and $\delta > 0$,

$$P(\tfrac{1}{2} - \delta < \theta_{n+i} < \tfrac{1}{2} + \delta;\, 0 \leqslant i \leqslant m \mid Y_n = y_n,\, Z_n = z_n,\, U_n = u_n) \to 1 \tag{23}$$

as $n \to \infty$.

For $(y, z, u), (y_n, z_n, u_n) \in \mathbb{R}_+^2 \times [0, 1)$, let

$$F_{n, n+m}((y_n, z_n, u_n), (y, z, u))$$
$$= P(Y_{n+m} \leqslant y, Z_{n+m} \leqslant z, U_{n+m} \leqslant u \mid Y_n = y_n, Z_n = z_n, U_n = u_n)$$

and let

$$F_m((y_n, z_n, u_n), (y, z, u))$$
$$= P\left(\frac{R(2^{-m}(2^m u_n (\bmod 1)))}{\sqrt{2^{-m}(2^m u_n (\bmod 1))}} \leqslant y \;\middle|\; \frac{R(u_n)}{\sqrt{u_n}} = y_n\right)$$
$$\cdot P\left(\frac{R'(2^{-m}(1-2^m u_n (\bmod 1)))}{\sqrt{2^{-m}(1-2^m u_n (\bmod 1))}} \leqslant z \;\middle|\; \frac{R'(1-u_n)}{\sqrt{1-u_n}} = z_n\right)$$
$$\cdot I_{\{2^m u_n (\bmod 1) \leqslant u\}}.$$

F_m represents the m-step transition probabilities for the chain that evolves according to the rule $U_{n+1} = 2U_n (\bmod 1)$ (i.e., $\theta_n = 1/2$ for all n). For fixed (y, z, u), (23) implies that

$$\lim_{n \to \infty} F_{n, n+m}((y_n, z_n, u_n), (y, z, u))) - F_m((y_n, z_n, u_n), (y, z, u))) = 0 \quad (24)$$

for each $(y_n, z_n, u_n) \in \mathbb{R}_+^2 \times ([0, 1) \backslash B_m)$, where

$$B_m = \{k2^{-m}, (k+z)\, 2^{-m}; k = 0, 1, ..., 2^m - 1\};$$

i.e., at all points except the discontinuities of $F_m(\cdot, (y, z, u))$.

Let μ denote the probability distribution of $(R(1), R'(1), U)$, and let μ_n denote the distribution of W_n. Since the family $\{\mu_n\}$ is tight, a theorem of Prohorov ([1], p. 37) implies that any subsequence of the $\{\mu_n\}$ has a further subsequence that converges weakly to a probability measure. We will show that any such subsequence must converge to μ, and thus μ_n converges weakly to μ.

Suppose that there exists a subsequence n_j and an absolutely continuous measure v such that $P(W_{n_j} \in A) = \mu_{n_j}(A) \to v(A)$ for all Borel sets $A \subset \mathbb{R}_+^2 \times [0, 1)$. Notice that each μ_{n_j} is absolutely continuous. Then (in the following let $w_n = (y_n, z_n, u_n)$, $w = (y, z, u)$, and take all integrals over $\mathbb{R}_+^2 \times [0, 1)$)

$$P(Y_{n_j+m} \leqslant y, Z_{n_j+m} \leqslant z, U_{n_j+m} \leqslant u) - P(R(1) \leqslant y)\, P(R(1) \leqslant z)\, u$$

$$= \int F_{n_j, n_j+m}(w_n, w)\, d\mu_{n_j}(w_n) - \int F_m(w_n, w)\, d\mu(w_n)$$

$$= \int F_{n_j, n_j+m}(w_n, w)\, d\mu_{n_j}(w_n) - \int F_m(w_n, w)\, d\mu_{n_j}(w_n) \tag{25}$$

$$+ \int F_m(w_n, w)\, d\mu_{n_j}(w_n) - \int F_m(w_n, w)\, dv(w_n) \tag{26}$$

$$+ \int F_m(w_n, w)\, dv(w_n) - \int F_m(w_n, w)\, d\mu(w_n). \tag{27}$$

The difference in (25) converges to 0 as $j \to \infty$ by the dominated convergence theorem and (24), since $\mu_{n_j}(B_m) = 0$ for each j. The difference in (26) goes to 0 as $j \to \infty$ since μ_{n_j} converges weakly to v and $v(B_m) = 0$, and the difference in (27) converges to 0 because v is continuous. Therefore, letting j and $m \to \infty$, we conclude that $v = \mu$; i.e.,

$$W_n = \left(\frac{R(t^* - t_L^n)}{\sqrt{t^* - t_L^n}}, \frac{R'(t_R^n - t^*)}{\sqrt{t_R^n - t^*}}, \frac{t^* - t_L^n}{t_R^n - t_L^n} \right) \xrightarrow{\mathscr{D}} (R(1), R'(1), U).$$

The triple

$$\left(\frac{X(t_L^n) - M}{\sqrt{t_R^n - t_L^n}}, \frac{X(t_R^n) - M}{\sqrt{t_R^n - t_L^n}}, \frac{t^* - t_L^n}{t_R^n - t_L^n} \right)$$

has the same limiting distribution as

$$\left(\frac{R(t^* - t_L^n)}{\sqrt{t^* - t_L^n}}, \frac{R'(t_R^n - t^*)}{\sqrt{t_R^n - t^*}}, \frac{t^* - t_L^n}{t_R^n - t_L^n} \right) = (Y_n \sqrt{U_n}, Z_n \sqrt{1 - U_n}, U_n).$$

By continuity of the map $(y, z, u) \to (y \sqrt{u}, z \sqrt{1-u}, u)$,

$$\begin{aligned} (Y_n \sqrt{U_n}, Z_n \sqrt{1 - U_n}, U_n) &\xrightarrow{\mathscr{D}} (R(1) \sqrt{U}, R'(1) \sqrt{1 - U}, U) \\ &\overset{\mathscr{D}}{=} (R(U), R'(1 - U), U), \end{aligned}$$

and the proof is complete. ∎

Using the continuity of $(x, y) \to (x \wedge y, x \vee y)$, we obtain

COROLLARY 4.1. *As* $n \to \infty$,

$$\frac{X(t_L^n) \vee X(t_R^n) - M}{\sqrt{T_n}} \xrightarrow{\mathscr{D}} R(U) \vee R'(1-U),$$

and

$$\frac{X(t_L^n) \wedge X(t_R^n) - M}{\sqrt{T_n}} = \frac{\hat{\Delta}_n}{\sqrt{T_n}} \xrightarrow{\mathscr{D}} R(U) \wedge R'(1-U). \tag{28}$$

In order to use (28) to establish how fast $\hat{\Delta}_n$ converges to 0, we need to determine how fast T_n converges to 0. We will determine how fast the τ_n converge to 0 in Theorem 6.1. The following Lemma shows that the T_n are not much larger than the τ_n, so the bounds we develop for the latter will imply bounds for the former.

The following lemma shows that the straddling subinterval is unlikely to be much larger than the smallest interval.

LEMMA 4.2. *For any* $\varepsilon > 0$,

$$P\left(\frac{T_{m_k}}{S_{m_k}} > 2 + \varepsilon\right) \to 0 \tag{29}$$

and

$$P\left(\frac{T_{m_k}}{\tau_{m_k}} > 4 + \varepsilon\right) \to 0. \tag{30}$$

Proof. Since by assumption at time m_k the interval $[s_L^{m_k}, s_R^{m_k}]$ is about to be split,

$$\rho^{m_k} = \frac{S_{m_k}}{Y_{m_k}(s_L^{m_k})\, Y_{m_k}(s_R^{m_k})} \geqslant \frac{T_{m_k}}{Y_{m_k}(t_L^{m_k})\, Y_{m_k}(t_R^{m_k})}.$$

This implies that

$$\begin{aligned}
\frac{T_{m_k}}{S_{m_k}} &\leqslant \frac{Y_{m_k}(t_L^{m_k})\, Y_{m_k}(t_R^{m_k})}{Y_{m_k}(s_L^{m_k})\, Y_{m_k}(s_R^{m_k})} \\
&= \frac{(X(t_L^{m_k}) - M_{m_k} + z_{m_k})(X(t_R^{m_k}) - M_{m_k} + z_{m_k})}{(X(s_L^{m_k}) - M_{m_k} + z_{m_k})(X(s_R^{m_k}) - M_{m_k} + z_{m_k})} \\
&\leqslant \left(1 + \frac{X(t_L^{m_k}) - M}{z_{m_k}}\right)\left(1 + \frac{X(t_R^{m_k}) - M}{z_{m_k}}\right)
\end{aligned}$$

$$
\begin{aligned}
&\leqslant\left(1+\frac{X(t_L^{m_k}) \vee X(t_R^{m_k})-M}{z_{m_k+1}}\right)^2\\
&=\left(1+\frac{X(t_L^{m_k}) \vee X(t_R^{m_k})-M}{\sqrt{T_{m_k}}}\frac{\sqrt{T_{m_k}}}{\sqrt{S_{m_k}}}\frac{\sqrt{S_{m_k}}}{\sqrt{\tau_{m_k+1}}}\frac{1}{\gamma_{m_k+1}}\right)^2\\
&\leqslant 2+2\left(\frac{X(t_L^{m_k}) \vee X(t_R^{m_k})-M}{\sqrt{T_{m_k}}}\right)^2\frac{T_{m_k}}{S_{m_k}}\frac{S_{m_k}}{\tau_{m_k+1}}\frac{1}{\gamma^2_{m_k+1}},
\end{aligned}\tag{31}
$$

using the inequality $(1+a)^2 \leqslant 2+2a^2$. Define

$$
\begin{aligned}
\psi_{m_k} &\triangleq 2\left(\frac{X(t_L^{m_k}) \vee X(t_R^{m_k})-M}{\sqrt{T_{m_k}}}\right)^2\frac{S_{m_k}}{\tau_{m_k+1}}\frac{1}{\gamma^2_{m_k+1}}\\
&=2\left(\frac{X(t_L^{m_k}) \vee X(t_R^{m_k})-M}{\sqrt{T_{m_k}}}\frac{1}{\gamma_{m_k+1}}\right)^2\frac{S_{m_k}}{\tau_{m_k+1}}.
\end{aligned}
$$

Then $\psi_{m_k} \xrightarrow{P} 0$, since

$$
\frac{X(t_L^{m_k}) \vee X(t_R^{m_k})-M}{\sqrt{T_{m_k}}}\frac{1}{\gamma_{m_k+1}} \xrightarrow{P} 0
$$

by Corollary 4.1, and

$$
\frac{S_{m_k}}{\tau_{m_k+1}} \xrightarrow{P} 2
$$

by Lemma 4.1. We can express (31) as

$$
\frac{T_{m_k}}{S_{m_k}} \leqslant 2+\psi_{m_k}\frac{T_{m_k}}{S_{m_k}}.
$$

Therefore, if $\varepsilon>0$, then

$$
P\left(\frac{T_{m_k}}{S_{m_k}}>2+\varepsilon\right) \leqslant P\left(\psi_{m_k}>\frac{\varepsilon}{2+\varepsilon}\right),\tag{32}
$$

which converges to 0 since $\psi_{m_k} \xrightarrow{P} 0$.

For the second part (30), for $\varepsilon>0$,

$$
P\left(\frac{T_{m_k}}{\tau_{m_k}}>4+\varepsilon\right)=P\left(\frac{T_{m_k}}{S_{m_k}}\frac{S_{m_k}}{\tau_{m_k}}>4+\varepsilon\right) \leqslant P\left(\frac{T_{m_k}}{S_{m_k}}\frac{S_{m_k}}{\tau_{m_k+1}}>4+\varepsilon\right).
$$

Now

$$\frac{S_{m_k}}{\tau_{m_k+1}} \xrightarrow{P} 2$$

by Lemma 4.1, and

$$\max\left(\frac{T_{m_k}}{S_{m_k}}, 2\right) \xrightarrow{P} 2$$

by (32). Therefore,

$$P\left(\frac{T_{m_k}}{S_{m_k}} \frac{S_{m_k}}{\tau_{m_k+1}} > 4+\varepsilon\right) \to 0,$$

which establishes (30). ∎

We will need the following lemma in the next section.

LEMMA 4.3. *As* $k \to \infty$,

$$\frac{X(t_L^{m_k}) \vee X(t_R^{m_k}) - M}{z_{m_k}} \xrightarrow{P} 0.$$

This implies, in particular, that

$$\frac{M_{m_k} - M}{z_{m_k}} \xrightarrow{P} 0.$$

Proof.

$$\begin{aligned}
\frac{X(t_L^{m_k}) \vee X(t_R^{m_k}) - M}{z_{m_k}} &\leqslant \frac{X(t_L^{m_k}) \vee X(t_R^{m_k}) - M}{z_{m_k+1}} \\
&= \frac{X(t_L^{m_k}) \vee X(t_R^{m_k}) - M}{\sqrt{T_{m_k}}} \frac{\sqrt{T_{m_k}}}{\sqrt{S_{m_k}}} \frac{\sqrt{S_{m_k}}}{\sqrt{\tau_{m_k+1}}} \frac{1}{\gamma_{m_k+1}} \\
&= \left(\frac{X(t_L^{m_k}) \vee X(t_R^{m_k}) - M}{\sqrt{T_{m_k}}} \frac{1}{\gamma_{m_k+1}}\right) \frac{\sqrt{T_{m_k}}}{\sqrt{S_{m_k}}} \frac{\sqrt{S_{m_k}}}{\sqrt{\tau_{m_k+1}}}.
\end{aligned}$$

The first term on the right converges to 0 in probability by Corollary 4.1, the second is bounded in probability by Lemma 4.2, and the third converges to $\sqrt{2}$ in probability by Lemma 4.1. Therefore the product converges in probability to 0. ∎

5. ASYMPTOTIC BEHAVIOR OF THE $\{\rho_i^n\}$

Corollary 4.1 describes the limit of the normalized "local" error $\hat{\Delta}_n T_n^{-1/2}$; to complete the picture we need to determine how fast T_n converges to 0. As previously noted, the convergence characteristics of the algorithm are determined by the triangular array $\{\rho_i^n\}$. Our next step is to use the auxiliary results derived so far to determine the speed at which the average of the ρ_i^n's converges to 0; from this we will be able to bound the rate that the smallest interval τ_n converges to zero. Since by (11)

$$\sum_{i=1}^{n} \rho_i^n = \int_{t=0}^{1} (L_n(t) - M_n + z_n)^{-2}\, dt, \tag{33}$$

it is natural to derive a bound based on the sequence of integrals

$$\int_{t=0}^{1} (X(t) - M_n + z_n)^{-2}\, dt. \tag{34}$$

In this section we first establish the rate at which the integrals in (34) converge to $+\infty$ and then show that the integrals in (34) and (33) do not differ by too much.

THEOREM 5.1. *As* $\varepsilon \downarrow 0$,

$$E\left(\frac{\int_{t=0}^{1} (X(t) - M + \varepsilon)^{-2}\, dt}{4\log(1/\varepsilon)}\right) \to 1. \tag{35}$$

Proof. The proof will involve various processes associated with the Wiener process. Let Z_k denote the square of a k-dimensional Bessel process. That is, Z_k is the diffusion equal in law to the sum of k squared one-dimensional Wiener processes.

We begin with a technical lemma.

LEMMA 5.1. *Fix* $\delta > 0$. *If* $\varepsilon \downarrow 0$, *then*

$$E\left(\frac{\int_{t=0}^{\delta} Z_k(t)(t+\varepsilon)^{-2}\, dt}{k\log(1/\varepsilon)}\right) \to 1. \tag{36}$$

Proof. Since Z_k is the squared modulus of a k-dimensional Brownian motion, it suffices to show that

$$E\left(\frac{\int_{t=0}^{\delta} B(t)^2 (t+\varepsilon)^{-2}\, dt}{\log(1/\varepsilon)}\right) \to 1, \tag{37}$$

where B is a standard one-dimensional Brownian motion. Since the integrand is nonnegative,

$$\frac{E\int_{t=0}^{\delta} B(t)^2 (t+\varepsilon)^{-2}\, dt}{\log(1/\varepsilon)} = \frac{\int_{t=0}^{\delta} EB(t)^2 (t+\varepsilon)^{-2}\, dt}{\log(1/\varepsilon)} = \frac{\int_{t=0}^{\delta} t(t+\varepsilon)^{-2}\, dt}{\log(1/\varepsilon)}$$

$$= \frac{\log(\delta+\varepsilon) - \log(\varepsilon) + \dfrac{\varepsilon}{\delta+\varepsilon} - 1}{\log(1/\varepsilon)} \to 1$$

as $\varepsilon \to 0$. ∎

Proof of Theorem 5.1. Conditional on M, t^*, and $X(1) = b$, the process

$$Y(t) = X(t^* + t) - M, \qquad 0 \leqslant t \leqslant 1 - t^*$$

is a 3-dimensional Bessel bridge from $(0, 0)$ to $(1 - t^*, b - M)$ (see [5]). Therefore, it suffices to show that for a (free) 3-dimensional Bessel process Y starting from 0,

$$\frac{E\int_{t=0}^{1-t^*} (Y(t)+\varepsilon)^{-2}\, dt}{2\log(1/\varepsilon)} \to 1 \tag{38}$$

(we consider only the integral to the right of t^* in the numerator; the integral to the left of t^* is handled in the same way). Fix $\delta < \min\{-M, b - M\}$. Since

$$\int_{t=0}^{1-t^*} (Y(t)+\varepsilon)^{-2}\, dt - \int_{t=0}^{1-t^*} I_{\{Y(t)\leqslant\delta\}} (Y(t)+\varepsilon)^{-2}\, dt \leqslant \frac{1}{\delta^2}, \tag{39}$$

it suffices to show that

$$\frac{E\int_{t=0}^{1-t^*} I_{\{Y(t)\leqslant\delta\}} (Y(t)+\varepsilon)^{-2}\, dt}{2\log(1/\varepsilon)} \to 1. \tag{40}$$

The local time process (occupation density) of Y is the square of a 2-dimensional Bessel process (see ([8], Ex. 2.5 of Chapt. XI), and so

$$\frac{\int_{t=0}^{\infty} I_{\{Y(t)\leqslant\delta\}}(Y(t)+\varepsilon)^{-2}\,dt}{2\log(1/\varepsilon)} = \frac{\int_{y=0}^{\delta} Z_2(y)(y+\varepsilon)^{-2}\,dt}{2\log(1/\varepsilon)}. \tag{41}$$

By Lemma 5.1,

$$E\left(\frac{\int_{t=0}^{\infty} I_{\{Y(t)\leqslant\delta\}}(Y(t)+\varepsilon)^{-2}\,dt}{2\log(1/\varepsilon)}\right) \to 1. \tag{42}$$

Therefore, to complete the proof of (38), it is enough to show that

$$\frac{E\int_{t=1-t^*}^{\infty} I_{\{Y(t)\leqslant\delta\}}(Y(t)+\varepsilon)^{-2}\,dt}{2\log(1/\varepsilon)} \to 0. \tag{43}$$

For this we will use Williams decomposition of the 3-dimensional Bessel process ([11]). Let be given a random variable β, uniformly distributed between 0 and $b-M$, a standard Brownian motion B starting at $b-M$, and a 3-dimensional Bessel process R starting at 0. Then we can write

$$Y(1-t^*+t) \overset{\mathscr{D}}{=} \begin{cases} B(t), & \text{if} \quad 0\leqslant t\leqslant T_\beta, \\ \beta+R(t-T_\beta), & \text{if} \quad T_\beta < t, \end{cases} \tag{44}$$

where $T_\beta = \inf\{t>0 : B(t)=\beta\}$; see Fig. 3.

If $\beta>\delta$, then the numerator in (40) is 0. Otherwise, if $\beta\leqslant\delta$, then

$$\begin{aligned}\int_{t=1-t^*}^{\infty} & I_{\{Y(t)\leqslant\delta\}}(Y(t)+\varepsilon)^{-2}\,dt \\ &= \int_{t=0}^{T_\beta} I_{\{B(t)\leqslant\delta\}}(B(t)+\varepsilon)^{-2}\,dt + \int_{t=0}^{\infty} I_{\{R(t)\leqslant\delta-\beta\}}(R(t)+\beta+\varepsilon)^{-2}\,dt.\end{aligned}$$

By the first Ray–Knight theorem ([8], Theorem 2.2 of Chapt. XI),

$$\int_{t=0}^{T_\beta} I_{\{B(t)\leqslant\delta\}}(B(t)+\varepsilon)^{-2} \overset{\mathscr{D}}{=} \int_{y=0}^{\delta-\beta} Z_2(y)(y+\beta+\varepsilon)^{-2}\,dy, \tag{45}$$

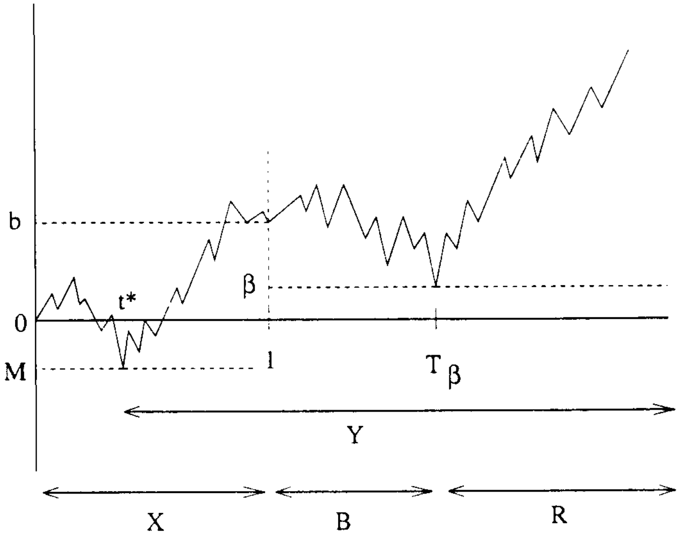

FIG. 3. Path decomposition.

where Z_2 is a square of a 2-dimensional Bessel process starting from 0 restricted to $[0, \delta - \beta]$. Similarly, since the local time of R is also Z_2,

$$\int_{t=0}^{\infty} I_{\{R(t) \leqslant \delta - \beta\}}(R(t) + \beta + \varepsilon)^{-2}\, dt \overset{\mathscr{D}}{=} \int_{y=0}^{\delta - \beta} Z_2(y)(y + \beta + \varepsilon)^{-2}\, dy. \tag{46}$$

Now use Lemma 5.1 to conclude that both of these have expectations that are $o(\log(1/\varepsilon))$ as $\varepsilon \to 0$. ∎

Next we compare the integrals in (34) and (33).

THEOREM 5.2. *As* $k \to \infty$,

$$\frac{\int_{t=0}^{1} (L_{m_k}(t) - M_{m_k} + z_{m_k})^{-2}\, dt}{\int_{t=0}^{1} (X(t) - M + z_{m_k})^{-2}\, dt} \xrightarrow{P} 1. \tag{47}$$

Proof. We will break the proof into two parts. First we will show that replacing M_{m_k} by M in the numerator has a negligible effect, and then we will show that replacing the linear interpolator L_{m_k} of X by X also has a negligible effect. To that end, it will be convenient to write the ratio in (47) as the product

$$\frac{\int_{t=0}^{1} (L_{m_k}(t) - M_{m_k} + z_{m_k})^{-2}\, dt}{\int_{t=0}^{1} (X(t) - M + z_{m_k})^{-2}\, dt}$$

$$= \frac{\int_{t=0}^{1} (L_{m_k}(t) - M + z_{m_k})^{-2}\, dt}{\int_{t=0}^{1} (X(t) - M + z_{m_k})^{-2}\, dt} \frac{\int_{t=0}^{1} (L_{m_k}(t) - Mm_k + z_{m_k})^{-2}\, dt}{\int_{t=0}^{1} (L_{m_k}(t) - M + z_{m_k})^{-2}\, dt}. \tag{48}$$

The proof consists of showing that each of the two fractions on the righthand side of (48) converges in probability to 1.

We begin by showing that as $k \to \infty$,

$$\frac{\int_{t=0}^{1} (X(t) - M + z_{m_k})^{-2}\, dt}{\int_{t=0}^{1} (L_{m_k}(t) - M + z_{m_k})^{-2}\, dt} \xrightarrow{P} 1 \tag{49}$$

(this is the reciprocal of the first factor in (48)). Using

$$\frac{1}{(X(t) - M + z_{m_k})^2} = \frac{1}{(L_{m_k}(t) - M + z_{m_k})^2} \frac{1}{\left(1 + \dfrac{X(t) - L_{m_k}(t)}{L_{m_k}(t) - M + z_{m_k}}\right)^2},$$

we can rewrite (49) as

$$\frac{\displaystyle\int_{t=0}^{1} (X(t) - M + z_{m_k})^{-2}\, dt}{\displaystyle\int_{t=0}^{1} (L_{m_k}(t) - M + z_{m_k})^{-2}\, dt}$$

$$= \frac{\displaystyle\sum_{i=1}^{m_k} \int_{s=t_{i}^{mk}-1}^{t_i^{mk}} (L_{m_k}(s) - M + z_{m_k})^{-2} \left(1 + \frac{X(s) - L_{m_k}(s)}{L_{m_k}(s) - M + z_{m_k}}\right)^{-2} ds}{\displaystyle\sum_{i=1}^{m_k} \int_{s=t_{i-1}^{mk}}^{t_i^{mk}} (L_{m_k}(s) - M + z_{m_k})^{-2}\, ds}$$

$$= \frac{\displaystyle\sum_{i=1}^{m_k} \int_{s=t_{i-1}^{mk}}^{t_i^{mk}} (L_{m_k}(s) - M + z_{m_k})^{-2} \left(1 + \frac{X(s_i^{m_k}) - L_{m_k}(s_i^{m_k})}{L_{m_k}(s_i^{m_k}) - M + z_{m_k}}\right)^{-2} ds}{\displaystyle\sum_{i=1}^{m_k} \int_{s=t_{i-1}^{mk}}^{t_i^{mk}} (L_{m_k}(s) - M + z_{m_k})^{-2}\, ds}$$

$$= \sum_{i=1}^{m_k} \lambda_i^{m_k} \left(1 + \frac{X(s_i^{m_k}) - L_{m_k}(s_i^{m_k})}{L_{m_k}(s_i^{m_k}) - M + z_{m_k}}\right)^{-2},$$

where the $\lambda_i^{m_k}$'s are non-negative and sum to 1, and each $s_i^{m_k} \in [t_{i-1}^{m_k}, t_i^{m_k}]$ (we used the mean-value theorem for integrals in the second equality). We will show that for each i,

$$\frac{X(s_i^{m_k}) - L_{m_k}(s_i^{m_k})}{L_{m_k}(s_i^{m_k}) - M + z_{m_k}} \xrightarrow{P} 0,$$

which will complete the proof of (49). Observe that

$$\begin{aligned}
&\left| \frac{X(s_i^{m_k}) - L_{m_k}(s_i^{m_k})}{L_{m_k}(s_i^{m_k}) - M + z_{m_k}} \right| \\
&\qquad \leqslant \max_{t_{i-1}^{mk} \leqslant s \leqslant t_i^{mk}} \frac{|X(s) - L_{m_k}(s)|}{L_{m_k}(s) - M + z_{m_k}} \\
&\qquad \leqslant \left(\max_{t_{i-1}^{mk} \leqslant s \leqslant t_i^{mk}} \frac{|X(s) - L_{m_k}(s)|}{\sqrt{t_i^{m_k} - t_{i-1}^{m_k}}} \right) \frac{\sqrt{t_i^{m_k} - t_{i-1}^{m_k}}}{X(t_{i-1}^{m_k}) \wedge X(t_{i-1}^{m_k}) - M + z_{m_k}}. \qquad (50)
\end{aligned}$$

The factor

$$\max_{t_{i-1}^{mk} \leqslant s \leqslant t_i^{mk}} \frac{|X(s) - L_{m_k}(s)|}{\sqrt{t_i^{m_k} - t_{i-1}^{m_k}}}$$

is the maximum of the absolute value of a standard Brownian bridge. The last term on the right-hand side of (50) converges in probability to 0, since

$$\begin{aligned}
&\frac{t_i^{m_k} - t_{i-1}^{m_k}}{(X(t_{i-1}^{m_k}) \wedge X(t_i^{m_k}) - M + z_{m_k})^2} \\
&\quad = \frac{t_i^{m_k} - t_{i-1}^{m_k}}{X(t_{i-1}^{m_k}) - M + z_{m_k})(X(t_i^{m_k}) - M + z_{m_k})} \frac{X(t_{i-1}^{m_k}) \vee X(t_i^{m_k}) - M + z_{m_k}}{X(t_{i-1}^{m_k}) \wedge X(t_i^{m_k}) - M + z_{m_k}} \\
&\quad = \frac{t_i^{m_k} - t_{i-1}^{m_k}}{(X(t_{i-1}^{m_k}) - M + z_{m_k})(X(t_i^{m_k}) - M + z_{m_k})} \\
&\qquad \cdot \left(1 + \frac{\sqrt{t_i^{m_k} - t_{i-1}^{m_k}}}{X(t_{i-1}^{m_k}) \wedge X(t_{i-1}^{m_k}) - M + z_{m_k}} \frac{|X(t_i^{m_k}) - X(t_{i-1}^{m_k})|}{\sqrt{t_i^{m_k} - t_{i-1}^{m_k}}} \right).
\end{aligned}$$

The last factor

$$\frac{|X(t_i^{m_k}) - X(t_{i-1}^{m_k})|}{\sqrt{t_i^{m_k} - t_{i-1}^{m_k}}}$$

has the distribution of the absolute value of a standard normal random variable. Therefore, since

$$\rho^{m_k} \geqslant \frac{t_i^{m_k} - t_{i-1}^{m_k}}{(X(t_{i-1}^{m_k}) - M + z_{m_k})(X(t_i^{m_k}) - M + z_{m_k})} \to 0,$$

it follows that

$$\frac{\sqrt{t_i^{m_k} - t_{i-1}^{m_k}}}{X(t_{i-1}^{m_k}) \wedge X(t_{i-1}^{m_k}) - M + z_{m_k}} \xrightarrow{P} 0.$$

Thus

$$\left| \frac{X(s_i^{m_k}) - L_{m_k}(s_i^{m_k})}{L_{m_k}(s_i^{m_k}) - M + z_{m_k}} \right| \xrightarrow{P} 0,$$

which establishes (49).

To complete the proof we need to show that the second factor on the right-hand side of (48) converges in probability to 1; i.e.,

$$\frac{\int_{t=0}^{1} (L_{m_k}(t) - M_{m_k} + z_{m_k})^{-2}\, dt}{\int_{t=0}^{1} (L_{m_k}(t) - M + z_{m_k})^{-2}\, dt} \xrightarrow{P} 1. \tag{51}$$

The ratio of the integrals is always at least 1, so we only need prove the upper bound. But this follows from the fact that $(M_{m_k} - M)/z_{m_k} \xrightarrow{P} 0$ by Lemma 4.3. More precisely, we can bound the integrand in the numerator by

$$\begin{aligned}
&(L_{m_k}(t) - M_{m_k} + z_{m_k})^{-2} \\
&\quad = (L_{m_k}(t) - M + z_{m_k})^{-2} \left(1 - \frac{M_{m_k} - M}{L_{m_k}(t) - M + z_{m_k}}\right)^{-2} \\
&\quad \leqslant (L_{m_k}(t) - M + z_{m_k})^{-2} \left(1 - \frac{M_{m_k} - M}{z_{m_k}}\right)^{-2},
\end{aligned}$$

and so

$$1 \leqslant \frac{\int_{t=0}^{1} (L_{m_k}(t) - M_{m_k} + z_{m_k})^{-2}\, dt}{\int_{t=0}^{1} (L_{m_k}(t) - M + z_{m_k})^{-2}\, dt} \leqslant \left(1 - \frac{M_{m_k} - M}{z_{m_k}}\right)^{-2} \xrightarrow{P} 1. \quad \blacksquare$$

Combining Theorems 5.1 and 5.2, we arrive at our main result for estimating the average of the ρ_i^n's.

COROLLARY 5.1. *As* $k \to \infty$,

$$\frac{\frac{1}{m_k}\sum_{i=1}^{m_k} \rho_i^{m_k}}{\frac{1}{m_k}\log(1/z_{m_k})} \xrightarrow{P} 4. \tag{52}$$

Proof. This follows directly from Theorem 5.2 and Theorem 5.1, since $z_{m_k} \to 0$ and

$$\frac{\frac{1}{m_k}\sum_{i=1}^{m_k} \rho_i^{m_k}}{\frac{1}{m_k}\log(1/z_{m_k})}$$

$$= \frac{\frac{1}{m_k}\int_{t=0}^{1} (L_{m_k}(t) - M_{m_k} + z_{m_k})^{-2}\, dt}{\frac{1}{m_k}\log(1/z_{m_k})}$$

$$= \frac{\int_{t=0}^{1} (L_{m_k}(t) - M_{m_k} + z_{m_k})^{-2}\, dt}{\int_{t=0}^{1} (X(t) - M + z_{m_k})^{-2}\, dt} \frac{\frac{1}{m_k}\int_{t=0}^{1} (X(t) - M + z_{m_k})^{-2}\, dt}{\frac{1}{m_k}\log(1/z_{m_k})}.$$

Theorem 5.2 implies that the first ratio on the right converges in probability to 1, and by Theorem 5.1,

$$\frac{\frac{1}{m_k}\int_{t=0}^{1} (X(t) - M + z_{m_k})^{-2}\, dt}{\frac{1}{m_k}\log(1/z_{m_k})}$$

converges in expectation (and therefore in probability) to 4. ∎

Corollary 5.1 gives sufficient information on the rate of convergence of the average of the ρ_i^n's. We will also need bounds for the maximum ρ^n, for which the following pair of technical lemmas will be useful.

LEMMA 5.2. *As* $k \to \infty$,

$$\min(\gamma_{m_k}^2 \rho^{m_k}, 1) \xrightarrow{P} 1 \tag{53}$$

and

$$\max(\gamma_{m_k}^2 \rho^{m_k}, 2) \xrightarrow{P} 2. \tag{54}$$

Proof. Observe the following inequalities:

$$\begin{aligned}
\frac{1}{\gamma_{m_k}^2} \leqslant \frac{\tau_{m_k}}{z_{m_k}^2} \leqslant \frac{T_{m_k}}{z_{m_k}^2} \\
&= \frac{T_{m_k}}{(X(t_L^{m_k}) - M + z_{m_k})(X(t_R^{m_k}) - M + z_{m_k})} \\
&\quad \cdot \frac{(X(t_L^{m_k}) - M + z_{m_k})(X(t_R^{m_k}) - M + z_{m_k})}{z_{m_k}^2} \\
&\leqslant \rho^{m_k}\left(1 + \frac{X(t_L^{m_k}) - M}{z_{m_k}}\right)\left(1 + \frac{X(t_R^{m_k}) - M}{z_{m_k}}\right).
\end{aligned}$$

Therefore,

$$\frac{1}{\gamma_{m_k}^2 \rho^{m_k}} \leqslant \left(1 + \frac{X(t_L^{m_k}) - M}{z_{m_k}}\right)\left(1 + \frac{X(t_R^{m_k}) - M}{z_{m_k}}\right) \xrightarrow{P} 1$$

by Lemma 4.3, which establishes (53).

For the second statement (54),

$$\begin{aligned}
\rho^{m_k} &= \frac{S_{m_k}}{(X(t_L^{m_k}) - M_{m_k} + z_{m_k})(X(t_R^{m_k}) - M_{m_k} + z_{m_k})} \leqslant \frac{S_{m_k}}{z_{m_k}^2} \\
&\leqslant \frac{S_{m_k}}{z_{m_k+1}^2} = \frac{S_{m_k}}{\tau_{m_k+1}} \frac{1}{\gamma_{m_k+1}^2} \leqslant \frac{S_{m_k}}{\tau_{m_k+1}} \frac{1}{\gamma_{m_k}^2}.
\end{aligned}$$

Therefore,

$$\frac{1}{\gamma_{m_k}^2 \rho^{m_k}} \geqslant \frac{\tau_{m_k+1}}{S_{m_k}} \xrightarrow{P} \frac{1}{2}$$

by Lemma 4.1. ∎

We now introduce the sequence of stopping times $\{n_k\}$ mentioned in the Introduction. Let n_k be the kth time that the sequence ρ^{m_j} reaches a new minimum. By the proof of Theorem 3.1, $\rho^{n_k} \downarrow 0$. We will be mainly interested in the error at the times $\{n_k\}$.

For large n, the average of the $\{\rho_i^n\}$ is not much less than the maximum ρ^n. This is because for the intervals that are not split, ρ_i^n does not change much, and the split interval replaces ρ^n with two values that are each about half as large. Therefore, the ρ_i^n tend to be spread out between $\rho^n/2$ and ρ^n. The following lemma gives a sufficiently accurate bound for our purposes.

LEMMA 5.3. *As $k \to \infty$,*

$$P\left(\frac{\frac{1}{n_k}\sum_{i=1}^{n_k} \rho_i^{n_k}}{\rho^{n_k}} \geqslant \frac{1}{4}\right) \to 1. \tag{55}$$

Proof. For simplicity we focus on one term and omit some subscripts. Suppose that $\rho_i^{n_k}$ is descended from ρ^{n-1} for some $n = n(i) < n_k$, and no intervening splits; say that ρ^{n-1} is split, resulting in $\hat{\rho}_i^n$, which eventually becomes $\rho_i^{n_k}$. (We will suppress the dependence of $n = n(i)$ on i.) Let j be such that $m_{j-1} \leqslant n < m_j \leqslant n_k$. Since the observations are dense, $m_j \to \infty$. Also, $\rho^{n-1} \geqslant \rho^{m_j}$, since the $\{z_i\}$ are constant between the times $\{m_j\}$ and the ρ's decrease due to decreases in M_n. Tracing the sequence of events, we start with

$$\rho^{n-1} \triangleq \frac{T}{(x_1 - M_{n-1} + z_{n-1})(x_2 - M_{n-1} + z_{n-1})}. \tag{56}$$

Let us suppose that the left child of the split results in $\hat{\rho}_i^n$ (the calculation for the right child is analogous), so that

$$\hat{\rho}_i^n = \frac{\theta T}{(x_1 - M_n + z_n)(X_3 - M_n + z_n)}, \tag{57}$$

where

$$\theta = \frac{x_1 - M_{n-1} + z_{n-1}}{x_1 - M_{n-1} + z_{n-1} + x_2 - M_{n-1} + z_{n-1}}$$

and X_3, the value of the function at the new evaluation point, can be expressed as

$$X_3 = (1-\theta)\, x_1 + \theta x_2 + \sqrt{\theta(1-\theta)\, T}\, V_i,$$

where $V_i \sim N(0, 1)$. We begin by showing that

$$\liminf P\left(\frac{\hat{\rho}_i^n}{\rho^{n-1}} \geqslant \frac{1}{2}\right) \geqslant \frac{1}{2}.$$

Using (56) and (57), we can express

$$\begin{aligned}\frac{\hat{\rho}_i^n}{\rho^{n-1}} &= \frac{\theta(x_1 - M_{n-1} + z_{n-1})(x_2 - M_{n-1} + z_{n-1})}{(x_1 - M_n + z_n)([(1-\theta)\, x_1 + \theta x_2] + \sqrt{\theta(1-\theta)\, T}\, V_i - M_n + z_n)} \\ &\geqslant \frac{\theta(x_1 - M_{n-1} + z_n)(x_2 - M_{n-1} + z_n)}{(x_1 - M_n + z_n)([(1-\theta)\, x_1 + \theta x_2] + \sqrt{\theta(1-\theta)\, T}\, V_i - M_n + z_n)},\end{aligned}$$

where the last inequality follows from the fact that the $\{z_n\}$ are nonincreasing. Therefore, on the event that $V_i \leqslant 0$ (which has probability 1/2),

$$\begin{aligned}\frac{\hat{\rho}_i^n}{\rho^{n-1}} &\geqslant \frac{(x_1 - M_{n-1} + z_n)(x_2 - M_{n-1} + z_n)}{(x_1 - M_n + z_n)\left(\dfrac{1-\theta}{\theta}(x_1 - M_n + z_n) + x_2 - M_n + z_n\right)} \\ &= \frac{(x_1 - M_{n-1} + z_n)(x_2 - M_{n-1} + z_n)}{(x_1 - M_n + z_n)\left(\dfrac{x_2 - M_{n-1} + z_n}{x_1 - M_{n-1} + z_n}(x_1 - M_n + z_n) + x_2 - M_n + z_n\right)} \\ &= \frac{(x_1 - M_{n-1} + z_n)(x_2 - M_{n-1} + z_n)}{(x_1 - M_n + z_n)(x_2 - M_n + z_n)\left(\dfrac{x_2 - M_{n-1} + z_n}{x_2 - M_n + z_n}\dfrac{x_1 - M_n + z_n}{x_1 - M_{n-1} + z_n} + 1\right)}.\end{aligned} \tag{58}$$

Notice that

$$\frac{z_{m_{j-1}}}{z_{m_j}} = \frac{\gamma_{m_{j-1}}\sqrt{\tau_{m_{j-1}}}}{\gamma_{m_j}\sqrt{\tau_{m_j}}} \leqslant \frac{\sqrt{\tau_{m_{j-1}}}}{\sqrt{\tau_{m_j}}} \leqslant \frac{\sqrt{S_{m_{j-1}}}}{\sqrt{\tau_{m_j}}} \xrightarrow{P} \sqrt{2}, \tag{59}$$

and so $z_{m_{j-1}}/z_{m_j}$ is bounded in probability.

Let us consider terms of the form

$$\frac{x_1 - M_{n-1} + z_n}{x_1 - M_n + z_n} = \left(1 + \frac{M_{n-1} - M_n}{x_1 - M_{n-1} + z_{n-1}}\right)^{-1}. \tag{60}$$

By Lemma 4.3 and (59),

$$0 \leqslant \frac{M_{n-1} - M_n}{x_1 - M_{n-1} + z_{n-1}} \leqslant \frac{M_{n-1} - M}{z_{n-1}} \leqslant \frac{M_{m_{j-1}} - M}{z_{m_{j-1}}} \frac{z_{m_{j-1}}}{z_{m_j}} \xrightarrow{P} 0.$$

Therefore,

$$\frac{x_1 - M_{n-1} + z_{n-1}}{x_1 - M_n + z_n} \xrightarrow{P} 1.$$

Applying this (4 times) to (58), we obtain

$$\frac{\hat{\rho}_i^n}{\rho^{n-1}} \geqslant \frac{1}{2} 1_{\{V_i < 0\}} + o_P(1),$$

and so

$$\liminf P\left(\frac{\hat{\rho}_i^n}{\rho^{n-1}} \geqslant \frac{1}{2}\right) \geqslant \frac{1}{2}. \tag{61}$$

Subsequently, $\rho_i^{n_k}$ is descended from $\hat{\rho}_i^n$ with no splits, so from (57),

$$\rho_i^{n_k} = \frac{\theta T}{(x_1 - M_{n_k} - z_{n_k})(X_3 - M_{n_k} + z_{n_k})},$$

and so

$$\frac{\rho_i^{n_k}}{\hat{\rho}_i^n} = \frac{(x_1 - M_n + z_n)(X_3 - M_n + z_n)}{(x_1 - M_{n_k} + z_{n_k})(X_3 - M_{n_k} + z_{n_k})}. \tag{62}$$

Observe that

$$\begin{aligned}\frac{x_1-M_n+z_n}{x_1-M_{n_k}+z_{n_k}} &= \frac{x_1-M-(M_n-M)+z_n}{x_1-M-(M_{n_k}-M)+z_{n_k}}\\ &\geqslant \frac{x_1-M-(M_{m_{j-1}}-M)+z_{m_j}}{x_1-M-(M_{n_k}-M)+z_{n_k}}\\ &= \frac{x_1-M+z_{m_j}(1-o_P(1))}{x_1-M+z_{n_k}(1-o_P(1))}\\ &\geqslant \frac{x_1-M+z_{n_k}(1-o_P(1))}{x_1-M+z_{n_k}(1-o_P(1))}\\ &= 1+o_P(1).\end{aligned}$$

Applying this estimate twice to (62) establishes that

$$\frac{\rho_i^{n_k}}{\hat{\rho}_i^n} \geqslant 1+o_P(1). \tag{63}$$

Therefore,

$$\begin{aligned}\frac{\frac{1}{n_k}\sum_{i=1}^{n_k}\rho_i^{n_k}}{\rho^{n_k}} &= \frac{\frac{1}{n_k}\sum_{i=1}^{n_k}\frac{\hat{\rho}_i^n}{\rho^{n-1}}\frac{\rho_i^{n_k}}{\hat{\rho}_i^n}\rho^{n-1}}{\rho^{n_k}}\\ &\geqslant \frac{\frac{1}{n_k}\sum_{i=1}^{n_k}\frac{\hat{\rho}_i^n}{\rho^{n-1}}\frac{\rho_i^{n_k}}{\hat{\rho}_i^n}\rho^{m_j}}{\rho^{n_k}}\\ &\geqslant \frac{1}{n_k}\sum_{i=1}^{n_k}\frac{1_{\{V_i<0\}}}{2}\frac{\rho_i^{n_k}}{\hat{\rho}_i^n}\frac{\rho^{m_j}}{\rho^{n_k}}\\ &\geqslant \frac{1}{n_k}\sum_{i=1}^{n_k}\frac{1_{\{V_i<0\}}}{2}(1+o_P(1))\end{aligned}$$

by (63) and the fact that $\rho^{m_j}>\rho^{n_k}$ by construction of the stopping times $\{n_k\}$.

At time n_k there are two types of subinterval; pairs that are split from the same parent, and subintervals whose siblings have been split previously. Let I_{n_k} denote the indices of the first subinterval of pairs of the first type. For each $i\in I_{n_k}$ there is one random variable V_i that determines the size of the pair, while for $j\in I_{n_k}^c$, there is a single random variable V_j. Therefore, the last sum can be written

$$\frac{1}{n_k}\left(\sum_{i\in I_{n_k}} 2B_i(1+o_P(1)) + \sum_{j\in I_{n_k}^c} B_j(1+o_P(1))\right),$$

where the B_i are independent Bernoulli(1/2), and this last expression is bounded below in probability by 1/4. ∎

6. ASYMPTOTIC ERROR ANALYSIS

We are now prepared to describe asymptotic limits on the length of the shortest interval, and then on the error $\hat{\Delta}_n$ and Δ_n, using the results of the previous sections. We begin with asymptotic probabilistic bounds on the length of the smallest subintervals.

THEOREM 6.1. *For $\varepsilon > 0$,*

$$\tau_{n_k} \exp\left(\frac{1}{8}\frac{n_k}{\gamma_{n_k}^2}(1-\varepsilon)\right) \xrightarrow{P} 0, \tag{64}$$

and

$$\tau_{n_k} \exp\left(\frac{n_k}{\gamma_{n_k}^2}(1+\varepsilon)\right) \xrightarrow{P} +\infty. \tag{65}$$

Proof. To prove (64), first observe that

$$\begin{aligned}
\frac{\gamma_{n_k}^2}{n_k}\log(1/z_{n_k}) &\geqslant \frac{\min(\gamma_{n_k}^2\rho^{n_k}, 1)}{n_k}\frac{\log(1/z_{n_k})}{\rho^{n_k}}\\
&= \min(\gamma_{n_k}^2\rho^{n_k}, 1)\frac{\frac{1}{n_k}\log(1/z_{n_k})}{\frac{1}{n_k}\sum_{i=1}^{n_k}\rho_i^{n_k}}\frac{\frac{1}{n_k}\sum_{i=1}^{n_k}\rho_i^{n_k}}{\rho^{n_k}}\\
&\geqslant \min(\gamma_{n_k}^2\rho^{n_k}, 1)\frac{\frac{1}{n_k}\log(1/z_{n_k})}{\frac{1}{n_k}\sum_{i=1}^{n_k}\rho_i^{n_k}}\min\left(\frac{\frac{1}{n_k}\sum_{i=1}^{n_k}\rho_i^{n_k}}{\rho^{n_k}}, \frac{1}{4}\right)\\
&\xrightarrow{P} \frac{1}{16},
\end{aligned} \tag{66}$$

since the first term in (66) converges in probability to 1 by Lemma 5.2, the second term converges in probability to 1/4 by Corollary 5.1, and the last term converges in probability to 1/4 by Lemma 5.3. Therefore,

$$P\left(\frac{\gamma_{n_k}^2}{n_k}\log(1/z_{n_k})<\frac{1}{16}-\frac{\varepsilon}{32}\right)$$
$$=P\left(-\frac{\gamma_{n_k}^2}{2n_k}\log(\tau_{n_k})-\frac{\gamma_{n_k}^2}{n_k}\log(\gamma_{n_k})<\frac{1}{16}-\frac{\varepsilon}{32}\right)\to 0. \tag{67}$$

Since

$$\frac{\gamma_{n_k}^2}{n_k}\log(\gamma_{n_k})\to 0 \tag{68}$$

by our assumption (5), we conclude that

$$P\left(-\frac{\gamma_{n_k}^2}{2n_k}\log(\tau_{n_k})<\frac{1}{16}-\frac{\varepsilon}{32}\right)\to 0. \tag{69}$$

Now

$$-\frac{\gamma_{n_k}^2}{2n_k}\log(\tau_{n_k})<\frac{1}{16}-\frac{\varepsilon}{32}$$

if and only if

$$\tau_{n_k}\exp\left(\frac{n_k}{8\gamma_{n_k}^2}(1-\varepsilon)\right)>\exp\left(-\frac{n_k}{\gamma_{n_k}^2}\frac{\varepsilon}{16}\right)$$

so (69) implies that

$$\lim P\left(\tau_{n_k}\exp\left(\frac{n_k}{8\gamma_{n_k}^2}(1-\varepsilon)\right)>\exp\left(-\frac{n_k}{\gamma_{n_k}^2}\frac{\varepsilon}{16}\right)\right)=0. \tag{70}$$

This proves (64), since $n_k/\gamma_{n_k}^2\to\infty$.

To prove (65), observe that

$$\frac{\gamma_{n_k}^2}{n_k}\log(1/z_{n_k}) \leqslant \frac{\max(\gamma_{n_k}^2\rho^{n_k}, 2)}{n_k}\,\frac{\log(1/z_{n_k})}{\rho^{n_k}}$$
$$\leqslant \frac{\max(\gamma_{n_k}^2\rho^{n_k}, 2)}{n_k}\,\frac{\log(1/z_{n_k})}{\frac{1}{n_k}\sum_{i=1}^{n_k}\rho_i^{n_k}}$$
$$= \max(\gamma_{n_k}^2\rho^{n_k}, 2)\,\frac{\frac{1}{n_k}\log(1/z_{n_k})}{\frac{1}{n_k}\sum_{i=1}^{n_k}\rho_i^{n_k}} \xrightarrow{P} \frac{1}{2}$$

by Lemma 5.2 and Corollary 5.1. Therefore,

$$P\left(\frac{\gamma_{n_k}^2}{n_k}\log(1/z_{n_k}) > \frac{1}{2}+\frac{\varepsilon}{2}\right)$$
$$= P\left(-\frac{\gamma_{n_k}^2}{2n_k}\log(\tau_{n_k}) - \frac{\gamma_{n_k}^2}{n_k}\log(\gamma_{n_k}) > \frac{1}{2}+\frac{\varepsilon}{2}\right) \to 0,$$

which, again using (68), implies that

$$P\left(-\frac{\gamma_{n_k}^2}{2n_k}\log(\tau_{n_k}) > \frac{1}{2}+\frac{\varepsilon}{2}\right) \to 0. \tag{71}$$

Since

$$-\frac{\gamma_{n_k}^2}{2n_k}\log(\tau_{n_k}) > \frac{1}{2}+\frac{\varepsilon}{2}$$

if and only if

$$\tau_{n_k}\exp\left(\frac{n_k}{\gamma_{n_k}^2}(1+\varepsilon)\right) < \exp\left(\frac{n_k}{2\gamma_{n_k}^2}\,\varepsilon\right),$$

(71) implies that

$$\lim P\left(\tau_{n_k}\exp\left(\frac{n_k}{\gamma_{n_k}^2}(1+\varepsilon)\right) < \exp\left(\frac{n_k}{2\gamma_{n_k}^2}\,\varepsilon\right)\right) = 0. \tag{72}$$

This proves (65), since $n_k/\gamma_{n_k}^2 \to \infty$. ∎

We are now ready to say something about the local error $\hat{\Delta}_{n_k}$.

PROPOSITION 6.1. *For any* $\varepsilon > 0$,

$$\exp\left(\frac{n_k}{2\gamma_{n_k}^2}(1+\varepsilon)\right)\hat{\Delta}_{n_k} \xrightarrow{P} +\infty, \tag{73}$$

and

$$\exp\left(\frac{n_k}{16\gamma_{n_k}^2}(1-\varepsilon)\right)\hat{\Delta}_{n_k} \xrightarrow{P} 0. \tag{74}$$

Proof. To prove (73),

$$\begin{aligned}\exp\left(\frac{n_k}{2\gamma_{n_k}^2}(1+\varepsilon)\right)\hat{\Delta}_{n_k} &= \left(\exp\left(\frac{n_k}{\gamma_{n_k}^2}(1+\varepsilon)\right)\tau_{n_k}\right)^{1/2}\frac{\sqrt{T_{n_k}}}{\sqrt{\tau_{n_k}}}\frac{\hat{\Delta}_{n_k}}{\sqrt{T_{n_k}}} \\ &\geqslant \left(\exp\left(\frac{n_k}{\gamma_{n_k}^2}(1+\varepsilon)\right)\tau_{n_k}\right)^{1/2}\frac{\hat{\Delta}_{n_k}}{\sqrt{T_{n_k}}}.\end{aligned}$$

The first term converges in probability to ∞ by (65) and the last term converges in distribution to $\min\{R(U), R'(1-U)\}$ by Corollary 4.1. Since $P(\min\{R(U), R'(1-U)\} > 0) = 1$, the product converges in probability to ∞.

To prove (74),

$$\exp\left(\frac{n_k}{16\gamma_{n_k}^2}(1+\varepsilon)\right)\hat{\Delta}_{n_k} = \left(\exp\left(\frac{n_k}{8\gamma_{n_k}^2}(1+\varepsilon)\right)\tau_{n_k}\right)^{1/2}\frac{\sqrt{T_{n_k}}}{\sqrt{\tau_{n_k}}}\frac{\hat{\Delta}_{n_k}}{\sqrt{T_{n_k}}}.$$

The first term converges to 0 in probability by (64), and the second term is bounded in probability by Lemma 4.2. The last term converges in distribution to $\min\{R(U), R'(1-U)\}$ by Corollary 4.1. Thus the product converges to 0 in probability. ∎

Recall that $0 \leqslant \Delta_{n_k} \leqslant \hat{\Delta}_{n_k}$. Theorem 6.2 therefore gives an upper bound on how fast we can approximate the minimum. We next turn our attention to obtaining a lower bound.

Define

$$U_R^n = \frac{\min_{s \geqslant t_R^n} X(s) - M}{X(t_R^n) - M}, \qquad U_L^n = \frac{\min_{s \geqslant t_L^n} X(s) - M}{X(t_L^n) - M}.$$

Then

$$\Delta_n \geqslant \min\{U_L^n(X(t_L^n)-M),\ U_R^n(X(t_R^n)-M)\} \geqslant \min\{U_L^n, U_R^n\}\ \hat{\Delta}_n. \tag{75}$$

We will show that U_R^n converges to a uniform random variable on $(0, 1)$, independent of $X(t_R^n)-M$ (a similar result holds, of course, for U_L^n).

PROPOSITION 6.2. *For* $z \in [0, 1]$,

$$P(U_R^n \leqslant z) \to z.$$

Proof. If X is a 3-dimensional Bessel bridge from $a>0$ at time 0 to $b>0$ at time T, and $v=\min_{0\leqslant t\leqslant T} X(t)$, then for $y<a \wedge b$,

$$P(v>y)=\frac{1-\exp\left(-\frac{2}{T}(a-y)(b-y)\right)}{1-\exp\left(-\frac{2}{T}ab\right)}. \tag{76}$$

This formula can be derived from formulas for the distribution of the minimum of diffusion processes developed in [4]. Let $0<z<1$, and let $P_{t,m,x}$ be a regular conditional probability for X given $t^*=t$, $M=m$, and $X(1)=x$, and let $E_{t,m,x}$ be the corresponding expectation. Under $P_{t,m,x}$ and conditional on $X(t_R^n)$, $\{X(t_R^n+s)-M : 0\leqslant s\leqslant 1-t_R^n\}$ is a 3-dimensional Bessel bridge from $X(t_R^n)$ at time 0 to x at time $1-t_R^n$, so we can apply (76) to obtain

$$\begin{aligned}
P_{t,m,x}(U_R^n>z) &= E_{t,m,x} P_{t,m,x}(U_R^n>z \mid X(t_R^n)) \\
&= E_{t,m,x}\left(\frac{1-\exp\left(-\frac{2}{1-t_R^n}(X(t_R^n)-M)(1-z)(X(1)-zX(t_R^n))\right)}{1-\exp\left(-\frac{2}{1-t_R^n}(X(t_R^n)-M)(X(1)-M)\right)}\right) \to 1-z
\end{aligned}$$

by the dominated convergence theorem, since $t_R^n \downarrow t^*$, $X(t_R^n)\to M$ by continuity, and $(1-e^{-\varepsilon a})/(1-e^{-\varepsilon}) \to a$ as $\varepsilon \to 0$. ∎

Putting all this together, we summarize the results of this section in our main theorem.

THEOREM 6.2. *For any* $\varepsilon>0$,

$$\exp\left(\frac{n_k}{2\gamma_{n_k}^2}(1+\varepsilon)\right)\Delta_{n_k} \xrightarrow{P} +\infty, \tag{77}$$

and

$$\exp\left(\frac{n_k}{16\gamma_{n_k}^2}(1-\varepsilon)\right)\Delta_{n_k}\xrightarrow{P}0. \tag{78}$$

Proof. By (73) and (75),

$$\exp\left(\frac{n_k}{2\gamma_{n_k}^2}(1+\varepsilon)\right)\Delta_{n_k}\geqslant\exp\left(\frac{n_k}{2\gamma_{n_k}^2}(1+\varepsilon)\right)\hat{\Delta}_{n_k}\min\{U_L^n,U_R^n\}.$$

Since $\min\{U_L^n, U_R^n\}$ converges in distribution to the minimum of two independent Uniform(0, 1) random variables (which is almost surely positive), the same logic used in the proof of (73) applies to prove (77).

Since $\Delta_n \leqslant \hat{\Delta}_n$, (78) is a direct consequence of (74). ∎

ACKNOWLEDGMENTS

The comments of two anonymous referees, and also M. Horn and K. Ritter, substantially improved the paper.

REFERENCES

1. P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1968.
2. J. Calvin, Average performance of a class of adaptive algorithms for global optimization, *Ann. Appl. Probab.* **7** (1997), 711–730.
3. J. Calvin, Convergence rate of the P-algorithm for optimization of continuous functions, *in* "Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems" (P. M. Pardalos, Ed.), Kluwer Academic, Dordrecht/Norwell, MA, 1999.
4. E. Csáki, A. Földes, and P. Salminen, On the joint distribution of the maximum and its location for a linear diffusion, *Ann. Inst. H. Poincaré* **23** (1987), 179–194.
5. P. J. Fitzsimmons, Excursions above the minimum for diffusions, Unpublished manuscript, 1985.
6. J. Kiefer, Sequential minimax search for a maximum, *Proc. Amer. Math. Soc.* **4** (1953), 502–506.
7. H. Kushner, A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, *J. Basic Engrg.* **86** (1964), 97–106.
8. D. Revuz and M. Yor, "Continuous Martingales and Brownian Motion," Springer-Verlag, Berlin/New York, 1991.

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

10. A. Törn and A. Žilinskas, “Global Optimization,” Springer-Verlag, Berlin, 1989.

11. D. Williams, Path decomposition and continuity of local time for one-dimensional diffusions, I, *Proc. London Math. Soc. Ser. 3* **28** (1974), 738–768.

12. A. Žilinskas, Axiomatic characterization of global optimization algorithm and investigation of its search strategy, *Oper. Res. Lett.* **4** (1985), 35–39.