



Comparison of One-dimensional Composite and Non-composite Passive Algorithms

HISHAM A. AL-MHARMAH¹ and JAMES M. CALVIN^{2*}

¹*Industrial Engineering Department, University of Jordan, Amman 11942, Jordan;* ²*Department of Computer and Information Science, New Jersey Institute of Technology, Newark, NJ 07102, USA*
(e-mail: mharmah@nets.com.jo)

(Received 11 November 1997; accepted in revised form 20 January 1999)

Abstract. In this paper we analyze composite non-adaptive algorithms for optimization of one-dimensional Brownian motion. We show that a composite deterministic algorithm has a better average performance than the best random one.

Key words: Brownian motion, One-dimensional global optimization, Passive algorithms

1. Introduction

This paper is a study of the average performance of non-adaptive (also called passive) algorithms for approximating the global minimum of one-dimensional real-valued functions defined on the unit interval. More specifically, we study the average error of composite algorithms as compared to non-composite algorithms. Our criterion for error is the difference between the observed minimum value and the actual global minimum. We compare algorithms based on average performance on a random objective function. For our comparison we use the Wiener measure on the continuous functions, so the objective function is taken to be a sample function of a Brownian motion (or Wiener) process. Since this process is the intersection of the classes of Gaussian processes, Markov processes, and martingales, it is both of modeling interest and there is also a large set of tools available for its analysis.

We consider only non-adaptive algorithms; that is, algorithms that make no use of prior information in choosing the next observation site. A composite non-adaptive algorithm is one that maintains its form as the number of observations increases; see [15]. If we denote the set of observations made by an algorithm up to time n by $T_n = \{t_1, t_2, \dots, t_n\}$, then we will call an algorithm composite if $T_n \subset T_{n+1}$. For example, if the observation points $\{t_i\}$ are chosen independently according to a fixed probability distribution, then the resulting algorithm is composite (the set T_n is random, and for a given sample point ω , $T_n(\omega) \subset T_{n+1}(\omega)$). A consequence of compositeness is that there is no need to determine in advance how many observations are to be taken in order to construct the observation set.

* The second author was supported in part by NSF Grant DMI-9500173.

In contrast, non-composite algorithms do not adapt gracefully as the number of observations changes. An example is the “uniform grid” non-composite algorithm that takes equally spaced observations; if a total of n observations are to be made, they are placed at $1/n, 2/n, \dots, 1$. However, if the number of observations is increased to $n + 1$, there is no way to add an observation point so as to maintain a uniform grid.

In the case of Brownian motion, it is shown in [4] that if observations form a deterministic equi-spaced grid, then the error is about 82% as large as if the points are chosen at random uniformly over the unit interval. However, if new observations are to be added the uniformity of the deterministic grid will not hold at all times. One might expect, for example, that if the grid is such that $2k$ points are equi-spaced to the left of $1/2$ and k points are equi-spaced to the right of $1/2$, then choosing $3k$ points at random uniformly over the interval might give a smaller error on average. This question is the motivation for the present paper.

Al-Mharmah and Calvin [1] studied randomized non-adaptive algorithms, and found that the optimal distribution from which to draw the independent observations is Beta(2/3,2/3). Calvin [5] showed that the n quantiles of the Beta(2/3,2/3) distribution are optimal within the class of deterministic non-adaptive algorithms; this algorithm has the disadvantage of being non-composite. The average normalized error for the deterministic version is about 82% of that for the random algorithm (with the number of observations n predetermined). The advantage of the random algorithms is compositeness, and the disadvantage is the random gaps. The largest gap in a set of uniformly distributed points is of order $\log(n)/n$, and because of length-biased sampling, the large gaps are more likely to contain the minimizer. Our main result is to show that a composite deterministic algorithm has a better average performance than the optimal random algorithm. Since the studies mentioned above were concerned with the limiting distribution of the normalized error, by “optimal” we mean minimal limiting normalized mean error. This implies that we do not distinguish between the performance of algorithms that differ by mean error of order $o(n^{-1/2})$.

Our study relies on a “splitting” theorem for Brownian motion (Theorem 3.2 below), which splits the Brownian path at the global minimum into two conditionally independent path fragments. While this result generalizes to other one-dimensional diffusion processes, there is no analog for multi-dimensional random fields. Therefore, the techniques of this paper do not allow us to compare composite and non-composite algorithms for multi-dimensional optimization.

In the next section we introduce the problem and the notation. In Section 3 we establish some background results concerning the distribution of the minimizer and path decompositions at the global minimum for Brownian motion. In Section 4 we describe the algorithm and analyze the limiting error. The main proof is presented in an appendix.

2. Notation and terminology

Suppose that the class of functions to be optimized is $C[0, 1]$, the continuous functions defined on the unit interval. For $X \in C[0, 1]$, let $M = \min\{X(t); t \in [0, 1]\}$ denote its global minimum, and let $T = \inf\{t \leq 1 : X(t) = M\}$ be the (first) location where the minimum is attained. Throughout this paper we will assume that we are allowed to observe (without error) the function X at the sequence of observation points t_1, t_2, \dots in $[0, 1]$.

The Wiener measure will be taken as the probability distribution; i.e., X is taken to be a sample path of a Brownian motion process. The Wiener measure is characterized as follows. For each $t \in [0, 1]$, $X(t)$ has the normal distribution with mean 0 and variance t , and for any $k \geq 1$ and

$$0 \leq t_0 \leq t_1 \leq \dots \leq t_k \leq 1, \quad (1)$$

the increments $X(t_1) - X(t_0), X(t_2) - X(t_1), \dots, X(t_k) - X(t_{k-1})$ are independent. It follows that the random variables $X(t_i) - X(t_{i-1})$ are normally distributed with mean 0 and variance $t_i - t_{i-1}$ (see [3]).

As long as the maximum distance between observations goes to 0 as $n \rightarrow \infty$, the error

$$\Delta_n = \min_{1 \leq i \leq n} X(t_i) - M$$

converges to 0 for any sample path. Ritter [10] established that the best non-adaptive algorithm has error of order $n^{-1/2}$ as $n \rightarrow \infty$. This suggests that in our comparison between the performance of both the deterministic composite algorithm and the best random one we examine the behavior of the sequence

$$E(\sqrt{n}\Delta_n) = E \sqrt{n} \left(\min_{1 \leq i \leq n} X(t_i) - M \right), \quad (2)$$

where the expectation is with respect to the joint distribution of the observations and X in the case of a randomized algorithm.

In order to obtain a detailed understanding of the behavior of the Brownian path near the global minimum, we need to introduce processes and random variables associated with the 3-dimensional Bessel process into our analysis. The 3-dimensional Bessel process is the diffusion process that is identical in law to the modulus of a 3-dimensional Brownian motion. A 3-dimensional Bessel bridge from $(0, 0)$ to (t, y) is a 3-dimensional Bessel process starting from 0 at time 0 “conditioned to take the value y at time t ”; see [11]. Define a “two-sided Bessel process” R by

$$R(t) = \begin{cases} R_1(t) & \text{if } t \geq 0, \\ R_2(-t) & \text{if } t < 0, \end{cases} \quad (3)$$

where R_1 and R_2 are independent 3-dimensional Bessel processes. We need to introduce a random variable that will appear in the limit results that we will derive. Let

$$W = \min_{i=0, \pm 1, \pm 2, \dots} R(i + U), \quad (4)$$

where U is a uniformly distributed random variable on the unit interval, independent of R . This random variable has expectation $E(W) = -\zeta(1/2)/\sqrt{2\pi}$ (see [2]), where ζ is Riemann's zeta function.

Finally, we will use \Rightarrow to denote convergence in distribution; i.e., $X_n \Rightarrow X$ means that $Ef(X_n) \rightarrow Ef(X)$ for all bounded continuous functions f .

3. Probabilities and path decomposition

In this section we describe how the Brownian motion path can be "split" at the global minimum, yielding two independent Markov processes, one to either side. This path decomposition completely describes the behavior of the Brownian motion path near the global minimum, and is basic to our analysis of algorithms in the next section. (For an example of how the decomposition can be used to analyze adaptive algorithms, see [6].)

Let $f(t; x, y)$ be the density of the first hitting time from x to y . These densities are given by (see [12])

$$f(t; x, y) = \frac{|y - x|}{\sqrt{2\pi t^3}} \exp\left(-\frac{(y - x)^2}{2t}\right). \quad (5)$$

We will have use for the joint distribution of M , T , and $X(1)$, which was derived by Shepp (see [13]). Theorem 3.1, which is proved in [9], expresses the joint density as the product of first hitting time densities.

THEOREM 3.1. For $x \geq y$, $0 \geq y$, and $0 \leq t \leq 1$,

$$P(M \in dy, X(1) \in dx, T \in dt) = f(t; 0, y)f(1 - t; x, y) dy dx dt. \quad (6)$$

The marginal density ξ of T is the "arc-sine" density (see Feller [7]);

$$P(T \in dt)/dt = \xi(t) = \frac{1}{\pi\sqrt{t(1-t)}}, \quad 0 < t < 1. \quad (7)$$

The following result is a special case of a general result of Fitzsimmons ([8]) that decomposes the path of a diffusion process at the minimum (this result generalizes an earlier result of Williams [14]).

THEOREM 3.2. Given $(M = y, T = t, X(1) = x)$ ($0 < t < 1, y < x$), the process $\{X(t + u) - y\}_{0 \leq u \leq 1-t}$ is a 3-dimensional Bessel bridge from 0 at time 0 to $x - y$ at time $1 - t$, independent of $\{X(t - u) - y\}_{0 \leq u \leq t}$, which is a 3-dimensional Bessel bridge from 0 at time 0 to $-y$ at time t .

The following result, which is proved in [2], will be used to determine the limits of certain subsequences of $E(\sqrt{n} \Delta_n)$ as shown in the next section. Roughly, it says that if we place the origin of our coordinate system at the global minimum, then rescale time by a factor of n and space by a factor of \sqrt{n} , then the resulting process will, in the limit, resemble the two-sided Bessel process R defined by (3).

THEOREM 3.3. *Let $A > 0$ be fixed. Then conditionally on $(M = y, T = t, X(1) = x)$,*

$$\begin{aligned} &(\sqrt{n} (X(t + u/n) - y), \sqrt{n} (X(t - u/n) - y), 0 \leq u \leq A) \\ &\Rightarrow ((R_1(u), R_2(u)); 0 \leq u \leq A) \end{aligned}$$

in $C[0, A] \times C[0, A]$ as $n \rightarrow \infty$.

Also, setting

$$\delta_n^A = \min_{|s-T| > A/n} X(s) - M, \tag{8}$$

we have that for any $a > 0$,

$$\lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\sqrt{n} \delta_n^A \leq a) = 0. \tag{9}$$

4. Algorithm and error analysis

To motivate the composite algorithm that is the main object of our study, we will first discuss a similar but simpler one. Consider the algorithm that chooses the following sequence of observation points (label the points as d_0, d_1, d_2, \dots):

$$1, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \frac{1}{16}, \dots \tag{10}$$

In general,

$$d_n = \frac{2(n - 2^{\lfloor \lg(n) \rfloor}) + 1}{2^{\lfloor \lg(n) \rfloor + 1}}, \quad n \geq 1,$$

where \lg denotes logarithm to the base 2. (Recall that $X(0) = 0$, so that in effect we start off with the observation at 0.) If the number of observations is a power of 2, then the points form an equi-spaced grid, which one would expect to be efficient. Otherwise, the grid has some intervals twice as wide as others, which is clearly inefficient. The algorithm that we will describe uses the images of the above sequence $\{d_i\}$ under a continuous transformation of the unit interval. This corresponds to making the points the quantiles of a beta distribution, which is in a sense optimal, as pointed out in the Introduction.

Let H be the cumulative density of the Beta(2/3, 2/3) distribution,

$$H(t) = I(2/3, 2/3, t) = \mathcal{B}(2/3, 2/3)^{-1} \int_{s=0}^t \frac{ds}{[s(1-s)]^{1/3}} \quad (11)$$

for $0 \leq t \leq 1$, where \mathcal{B} denotes the Beta and I the incomplete Beta function. Let $h(t) = H'(t) = (\mathcal{B}(2/3, 2/3)[t(1-t)]^{1/3})^{-1}$ denote the corresponding probability density function, and define an algorithm by $t_0 = 1 = H^{-1}(1)$, and

$$t_n = H^{-1}(d_n), \quad n \geq 1. \quad (12)$$

Thus the t_n 's are the images under H^{-1} of the grid points defined by (10). We will show that this algorithm dominates the best random composite algorithm in the limit. That is, the sequence defined by (12) performs better than choosing the points independently from the Beta(2/3,2/3) distribution.

The sequence $E(\sqrt{n}\Delta_n)$ does not converge, and our aim is to determine its least upper bound and greatest lower bound (Theorem 4.2). In a first step, we will determine the limits of certain subsequences of $E(\sqrt{n}\Delta_n)$ as shown in Theorem 4.1.

THEOREM 4.1. *Let $\{n_k : k \geq 1\}$ be an increasing sequence of integers such that*

$$2^k \leq n_k < 2^{k+1}, \quad k \geq 1,$$

and

$$t_{n_k} \rightarrow \tau \in (0, 1)$$

as $k \rightarrow \infty$. Let

$$\beta(T) = \begin{cases} 1 & \text{if } T \geq \tau, \\ 2 & \text{if } T < \tau. \end{cases}$$

Then

$$\sqrt{2^k \beta(T) h(T)} \Delta_{n_k} \Rightarrow W \quad (13)$$

as $k \rightarrow \infty$, where W is defined by (4).

Furthermore,

$$E(\sqrt{n_k} \Delta_{n_k}) \rightarrow \sqrt{1 + H(\tau)} E(W) \frac{\mathcal{B}(2/3, 2/3)^{3/2}}{\pi} (1 - (1 - 2^{-1/2}) H(\tau)). \quad (14)$$

The proof of Theorem 4.1 is deferred to the appendix.

To picture the content of Theorem 4.1, it may be helpful to consider the simpler result obtained by replacing H by the identity function (in the Theorem, replace

$H(\tau)$ by τ and $h(T)$ by 1). The observation sequence is then $t_i = d_i$; i.e., choose the next point by bisecting the largest subinterval, choosing the leftmost subinterval in case of ties. Suppose we look at the grid at those time points $n_k = 2^k + 2^{k-1}$; we have a uniform grid of mesh 2^{-k} , and in addition have bisected all the intervals to the left of $1/2$. Basically, the grid to the right of $1/2$ is of mesh 2^{-k} and to the left it is 2^{-k-1} , and $t_{n_k} \rightarrow \tau = 1/2$. Then the theorem says that if $T < 1/2$, the asymptotic error is as if the observations formed a uniform grid of mesh 2^{-k-1} , while if $T > 1/2$, then the asymptotic error is as if the observations formed a uniform grid of mesh 2^{-k} . The effect of nonlinear H is captured by the h term in (13).

We are now ready for our main result.

THEOREM 4.2. *Under the algorithm described in this section,*

$$\begin{aligned} \limsup_{n \rightarrow \infty} E(\sqrt{n} \Delta_n) & \qquad \qquad \qquad (15) \\ &= E(W) \frac{\mathcal{B}(2/3, 2/3)^{3/2}}{\pi} \frac{4\sqrt{2} - 2}{3\sqrt{2}} \sqrt{\frac{2\sqrt{2} - 1}{3\sqrt{2} - 3}} \approx 0.5705, \end{aligned}$$

and

$$\liminf_{n \rightarrow \infty} E(\sqrt{n} \Delta_n) = E(W) \frac{\mathcal{B}(2/3, 2/3)^{3/2}}{\pi} \approx 0.5457. \qquad (16)$$

Proof. Denote the function of τ given by the limit in (14) by $\Psi(\tau)$; i.e.,

$$\Psi(\tau) = \sqrt{1 + H(\tau)} E(W) \frac{\mathcal{B}(2/3, 2/3)^{3/2}}{\pi} (1 - (1 - 2^{-1/2}) H(\tau)). \qquad (17)$$

Since H is differentiable and strictly increasing, we can find the value $\hat{\tau}$ maximizing Ψ , which is

$$\hat{\tau} = H^{-1} \left(\frac{\sqrt{2}}{3} \right),$$

with corresponding value given by the right hand side of (15). Therefore, since we know how to construct a subsequence with this limit, we know that

$$\limsup_{n \rightarrow \infty} E(\sqrt{n} \Delta_n) \geq \Psi(\hat{\tau}). \qquad (18)$$

To show that this is in fact the lim sup, suppose that there is a subsequence n_k such that

$$E(\sqrt{n_k} \Delta_{n_k}) \rightarrow \gamma > \Psi(\hat{\tau}).$$

Consider the sequence $\{t_{n_k}\}$. There is a subsequence n'_k such that $t_{n'_k} \rightarrow \tau$, say. By (14), which holds even if the n'_k are not all in the interval $[2^k, 2^{k+1}]$,

$$E\left(\sqrt{n'_k} \Delta_{n'_k}\right) \rightarrow \Psi(\tau) \leq \Psi(\hat{\tau}),$$

contradicting the assumption that

$$E(\sqrt{n_k} \Delta_{n_k}) \rightarrow \gamma > \Psi(\hat{\tau}),$$

and (15) is proved.

The lim inf occurs when $H(\tau) \in \{0, 1\}$ (i.e., $\tau \in \{0, 1\}$), which gives (16).

As mentioned in the Introduction, the best performance among algorithms that choose observations independently from a fixed probability distribution is obtained from the Beta(2/3, 2/3) distribution. The corresponding limiting normalized mean error for this random algorithm is

$$E(\sqrt{n} \Delta_n) \rightarrow \frac{1}{\pi \sqrt{2}} \mathcal{B}(2/3, 2/3)^{3/2} \approx 0.6623; \tag{19}$$

see [1]. Comparing this result with Theorem 4.2 shows that the deterministic composite algorithm has a better limiting performance in the sense that the lim sup of the normalized mean error is ≈ 0.5705 , considerably less than the limit in (19). In this case, the benefit of deterministic gaps outweighs the penalty of grid non-uniformity when n is not a power of 2.

Appendix

PROOF OF THEOREM 4.1

Let $P_{t,y,x}$ be a regular version of the conditional probability $P(\cdot | T = t, M = y, X(1) = x)$, and let

$$\Delta_{n_k}^A = \min\{X(t_{n_i}) - M; 1 \leq n_i \leq n_k, |t_{n_i} - T| \leq A/n_k\}.$$

Then,

$$\begin{aligned} &P\left(\sqrt{2^k \beta(T)h(T)} \Delta_{n_k}^A \leq z\right) \\ &= \int_{t,y,x} P_{t,y,x}\left(\sqrt{2^k \beta(t)h(t)} \Delta_{n_k}^A \leq z\right) P(T \in dt, M \in dy, X(1) \in dx), \\ &= \int_{t,y,x} P_{t,y,x}\left(\sqrt{2^k \beta(t)h(t)} \min_{|t_{n_i}-t| \leq A/n_k} (X(t_{n_i}) - y) \leq z\right) \\ &\quad \cdot P(T \in dt, M \in dy, X(1) \in dx). \end{aligned}$$

To simplify notation, it is convenient to relabel the observation sites according to their location with respect to T . Let t'_1 be the first observation to the right of T , t'_2 the next observation to the right of t'_1 , and so on. Similarly, Let t'_{-1} be the first observation to the left of T , t'_{-2} the next observation to the left of t'_{-1} , and so on. The t'_i depend on n , but to keep the notation simple we do not use the additional superscript. Considering the observation sites to the left and to the right of t , the integrand

$$P_{t,y,x} \left(\sqrt{2^k \beta(t)h(t)} \min_{|t_{n_i}-t| \leq A/n_k} (X(t_{n_i}) - y) \leq z \right)$$

becomes

$$P_{t,y,x} \left(\min_{|t'_j-t| \leq A/n_k} \sqrt{2^k \beta(t)h(t)} \cdot \left(X \left(t + \frac{2^k \beta(t)h(t)(t'_1 - t) + \sum_j \varphi_j(t)}{2^k \beta(t)h(t)} \right) - y \right) \leq z \right),$$

where

$$\varphi_j(t) = 2^k \beta(t)h(t)[t'_{j+1} - t'_j]. \tag{20}$$

Since $|t'_j - t| \leq A/n_k$, $\varphi_j(t) \rightarrow 1$ as $n_k \rightarrow \infty$ for each such j . This is because near t , the t'_j are the images under H^{-1} of points separated by $[2^k \beta(t)]^{-1}$. Therefore,

$$\begin{aligned} t'_{j+1} - t'_j &\approx \frac{1}{2^k \beta(t)} \frac{d}{ds} H^{-1}(s)|_{s=H(t)} \\ &= \frac{1}{2^k \beta(t)} \frac{1}{h(t)}. \end{aligned}$$

The process

$$\sqrt{2^k \beta(t)h(t)} \left(X \left(t + \frac{u}{2^k \beta(t)h(t)} \right) - y \right)_{-A \leq u \leq A} \tag{21}$$

converges in distribution to $\{R(u) : -A \leq u \leq A\}$ by Theorem 3.3 and the Bessel scaling property (if X is a 3-dimensional Bessel process, then so is $\sqrt{c}X(\cdot/c)$ for any $c > 0$). Also, $2^k \beta(T)h(T)(t'_1 - T) \Rightarrow U$, where U is uniform $(0, 1)$, since

$$\begin{aligned} &\int_{t,y,x} P_{t,y,x} (2^k \beta(t)h(t)(t'_1 - t) \leq z) P(T \in dt, M \in dy, X(1) \in dx) \\ &= \sum_{i=1}^n \int_{t=t'_i - z/2^k \beta(t)h(t)}^{t'_i} \xi(t) dt \rightarrow z, \end{aligned}$$

where the $\{t_i''\}$ are the $\{t_i\}$ relabeled in increasing order. The last limit follows from the fact that for $t_{i-1}'' < t < t_i''$,

$$2^k \beta(t)h(t)[t_i'' - t_{i-1}''] \rightarrow 1.$$

Therefore,

$$\sqrt{2^k \beta(T)h(T)} \Delta_{n_k}^A \Rightarrow W^A, \tag{22}$$

where

$$W^A = \min_{\substack{i=0,\pm 1,\pm 2,\dots \\ |i+U| \leq A}} R(i + U). \tag{23}$$

So, by the second part of Theorem 3.3,

$$\begin{aligned} & | P \left(\sqrt{2^k \beta(T)h(T)} \Delta_{n_k}^A \leq z \right) - P \left(\sqrt{2^k \beta(T)h(T)} \Delta_{n_k} \leq z \right) | \\ & \leq P \left(\sqrt{n_k} \delta_{n_k}^A \leq z \right) \end{aligned}$$

converges to 0 as n_k and then $A \rightarrow \infty$. Clearly $W^A \rightarrow W$ as $A \uparrow \infty$. Combining these facts with (22), we have

$$\begin{aligned} & | P \left(\sqrt{2^k \beta(T)h(T)} \Delta_{n_k} \leq z \right) - P \left(W \leq z \right) | \\ & \leq | P \left(\sqrt{2^k \beta(T)h(T)} \Delta_{n_k} \leq z \right) - P \left(\sqrt{2^k \beta(T)h(T)} \Delta_{n_k}^A \leq z \right) | \\ & \quad + | P \left(\sqrt{2^k \beta(T)h(T)} \Delta_{n_k}^A \leq z \right) - P \left(W^A \leq z \right) | \\ & \quad + | P \left(W^A \leq z \right) - P \left(W \leq z \right) | . \end{aligned}$$

Since all three terms on the right hand side of the inequality converge to 0, the first part of the theorem is proved.

The above steps actually show that

$$\left(\sqrt{\beta(T)h(T)}, \sqrt{2^k \beta(T)h(T)} \Delta_{n_k} \right) \Rightarrow \left(\sqrt{\beta(T)h(T)}, W \right), \tag{24}$$

where the limit random variables are independent. Therefore, by the continuous mapping theorem (see [3]),

$$\frac{\sqrt{2^k \beta(T)h(T)} \Delta_{n_k}}{\sqrt{\beta(T)h(T)}} = 2^{k/2} \Delta_{n_k} \Rightarrow \frac{W}{\sqrt{\beta(T)h(T)}}, \tag{25}$$

where W and $\sqrt{\beta(T)h(T)}$ are independent. It can be shown (see [2] for a similar calculation) that $\{2^{k/2} \Delta_{n_k} : n_k \geq 1\}$ is a uniformly integrable family of random variables. Therefore, we can conclude that

$$E \left(2^{k/2} \Delta_{n_k} \right) = \left(\frac{2^k}{n_k} \right)^{1/2} E \left(\sqrt{n_k} \Delta_{n_k} \right) \rightarrow E(W) E \left(\frac{1}{\sqrt{\beta(T)h(T)}} \right) \tag{26}$$

as $k \rightarrow \infty$. Since $t_{n_k} \rightarrow \tau$ and H is continuous, $H(t_{n_k}) \rightarrow H(\tau)$, or, by (12),

$$H(t_{n_k}) = \frac{2(n_k - 2^k) + 1}{2^{k+1}} = \frac{n_k}{2^k} - 1 + 2^{-k-1} \rightarrow H(\tau),$$

which implies that

$$\frac{n_k}{2^k} \rightarrow 1 + H(\tau)$$

as $k \rightarrow \infty$. Combining this with (26), we conclude that

$$E(\sqrt{n_k} \Delta_{n_k}) \rightarrow \sqrt{1 + H(\tau)} E(W) E\left(\frac{1}{\sqrt{\beta(T)h(T)}}\right). \quad (27)$$

Finally, we evaluate the last expectation appearing in (27):

$$\begin{aligned} & E\left(\frac{1}{\sqrt{\beta(T)h(T)}}\right) \\ &= \int_{t=0}^{\tau} \frac{\xi(t)}{\sqrt{2h(t)}} dt + \int_{t=\tau}^1 \frac{\xi(t)}{\sqrt{h(t)}} dt \\ &= \frac{\sqrt{\mathcal{B}(2/3, 2/3)}}{\pi} \left(\frac{1}{\sqrt{2}} \int_{t=0}^{\tau} [t(1-t)]^{-1/3} dt + \int_{t=\tau}^1 [t(1-t)]^{-1/3} dt \right) \\ &= \frac{\mathcal{B}(2/3, 2/3)^{3/2}}{\pi} \left(\frac{1}{\sqrt{2}} I(2/3, 2/3, \tau) + I(2/3, 2/3, 1) - I(2/3, 2/3, \tau) \right) \\ &= \frac{\mathcal{B}(2/3, 2/3)^{3/2}}{\pi} (1 - (1 - 2^{-1/2}) I(2/3, 2/3, \tau)) \\ &= \frac{\mathcal{B}(2/3, 2/3)^{3/2}}{\pi} (1 - (1 - 2^{-1/2}) H(\tau)). \end{aligned}$$

We can therefore rewrite (27) as

$$E(\sqrt{n_k} \Delta_{n_k}) \rightarrow \sqrt{1 + H(\tau)} E(W) \frac{\mathcal{B}(2/3, 2/3)^{3/2}}{\pi} (1 - (1 - 2^{-1/2}) H(\tau)),$$

which is (14).

References

1. Al-Mharmah, H. and Calvin, J. (1996), Optimal random non-adaptive algorithm for optimization of Brownian motion, *Journal of Global Optimization* 8: 81–90.
2. Asmussen, S., Glynn, P. and Pitman, J. (1995), Discretization error in simulation of one-dimensional reflecting Brownian motion, *The Annals of Applied Probability* 5: 875–896.
3. Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.
4. Calvin, J. (1995), Average performance of passive algorithms for global optimization, *J. Math. Anal. Appl.* 191: 608–617.

5. Calvin, J. (1996), Asymptotically optimal non-adaptive algorithms for minimization of Brownian motion. In *The Mathematics of Numerical Analysis* (J. Renegar, M. Shub, S. Smale, eds.) American Mathematical Society, Lectures in Applied Mathematics, Vol. 32.
6. Calvin, J. (1997), Average performance of a class of adaptive algorithms for global optimization, *The Annals of Applied Prob.* 7: 711–730.
7. Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, II, Wiley, New York.
8. Fitzsimmons, P. J., Pitman, J. W. and Yor, M. (1992), Markovian bridges: Construction, Palm interpretation, and Splicing, *Sem. Stochastic Processes*, Birkhäuser, Boston.
9. Imhof, J. P. (1984), Density factorization for Brownian motion, meander and the three-dimensional Bessel process, and applications, *J. Appl. Prob.* 21: 500–510.
10. Ritter, K. (1990), Approximation and optimization on the Wiener space, *J. Complexity* 6: 337–364.
11. Revuz, D. and M. Yor (1991), *Continuous Martingales and Brownian Motion*, Springer Verlag, Berlin/New York.
12. Karlin, S. and Taylor, H. (1975), *A First Course in Stochastic Processes*, Academic Press, New York.
13. Shepp, L. A. (1976), The joint density of the maximum and its location for a Wiener process with drift, *J. Appl. Prob.* 16: 423–427.
14. Williams, D. (1974), Path decomposition and continuity of local time for one-dimensional diffusions, I, *Proc. London Math. Soc.*, Ser. 3 28: 738–68.
15. Zhigljavsky, A. (1991), *Theory of Global Random Search*, Kluwer Academic Publishers, Dordrecht/Boston/London.