ADAPTIVE APPROXIMATION OF THE MINIMUM OF BROWNIAN MOTION

JAMES M. CALVIN, MARIO HEFTER, AND ANDRÉ HERZWURM

Abstract. We study the error in approximating the minimum of a Brownian motion on the unit interval based on finitely many point evaluations. We construct an algorithm that adaptively chooses the points at which to evaluate the Brownian path. In contrast to the 1/2 convergence rate of optimal nonadaptive algorithms, the proposed adaptive algorithm converges at an arbitrarily high polynomial rate.

1. Introduction

We study the pathwise approximation of the minimum

$$M = \inf_{0 \le t \le 1} W(t)$$

of a Brownian motion W on the unit interval $[0,1]$ based on adaptively chosen function values of W. In contrast to nonadaptive algorithms, which evaluate a function always at the same points, adaptive algorithms may sequentially choose points at which to evaluate the function. For the present problem, this means that the n-th evaluation site may depend on the first $n-1$ observed values of the Brownian path W. Given a number of evaluation sites, we are interested in algorithms that have a small error in the residual sense with respect to the L_p-norm.

A key motivation for studying this approximation problem stems from numerics for the reflected Brownian motion given by

$$\hat{W}(t) = W(t) - \inf_{0 \le s \le t} W(s).$$

Apart from its use in queueing theory [7], the reflected Brownian motion also appears in the context of nonlinear stochastic differential equations. More precisely, the solution process of a particular instance of a Cox-Ingersoll-Ross process is given by the square of $\hat{W}$. Hence numerical methods for the approximation of M can be used for the approximation of $\hat{W}$ and thus for the corresponding Cox-Ingersoll-Ross process. We refer to [8] for such an application of the algorithm proposed in this paper.

The complexity analysis of pathwise approximation of the Brownian minimum M based on finitely many function evaluations was initiated in [17], where it was shown that for any nonadaptive algorithm using n function evaluations the average error is at least of order $n^{-1/2}$. Moreover, a simple equidistant discretisation already has an error of order $n^{-1/2}$, and thus achieves the lower bound for nonadaptive algorithms. A detailed analysis of the asymptotics of the pathwise error in case of an equidistant discretisation was undertaken in [1].

The situation regarding adaptive algorithms for the pathwise approximation of M is rather different. In [5], it was shown that for any (adaptive) algorithm using n function

Key words and phrases. Brownian motion; global optimization; pathwise approximation; adaptive algorithm.

evaluations the average error is at least of order $\exp(-c\,n/\log(n))$ for some positive constant c. In contrast to the nonadaptive case, we are unaware of algorithms with error bounds matching the lower bound for adaptive algorithms. In this paper we analyze an adaptive algorithm that has an average error at most of order n^{-r}, for any positive number r. Hence this algorithm converges at an arbitrarily high polynomial rate. In [6], the same algorithm was shown to converge in a probabilistic sense. We are unaware of previous results showing the increased power of adaptive methods relative to nonadaptive methods with respect to the L_p error.

Several optimization algorithms have been proposed that use the Brownian motion as a model for an unknown function to be minimized, including [10, 12, 21, 3]. One of the ideas proposed in [10] is to evaluate the function next at the point where the function has the maximum conditional probability of having a value less than the minimum of the conditional mean, minus some positive amount (tending to zero). This is the same idea behind our algorithm, described in Section 2. The question of convergence of such (Bayesian) methods in general is addressed in [13]. Several algorithms, with an emphasis on the question of convergence, are described in [19].

In global optimization, the function to be optimized is typically assumed to be a member of some class of functions. Often, the worst-case error of algorithms on such a class of functions is studied. However, if the function class is convex and symmetric, then the worst-case error for any method using n function evaluations is at least as large as the error of a suitable nonadaptive method using $n+1$ evaluations, see, e.g., [14, Chap. 1.3]. In this case, a worst-case analysis cannot justify the use of adaptive algorithms for global optimization. An average-case analysis, where it is assumed that the function to be optimized is drawn from a probability distribution, is an alternative to justify adaptive algorithms for general function classes. Brownian motion is suitable for such an average-case study since its analysis is tractable, yet the answers to the complexity questions are far from obvious. As already explained, adaptive methods are much more powerful than nonadaptive methods for optimization of Brownian motion.

This paper is organized as follows. In Section 2 we present our algorithm with corresponding error bound, see Theorem 1. In Section 3 we illustrate our results by numerical experiments. The rest of the paper is devoted to proving Theorem 1.

2. Algorithm and Main Result

Let $f\colon [0,1] \to \mathbb{R}$ be a continuous function with $f(0) = 0$. We will recursively define a sequence

$$t_0, t_1, \ldots \in [0,1] \tag{1}$$

of pairwise distinct points from the unit interval. These points are chosen adaptively, i.e., the k-th evaluation site t_k may depend on the previous values $t_0, f(t_0), \ldots, t_{k-1}, f(t_{k-1})$. We use the discrete minimum over these points given by

$$M_n = \min_{0 \le i \le n} f(t_i)$$

for $n \in \mathbb{N}_0$, as an approximation of the global minimum

$$M = \inf_{0 \le t \le 1} f(t)$$

of f. The aim is that M_n is a "good" approximation of M on average if f is a Brownian motion.

We begin by introducing some notation. For $n \in \mathbb{N}_0$ we denote the ordered first n evaluation sites by

$$0 \leq t_0^n < t_1^n < \ldots < t_n^n \leq 1$$

such that $\{t_i^n : 0 \leq i \leq n\} = \{t_i : 0 \leq i \leq n\}$. Furthermore, for $n \in \mathbb{N}$ let

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

be the smallest distance between two evaluation sites. Moreover, we define $g\colon]0,1] \to [0,\infty[$ by

$$g(x) = \sqrt{\lambda x \log(1/x)}\,,$$

where log denotes the natural logarithm. Here, $\lambda \in [1,\infty[$ is a fixed parameter, which is convenient to be left unspecified at this point.

Now, we define the sequence appearing in (1). The first two evaluation sites are non-adaptively chosen to be $t_1 = 1$ and $t_2 = 1/2$. Moreover, for notational convenience we set $t_0 = 0$.

Let $n \geq 2$, and suppose that the algorithm has already constructed the first n points $t_0, \ldots, t_n$. The key quantity for choosing the next evaluation site is given by

$$\rho_i^n = \frac{t_i^n - t_{i-1}^n}{\big(f(t_{i-1}^n) - M_n + g(\tau_n)\big)\left(f(t_i^n) - M_n + g(\tau_n)\right)} \tag{2}$$

for $i \in \{1, \ldots, n\}$. The algorithm splits the interval with the largest value of ρ_i^n at the midpoint. More precisely, let $j \in \{1, \ldots, n\}$ be the smallest index such that $\rho_j^n = \rho^n$ where

$$\rho^n = \max_{1 \leq i \leq n} \rho_i^n.$$

The next function evaluation is then made at the midpoint

$$t_{n+1} = \frac{1}{2}\left(t_{j-1}^n + t_j^n\right)$$

of the corresponding subinterval.

As we use the discrete minimum M_n as an approximation of the global minimum M, the error of the proposed algorithm is given by

$$\Delta_n = \Delta_{n,\lambda}(f) = M_n - M$$

for $n \in \mathbb{N}_0$.

We stress that all quantities defined above depend on the prespecified choice of the parameter $\lambda \in [1,\infty[$. In particular, λ affects all adaptively chosen evaluation sites $t_3, t_4, \ldots$ and hence M_n. However, we often do not explicitly indicate this dependence.

The following theorem shows that this algorithm achieves an arbitrarily high polynomial convergence rate w.r.t. the L_p-norm in case of a Brownian motion $W = (W(t))_{0 \leq t \leq 1}$.

Theorem 1. *For all $r \in [1,\infty[$ and for all $p \in [1,\infty[$ there exist $\lambda \in [1,\infty[$ and $c > 0$ such that*

$$\left(\mathrm{E}\left(|\Delta_{n,\lambda}(W)|^p\right)\right)^{1/p} \leq c \cdot n^{-r}$$

for all $n \in \mathbb{N}$.

Remark 1. Our analysis shows that

$$\lambda \geq 144 \cdot (1 + p \cdot r)$$

is sufficient to obtain convergence order r w.r.t. the L_p-norm in Theorem 1. However, numerical experiments indicate an exponential decay even for small values of λ, see Figure 2.

Remark 2. The number of function evaluations made by the algorithm to produce the approximation M_n is a fixed number $n \in \mathbb{N}$ (we assume that $f(0) = 0$ and so we do not count t_0). Thus we do not consider adaptive stopping rules. A straightforward implementation of this algorithm on a computer requires operations of order n^2.

An intuitive explanation why this algorithm works in the case of Brownian motion is as follows. The function g is chosen such that $M_n - M \leq g(\tau_n)$ with high probability if f is a Brownian path. The idea of the algorithm is to next evaluate the function at the midpoint of the subinterval that is most likely to have a value less than $M_n - g(\tau_n)$. Conditional on the values observed up to time n, the probability that the minimum over $[t_{i-1}^n, t_i^n]$ is less than $M_n - g(\tau_n)$ is

$$\exp\left(-\frac{2}{\rho_i^n}\right), \tag{3}$$

see [4]. The behavior of the $\{\rho_i^n\}$ defined in (2) is more convenient to characterize under the proposed algorithm than the probabilities given in (3).

The proof of Theorem 1 relies on two sets of preliminary results. Section 4 establishes upper bounds for the error when the algorithm is applied to certain sets of functions, culminating in Corollary 2. In Section 5, we bound the Wiener measure of these sets of functions, leading to Corollary 3. Section 6 combines these results to prove Theorem 1.

3. Numerical Results

In this section we present numerical results of the proposed algorithm for different values of the parameter λ. Figure 1 shows the error Δ_n for each of three independently generated Wiener paths using $\lambda = 1$.

We also performed numerical experiments to estimate $(\mathrm{E}(|\Delta_{n,\lambda}(W)|^p))^{1/p}$ using 10^3 replications. Figure 2 shows the results for $p = 2$ and $\lambda \in \{1, 4, 8\}$. We observe an exponential decay of the L_2 error for each value of λ. Let us recall that Theorem 1 and Remark 1 only show that sufficiently large values of λ ensure a "high" polynomial convergence rate of the L_p error. However, from a numerical point of view one might prefer choosing a small λ since the numerically observed error in Figure 2 is increasing in λ for a fixed number of evaluation sites. Let us mention that a small λ corresponds to a small offset $g(\tau_n)$ to the discrete minimum M_n in (2). Hence a small λ results in a "more local search" around the discrete minimum.

4. Non-probabilistic Arguments

In this section we will define a sequence of subsets of "favorable" functions for which we show that the error of the algorithm decreases at an exponential rate.

First, let us mention some basic facts, which will be frequently used in this paper. Due to the bisection strategy, the lengths of all subintervals satisfy

$$t_i^n - t_{i-1}^n \in \mathcal{A} = \{1/2^k : k \in \mathbb{N}\}$$

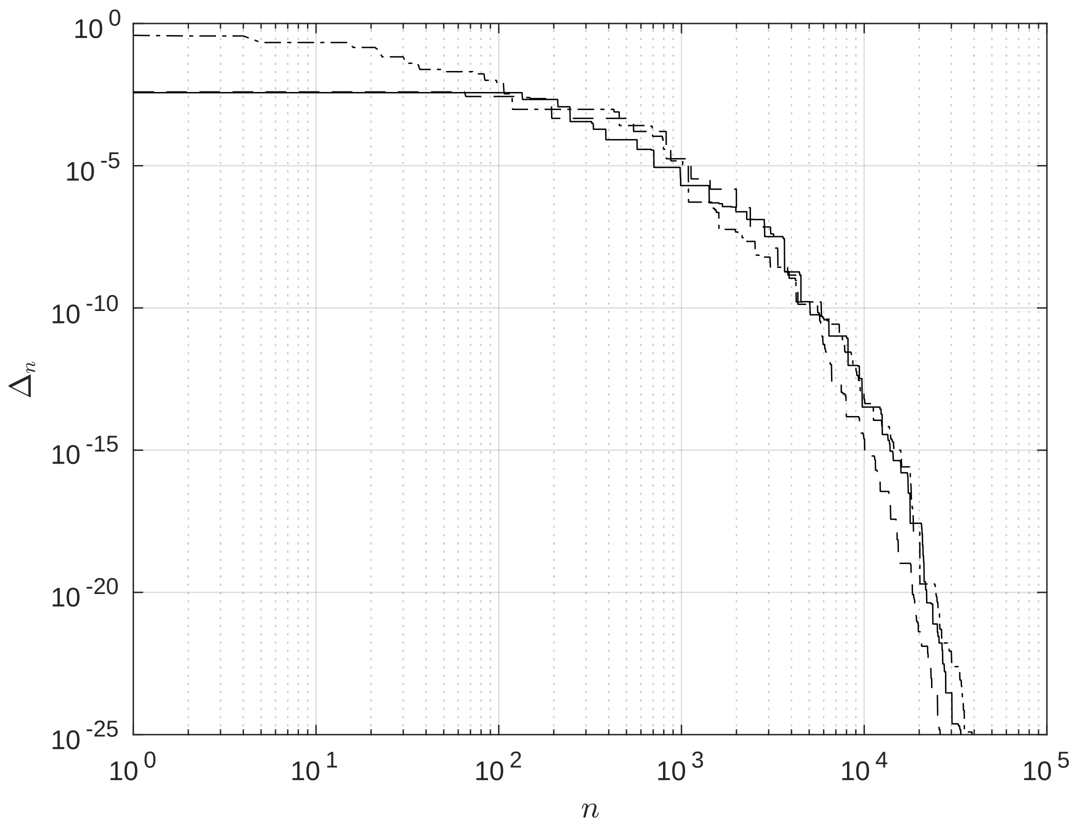

FIGURE 1. Errors for 3 sample paths using $\lambda = 1$.

for all $n \geq 2$ and $i \in \{1, \ldots, n\}$, and consequently $\tau_n \in \mathcal{A}$ for all $n \geq 2$. Let us stress that g is non-decreasing on $\mathcal{A}$. Furthermore, we have $\lim_{x \to 0} g(x) = 0$.

Let

$$F = \{f \colon [0,1] \to \mathbb{R},\ f \text{ continuous with } f(0) = 0\}.$$

Moreover, for $n \geq 2$ and $\lambda \in [1, \infty[$ we define

$$F_n = F_{\lambda,n} = \left\{ f \in F : \max_{1 \leq k \leq n} \max_{1 \leq i \leq k} \frac{\left| f(t_i^k) - f(t_{i-1}^k) \right|}{\sqrt{t_i^k - t_{i-1}^k}} \leq \sqrt{\lambda \log(n)/4} \right\}.$$

The sets of "favorable" functions, defined in (17) below, will be the intersection of several sets, including F_n, which depend on the prespecified parameter λ of the algorithm. To simplify the notation, we will suppress the dependence of these sets on λ after their definition. Recall that most quantities defined above depend on λ, n, and f simultaneously. However, we typically only highlight the dependence on n. For instance, ρ_i^n also depends on the corresponding function f as well as on the parameter λ.

In the following we present some properties of the algorithm applied to functions $f \in F_n$, which will be frequently used in this paper.

Lemma 1. *For all $\lambda \geq 1$, $n \geq 2$, and $f \in F_n$ we have*

$$\rho^n \leq \frac{2}{\lambda \log(1/\tau_n)}.$$

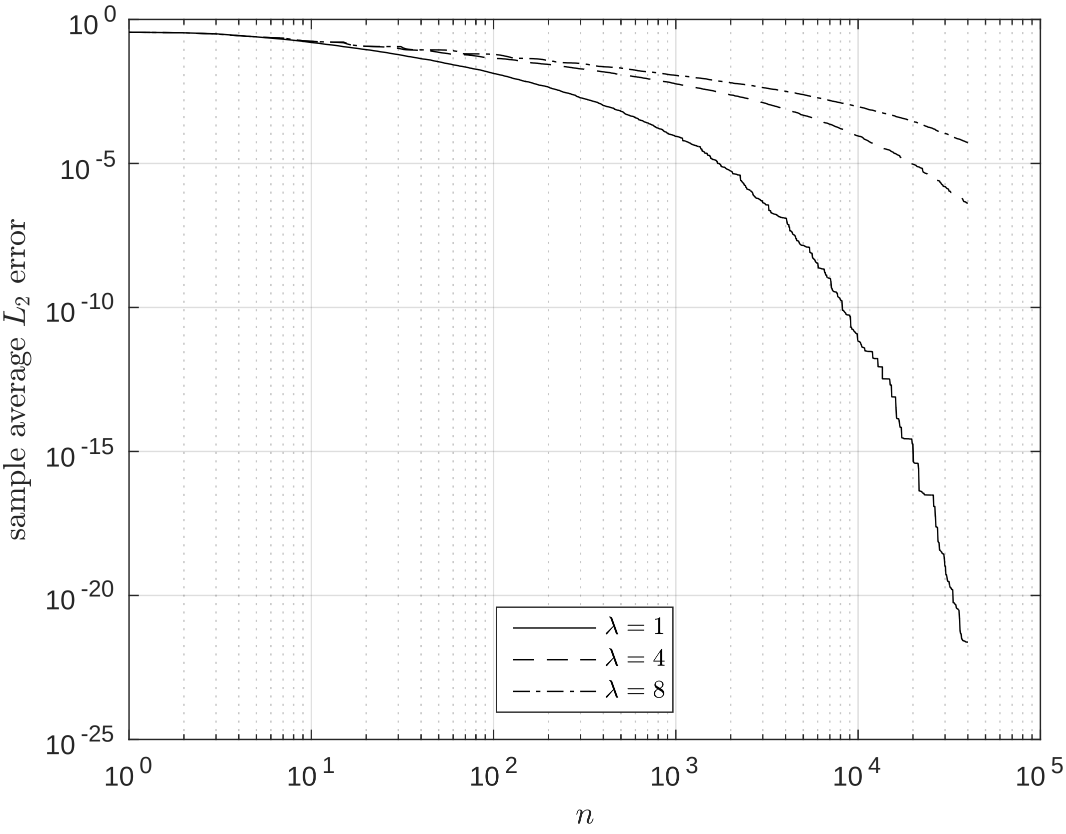

FIGURE 2. Sample L_2 error for various λ.

In particular, $\rho^n \leq 2/(\lambda \log(n))$.

Proof. First, we observe that

$$\rho^m \leq \frac{\tau_m}{g(\tau_m)^2} = \frac{1}{\lambda \log(1/\tau_m)}, \tag{4}$$

whenever the algorithm is about to split a smallest subinterval at step $m \geq 2$. In the following step $m+1$, we also clearly have

$$\rho_i^{m+1} \leq \frac{\tau_{m+1}}{g(\tau_{m+1})^2} = \frac{1}{\lambda \log(1/\tau_{m+1})},$$

if $i \in \{1, \ldots, m+1\}$ corresponds to one of the newly created smallest subintervals. If $j \in \{1, \ldots, m+1\}$ denotes a subinterval that has not been split at step m we obtain

$$\begin{aligned}
\rho_j^{m+1} &= \frac{t_j^{m+1} - t_{j-1}^{m+1}}{\big(f(t_{j-1}^{m+1}) - M_{m+1} + g(\tau_{m+1})\big)\big(f(t_j^{m+1}) - M_{m+1} + g(\tau_{m+1})\big)} \\
&\leq \left(\frac{g(\tau_m)}{g(\tau_{m+1})}\right)^2 \cdot \frac{t_j^{m+1} - t_{j-1}^{m+1}}{\big(f(t_{j-1}^{m+1}) - M_m + g(\tau_m)\big)\big(f(t_j^{m+1}) - M_m + g(\tau_m)\big)} \\
&\leq \left(\frac{g(\tau_m)}{g(\tau_{m+1})}\right)^2 \cdot \rho^m,
\end{aligned}$$

since g is non-decreasing on $\mathcal{A}$. Moreover, $\tau_{m+1} = \tau_m/2$ and (4) imply

$$\rho_j^{m+1} \le \frac{2\tau_{m+1}\log(1/\tau_m)}{\tau_{m+1}\log(1/\tau_{m+1})} \cdot \frac{1}{\lambda\log(1/\tau_m)} = \frac{2}{\lambda\log(1/\tau_{m+1})}$$

and thus

$$\rho^{m+1} \le \frac{2}{\lambda\log(1/\tau_{m+1})}. \tag{5}$$

Let $n \ge 2$ be arbitrary and let $m \in \{2,\dots,n\}$ be the last time that the algorithm was about to split a smallest subinterval, thus (4) holds. Let us stress that $\tau_{m+k} = \tau_n$ for all $k \in \{1,\dots,n-m\}$. We will show by induction that

$$\rho^{m+k} \le \frac{2}{\lambda\log(1/\tau_{m+k})} \tag{6}$$

for all $k \in \{0,\dots,n-m\}$.

We consider the non-trivial case of $m < n-1$, and we assume that (6) holds for some $k \in \{1,\dots,n-m-1\}$. At iteration $m+k+1$, we suppose that the i-th subinterval was split at step $m+k$, thus $\rho^{m+k} = \rho_i^{m+k}$, and consequently $t_{m+k+1} = (t_{i-1}^{m+k} + t_i^{m+k})/2$. Then we have

$$f(t_{m+k+1}) = \frac{f(t_{i-1}^{m+k}) + f(t_i^{m+k})}{2} + \delta,$$

for some $\delta \in \mathbb{R}$. Furthermore, $f \in F_n$ implies

$$\left| \frac{f(t_{i-1}^{m+k}) + f(t_i^{m+k})}{2} + \delta - f(t_{i-1}^{m+k}) \right| \le \sqrt{T_i/2} \cdot \sqrt{\lambda\log(n)/4}$$

and

$$\left| \frac{f(t_{i-1}^{m+k}) + f(t_i^{m+k})}{2} + \delta - f(t_i^{m+k}) \right| \le \sqrt{T_i/2} \cdot \sqrt{\lambda\log(n)/4}$$

for $T_i = t_i^{m+k} - t_{i-1}^{m+k}$. In particular, this yields

$$\delta \ge \frac{\left| f(t_i^{m+k}) - f(t_{i-1}^{m+k}) \right|}{2} - \sqrt{T_i/2} \cdot \sqrt{\lambda\log(1/\tau_n)/4} \tag{7}$$

since $\tau_n \le 1/n$. We obtain

$$\begin{aligned} \frac{\rho_i^{m+k}}{\rho_i^{m+k+1}} &= 2 \cdot \frac{\big(f(t_{i-1}^{m+k}) - M_{m+k+1} + g(\tau_n)\big)\,(f(t_{m+k+1}) - M_{m+k+1} + g(\tau_n))}{\big(f(t_{i-1}^{m+k}) - M_{m+k} + g(\tau_n)\big)\,\big(f(t_i^{m+k}) - M_{m+k} + g(\tau_n)\big)} \\ &\ge \frac{2\,(f(t_{m+k+1}) - M_{m+k+1} + g(\tau_n))}{\big(f(t_i^{m+k}) - M_{m+k} + g(\tau_n)\big)}. \end{aligned}$$

Moreover, we have

$$\begin{aligned} &2\,(f(t_{m+k+1}) - M_{m+k+1} + g(\tau_n)) \\ &\quad \ge f(t_{i-1}^{m+k}) + f(t_i^{m+k}) + \left| f(t_i^{m+k}) - f(t_{i-1}^{m+k}) \right| - \sqrt{\lambda T_i \log(1/\tau_n)/2} - 2M_{m+k} + 2g(\tau_n) \\ &\quad = \big(f(t_{i-1}^{m+k}) - M_{m+k} + g(\tau_n)\big) + \big(f(t_i^{m+k}) - M_{m+k} + g(\tau_n)\big) \\ &\qquad + \left| \big(f(t_i^{m+k}) - M_{m+k} + g(\tau_n)\big) - \big(f(t_{i-1}^{m+k}) - M_{m+k} + g(\tau_n)\big) \right| - \sqrt{\lambda T_i \log(1/\tau_n)/2} \end{aligned}$$

due to (7). Making the substitution

$$a = f(t_i^{m+k}) - M_{m+k} + g(\tau_n) > 0 \quad \wedge \quad b = f(t_{i-1}^{m+k}) - M_{m+k} + g(\tau_n) > 0$$

and $x = \sqrt{a/b} > 0$ we get

$$\begin{aligned} \frac{\rho_i^{m+k}}{\rho_i^{m+k+1}} &\geq \frac{a + b + |a-b| - \sqrt{T_i} \cdot \sqrt{\lambda \log(1/\tau_n)/2}}{a} \\ &= \frac{x + \frac{1}{x} + \left|x - \frac{1}{x}\right| - \sqrt{\rho_i^{m+k}} \cdot \sqrt{\lambda \log(1/\tau_n)/2}}{x} \\ &\geq \frac{x + \frac{1}{x} + \left|x - \frac{1}{x}\right| - 1}{x} \\ &\geq 1, \end{aligned}$$

where the second inequality holds by the induction hypothesis. Hence we get

$$\rho_i^{m+k+1} \leq \rho_i^{m+k}.$$

Similarly, we obtain $\rho_{i+1}^{m+k+1} \leq \rho_i^{m+k}$ and thus $\rho^{m+k+1} \leq \rho^{m+k}$. □

Lemma 2. *For all $\lambda \geq 1$, $n \geq 2$, and $f \in F_n$ we have*

$$\max_{1 \leq i \leq n} \frac{t_i^n - t_{i-1}^n}{(\min\{f(t_{i-1}^n), f(t_i^n)\} - M_n + g(\tau_n))^2} \leq \frac{4}{\lambda \log(n)}.$$

Proof. Let $n \geq 2$. At first, observe that

$$\frac{t_i^n - t_{i-1}^n}{(\min\{f(t_{i-1}^n), f(t_i^n)\} - M_n + g(\tau_n))^2} = \rho_i^n \cdot \left(1 + \frac{\left|f(t_i^n) - f(t_{i-1}^n)\right|}{\min\{f(t_{i-1}^n), f(t_i^n)\} - M_n + g(\tau_n)}\right)$$

for $i \in \{1, \ldots, n\}$, and thus

$$\begin{aligned} &\max_{1 \leq i \leq n} \frac{t_i^n - t_{i-1}^n}{(\min\{f(t_{i-1}^n), f(t_i^n)\} - M_n + g(\tau_n))^2} \\ &\quad \leq \rho^n \cdot \left(1 + \max_{1 \leq i \leq n} \frac{\left|f(t_i^n) - f(t_{i-1}^n)\right|}{\min\{f(t_{i-1}^n), f(t_i^n)\} - M_n + g(\tau_n)}\right) \\ &\quad = \rho^n \cdot \left(1 + \max_{1 \leq i \leq n} \frac{\left|f(t_i^n) - f(t_{i-1}^n)\right| \cdot \sqrt{t_i^n - t_{i-1}^n}}{\sqrt{t_i^n - t_{i-1}^n} \cdot \left(\min\{f(t_{i-1}^n), f(t_i^n)\} - M_n + g(\tau_n)\right)}\right) \\ &\quad \leq \rho^n \cdot \left(1 + \max_{1 \leq i \leq n} \frac{\left|f(t_i^n) - f(t_{i-1}^n)\right|}{\sqrt{t_i^n - t_{i-1}^n}} \cdot \max_{1 \leq i \leq n} \frac{\sqrt{t_i^n - t_{i-1}^n}}{\min\{f(t_{i-1}^n), f(t_i^n)\} - M_n + g(\tau_n)}\right). \end{aligned}$$

Setting

$$z_n = \max_{1 \leq i \leq n} \frac{\sqrt{t_i^n - t_{i-1}^n}}{\min\{f(t_{i-1}^n), f(t_i^n)\} - M_n + g(\tau_n)},$$

the last inequality reads

$$z_n^2 \leq \rho^n \cdot \left(1 + z_n \cdot \max_{1 \leq i \leq n} \frac{\left|f(t_i^n) - f(t_{i-1}^n)\right|}{\sqrt{t_i^n - t_{i-1}^n}}\right).$$

Now use the fact that $z_n > 0$, $f \in F_n$ and $\rho^n \leq 2/(\lambda \log(n))$ from Lemma 1 to obtain

$$z_n^2 \leq \frac{2}{\lambda \log(n)} \cdot \left(1 + z_n \cdot \sqrt{\lambda \log(n)/4}\right),$$

or equivalently,

$$\left(z_n - 1/\sqrt{4\lambda \log(n)}\right)^2 \leq \left(2 + \frac{1}{4}\right) \cdot \frac{1}{\lambda \log(n)}.$$

This implies that

$$z_n \leq \frac{1}{\sqrt{\lambda \log(n)}} \left(1/2 + \sqrt{2 + 1/4}\right) = \frac{2}{\sqrt{\lambda \log(n)}}$$

and hence $z_n^2 \leq 4/(\lambda \log(n))$. □

Lemma 3. *For all $\lambda \geq 1$, $n \geq 2$, and $f \in F_n$ we have*

$$\rho^k \geq \frac{2}{3\lambda \log(1/\tau_n)}$$

for all $k \in \{2, \dots, n\}$.

Proof. Let $n \geq 2$, $k \in \{2, \dots, n\}$ and $f \in F_n$. Moreover, let $i \in \{1, \dots, k\}$ be an index with $M_k \in \{f(t_{i-1}^k), f(t_i^k)\}$ and $T_i = t_i^k - t_{i-1}^k$. Then we have

$$\begin{aligned}
\rho^k \geq \rho_i^k &= \frac{T_i}{g(\tau_k) \cdot \left(\max\{f(t_{i-1}^k), f(t_i^k)\} - \min\{f(t_{i-1}^k), f(t_i^k)\} + g(\tau_k)\right)} \\
&= \frac{1}{\sqrt{\lambda \log(1/\tau_k) \cdot (\tau_k/T_i)} \cdot \left(\left|f(t_i^k) - f(t_{i-1}^k)\right| / \sqrt{T_i} + \sqrt{\lambda \log(1/\tau_k) \cdot (\tau_k/T_i)}\right)} \\
&\geq \frac{1}{\sqrt{\lambda \log(1/\tau_k)} \left(\sqrt{\lambda \log(n)/4} + \sqrt{\lambda \log(1/\tau_k)}\right)} \\
&\geq \frac{1}{\sqrt{\lambda \log(1/\tau_n)} \left(\sqrt{\lambda \log(1/\tau_n)/4} + \sqrt{\lambda \log(1/\tau_n)}\right)} \\
&\geq \frac{2}{3\lambda \log(1/\tau_n)},
\end{aligned}$$

since $\tau_n \leq 1/n$ and $\tau_n \leq \tau_k$. □

4.1. **Upper Bound on $\sum_{i=1}^n \rho_i^n$.** For $\lambda \in [1, \infty[$, $n \in \mathbb{N}$, and $f \in F$ we consider the linear interpolation L_n of the $\{f(t_i^n)\}$ defined by

$$L_n(s) = \frac{t_i^n - s}{t_i^n - t_{i-1}^n} \cdot f(t_{i-1}^n) + \frac{s - t_{i-1}^n}{t_i^n - t_{i-1}^n} \cdot f(t_i^n), \quad s \in [t_{i-1}^n, t_i^n], \quad 1 \leq i \leq n. \tag{8}$$

Remark 3. A simple computation shows that

$$\int_s^t \frac{1}{(h(x))^2} \, \mathrm{d}x = \frac{t - s}{h(s) \cdot h(t)}$$

for $s < t$ and $h\colon [s, t] \to \mathbb{R}$ affine linear with $h(s), h(t) > 0$. This yields

$$\rho_i^n = \int_{t_{i-1}^n}^{t_i^n} \frac{1}{(L_n(s) - M_n + g(\tau_n))^2} \, \mathrm{d}s$$

and hence

$$\sum_{i=1}^{n} \rho_i^n = \int_0^1 \frac{1}{(L_n(t) - M_n + g(\tau_n))^2} \, \mathrm{d}t \tag{9}$$

for $f \in F$, $n \geq 2$, and $\lambda \geq 1$.

Replacing the discrete minimum by the global minimum in (9) clearly yields the lower bound

$$\int_0^1 \frac{1}{(L_n(t) - M + g(\tau_n))^2} \, \mathrm{d}t \leq \sum_{i=1}^{n} \rho_i^n.$$

In the following we provide an upper bound of similar structure.

For $n \geq 2$ and $\lambda \in [1, \infty[$ we define

$$G_{1,n} = G_{\lambda,1,n} = \{f \in F : \Delta_n \leq g(\tau_n)\}$$

and

$$G_{1/2,n} = G_{\lambda,1/2,n} = \left\{ f \in F : \Delta_n \leq \frac{1}{2}\, g(\tau_n) \right\}.$$

We clearly have $G_{1/2,n} \subseteq G_{1,n}$.

Lemma 4. *For all $\lambda \geq 1$, $n \geq 2$, and $f \in G_{1/2,n}$ we have*

$$\sum_{i=1}^{n} \rho_i^n \leq 4 \cdot \int_0^1 \frac{1}{(L_n(t) - M + g(\tau_n))^2} \, \mathrm{d}t.$$

Proof. Using (9) we obtain

$$\begin{aligned} \sum_{i=1}^{n} \rho_i^n &= \int_0^1 \frac{1}{(L_n(t) - M + g(\tau_n) - \Delta_n)^2} \, \mathrm{d}t \\ &\leq \int_0^1 \frac{1}{\left(L_n(t) - M + \frac{1}{2}\, g(\tau_n)\right)^2} \, \mathrm{d}t \\ &\leq \frac{1}{(1/2)^2} \cdot \int_0^1 \frac{1}{(L_n(t) - M + g(\tau_n))^2} \, \mathrm{d}t. \end{aligned}$$

□

In the next step we bound $\sum_{i=1}^{n} \rho_i^n$ in terms of

$$\int_0^1 \frac{1}{(f(t) - M + g(\tau_n))^2} \, \mathrm{d}t.$$

For $n \geq 2$ and $\lambda \in [1, \infty[$ we define

$$J_n^+ = J_{\lambda,n}^+ = \left\{ f \in F : \max_{1 \leq i \leq n} \sup_{s \in [t_{i-1}^n, t_i^n]} \frac{f(s) - L_n(s)}{\sqrt{t_i^n - t_{i-1}^n}} \leq \frac{1}{2}\sqrt{\lambda \log(n)/4} \right\} \tag{10}$$

and

$$J_n^- = J_{\lambda,n}^- = \left\{ f \in F : \min_{1 \leq i \leq n} \inf_{s \in [t_{i-1}^n, t_i^n]} \frac{f(s) - L_n(s)}{\sqrt{t_i^n - t_{i-1}^n}} \geq -\frac{1}{2}\sqrt{\lambda \log(n)/4} \right\}. \tag{11}$$

Lemma 5. *For all $\lambda \geq 1$, $n \geq 2$, and $f \in F_n \cap J_n^+ \cap J_n^-$ we have*

$$\int_0^1 \frac{1}{(L_n(t) - M + g(\tau_n))^2} \,\mathrm{d}t \leq \frac{9}{4} \cdot \int_0^1 \frac{1}{(f(t) - M + g(\tau_n))^2} \,\mathrm{d}t.$$

Proof. Let $i \in \{1, \dots, n\}$. We have for all $s \in [t_{i-1}^n, t_i^n]$ that

$$\begin{aligned} \frac{|f(s) - L_n(s)|}{L_n(s) - M + g(\tau_n)} &\leq \frac{|f(s) - L_n(s)|}{\sqrt{t_i^n - t_{i-1}^n}} \cdot \frac{\sqrt{t_i^n - t_{i-1}^n}}{\min\{f(t_{i-1}^n), f(t_i^n)\} - M_n + g(\tau_n)} \\ &\leq \frac{1}{2}\sqrt{\lambda \log(n)/4} \cdot \frac{2}{\sqrt{\lambda \log(n)}} \\ &= \frac{1}{2} \end{aligned}$$

due to Lemma 2. This yields

$$\begin{aligned} \int_0^1 \frac{1}{(f(t) - M + g(\tau_n))^2} \,\mathrm{d}t &= \int_0^1 \frac{1}{(L_n(t) - M + g(\tau_n))^2 \cdot \left(1 + \frac{f(t) - L_n(t)}{L_n(t) - M + g(\tau_n)}\right)^2} \,\mathrm{d}t \\ &\geq \frac{4}{9} \cdot \int_0^1 \frac{1}{(L_n(t) - M + g(\tau_n))^2} \,\mathrm{d}t. \end{aligned}$$

□

For $\lambda \in [1, \infty[$, $n \geq 2$, and $C > 0$ we define

$$H_{C,n} = H_{\lambda,C,n} = \left\{ f \in F : \int_0^1 \frac{1}{(f(t) - M + g(\tau_n))^2} \,\mathrm{d}t \leq C \cdot (\log(1/g(\tau_n)))^4 \right\}. \tag{12}$$

Note that $H_{C_1,n} \subseteq H_{C_2,n}$, if $0 < C_1 \leq C_2$.

Corollary 1. *For all $\lambda \geq 1$, $n \geq 2$, $C > 0$, and $f \in F_n \cap J_n^+ \cap J_n^- \cap G_{1/2,n} \cap H_{C,n}$ we have*

$$\sum_{i=1}^n \rho_i^n \leq 9 \cdot C \cdot (\log(1/g(\tau_n)))^4 .$$

Proof. This follows directly from Lemma 4, Lemma 5, and (12). □

4.2. **Lower Bound on $\sum_{i=1}^n \rho_i^n$.** By $\lceil \cdot \rceil$ we denote the ceiling function, e.g., $\lceil 2 \rceil = 2$ and $\lceil 5/2 \rceil = 3$.

Proposition 1. *For all $\lambda \geq 1$, $n \geq 4$, and $f \in \bigcap_{k=\lceil n/2 \rceil}^n (G_{1,k} \cap F_k)$ we have*

$$\sum_{i=1}^n \rho_i^n \geq \frac{1}{288} \cdot \frac{n}{\lambda \log(1/\tau_n)}.$$

Proof. At iteration n, there are n subintervals

$$[t_0^n, t_1^n], [t_1^n, t_2^n], \dots, [t_{n-1}^n, t_n^n].$$

At first, we observe that at least $\lceil n/2 \rceil$ of these subintervals resulted from the iterations $\lceil n/2 \rceil, \dots, n-1$. Now suppose that such a subinterval, say $\left[t_{i-1}^n, t_i^n\right]$ for $i \in \{1, \dots, n\}$, resulted from the split of the interval $\left[t_{j_i-1}^{k_i}, t_{j_i}^{k_i}\right]$ at the k_i-th iteration of the algorithm, i.e.,

$$\rho^{k_i} = \rho_{j_i}^{k_i} \quad \wedge \quad t_{k_i+1} = t_{j_i}^{k_i+1} = \left(t_{j_i-1}^{k_i} + t_{j_i}^{k_i}\right)/2$$

with $k_i \in \{\lceil n/2 \rceil, \dots, n-1\}$ and $j_i \in \{1, \dots, k_i\}$. Thus we have

$$\left[t_{i-1}^n, t_i^n\right] = \left[t_{j_i-1}^{k_i}, t_{k_i+1}\right] = \left[t_{j_i-1}^{k_i+1}, t_{j_i}^{k_i+1}\right] \quad \vee \quad \left[t_{i-1}^n, t_i^n\right] = \left[t_{k_i+1}, t_{j_i}^{k_i}\right] = \left[t_{j_i}^{k_i+1}, t_{j_i+1}^{k_i+1}\right],$$

as depicted in Figure 3.

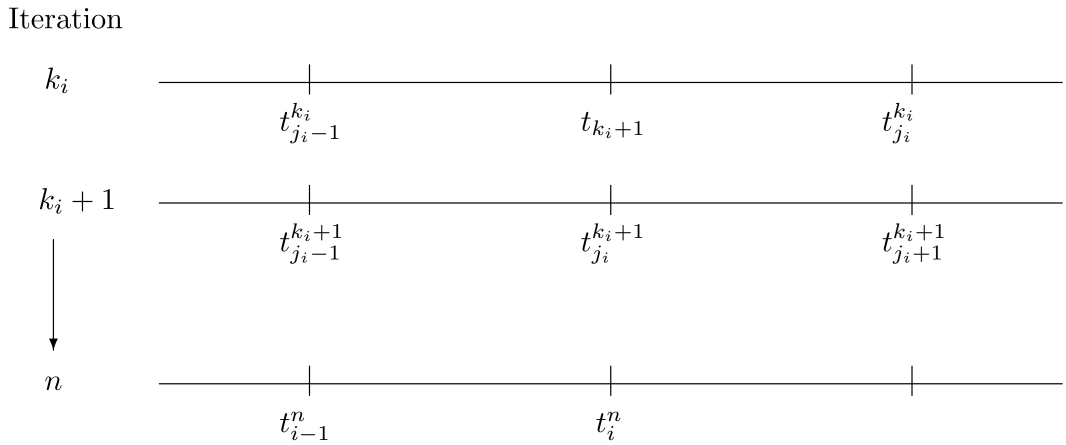

FIGURE 3. Situation from the proof of Proposition 1.

Without loss of generality we may consider the case where $\left[t_{i-1}^n, t_i^n\right] = \left[t_{j_i-1}^{k_i+1}, t_{j_i}^{k_i+1}\right]$ is given by the left child. First we show

$$\rho_{j_i}^{k_i+1} \geq \frac{1}{24}\,\rho^{k_i}. \tag{13}$$

For this we may without loss of generality assume that $f\left(t_{j_i-1}^{k_i}\right) \geq f\left(t_{j_i}^{k_i}\right)$ and hence $\rho_{j_i}^{k_i+1} \leq \rho_{j_i+1}^{k_i+1}$. We have

$$\begin{aligned}
\rho_{j_i}^{k_i+1} &= \frac{1}{2} \cdot \frac{t_{j_i}^{k_i} - t_{j_i-1}^{k_i}}{\left(f(t_{j_i-1}^{k_i}) - M_{k_i+1} + g(\tau_{k_i+1})\right)\left(f(t_{k_i+1}) - M_{k_i+1} + g(\tau_{k_i+1})\right)} \\
&= \frac{1}{2}\,\rho^{k_i} \cdot \frac{\left(f(t_{j_i-1}^{k_i}) - M_{k_i} + g(\tau_{k_i})\right)\left(f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})\right)}{\left(f(t_{j_i-1}^{k_i}) - M_{k_i+1} + g(\tau_{k_i+1})\right)\left(f(t_{k_i+1}) - M_{k_i+1} + g(\tau_{k_i+1})\right)} \\
&\geq \frac{1}{2}\,\rho^{k_i} \cdot \frac{\left(f(t_{j_i-1}^{k_i}) - M_{k_i} + g(\tau_{k_i})\right)\left(f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})\right)}{\left(f(t_{j_i-1}^{k_i}) - M_{k_i+1} + g(\tau_{k_i})\right)\left(f(t_{k_i+1}) - M_{k_i+1} + g(\tau_{k_i})\right)},
\end{aligned}$$

since g is non-decreasing on $\mathcal{A}$. Moreover, $f \in F_{k_i}$ and $f(t_{j_i-1}^{k_i}) \geq f(t_{j_i}^{k_i})$ imply

$$1 \leq \frac{f(t_{j_i-1}^{k_i}) - M_{k_i} + g(\tau_{k_i})}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})} = 1 + \frac{f(t_{j_i-1}^{k_i}) - f(t_{j_i}^{k_i})}{\sqrt{t_{j_i}^{k_i} - t_{j_i-1}^{k_i}}} \frac{\sqrt{t_{j_i}^{k_i} - t_{j_i-1}^{k_i}}}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})} \leq 2 \tag{14}$$

due to Lemma 2. Analogously, $f \in F_{k_i+1}$ yields

$$f(t_{k_i+1}) \leq f(t_{j_i-1}^{k_i}) + \sqrt{\lambda \log(k_i+1)/4} \cdot \sqrt{\left(t_{j_i}^{k_i} - t_{j_i-1}^{k_i}\right)/2}$$

and thus ($n \geq 4$ thus $k_i \geq 2$ and $\log(k_i + 1) \leq 2\log(k_i)$)

$$\frac{f(t_{k_i+1}) - M_{k_i} + g(\tau_{k_i})}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})} \tag{15}$$
$$\leq \frac{f(t_{j_i-1}^{k_i}) - M_{k_i} + g(\tau_{k_i})}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})} + \sqrt{\lambda \log(k_i)/4} \cdot \frac{\sqrt{t_{j_i}^{k_i} - t_{j_i-1}^{k_i}}}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})} \leq 3$$

due to (14) and Lemma 2. Furthermore, $M_{k_i} - M \leq g(\tau_{k_i})$ shows

$$0 \leq \frac{M_{k_i} - M_{k_i+1}}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})} \leq \frac{M_{k_i} - M}{g(\tau_{k_i})} \leq 1. \tag{16}$$

Combining (14), (15) and (16) yields

$$\frac{\left(f(t_{j_i-1}^{k_i}) - M_{k_i} + g(\tau_{k_i})\right)\left(f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})\right)}{\left(f(t_{j_i-1}^{k_i}) - M_{k_i+1} + g(\tau_{k_i})\right)\left(f(t_{k_i+1}) - M_{k_i+1} + g(\tau_{k_i})\right)}$$
$$= \frac{\frac{f(t_{j_i-1}^{k_i}) - M_{k_i} + g(\tau_{k_i})}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})}}{\left(\frac{f(t_{j_i-1}^{k_i}) - M_{k_i} + g(\tau_{k_i})}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})} + \frac{M_{k_i} - M_{k_i+1}}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})}\right)\left(\frac{f(t_{k_i+1}) - M_{k_i} + g(\tau_{k_i})}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})} + \frac{M_{k_i} - M_{k_i+1}}{f(t_{j_i}^{k_i}) - M_{k_i} + g(\tau_{k_i})}\right)}$$
$$\geq \frac{1}{(2+1)\cdot(3+1)}$$

and hence we get (13).

Now, we exploit

$$\rho_i^n = \rho_{j_i}^{k_i+1} \cdot \frac{\left(f(t_{j_i-1}^{k_i}) - M_{k_i+1} + g(\tau_{k_i+1})\right)\left(f(t_{k_i+1}) - M_{k_i+1} + g(\tau_{k_i+1})\right)}{\left(f(t_{j_i-1}^{k_i}) - M_n + g(\tau_n)\right)\left(f(t_{k_i+1}) - M_n + g(\tau_n)\right)}$$
$$\geq \rho_{j_i}^{k_i+1} \cdot \frac{\left(f(t_{j_i-1}^{k_i}) - M_{k_i+1} + g(\tau_{k_i+1})\right)}{\left(f(t_{j_i-1}^{k_i}) - M_{k_i+1} + g(\tau_{k_i+1}) + (M_{k_i+1} - M_n)\right)}$$
$$\cdot \frac{\left(f(t_{k_i+1}) - M_{k_i+1} + g(\tau_{k_i+1})\right)}{\left(f(t_{k_i+1}) - M_{k_i+1} + g(\tau_{k_i+1}) + (M_{k_i+1} - M_n)\right)}$$
$$\geq \rho_{j_i}^{k_i+1} \cdot \frac{1}{1+1} \cdot \frac{1}{1+1},$$

where the last inequality holds due to $M_{k_i+1} - M_n \leq M_{k_i+1} - M \leq g(\tau_{k_i+1})$, and use Lemma 3 to conclude that

$$\rho_i^n \geq \frac{1}{4}\,\rho_{j_i}^{k_i+1} \geq \frac{1}{96}\,\rho^{k_i} \geq \frac{1}{96} \cdot \frac{2}{3\lambda \log(1/\tau_n)}.$$

This shows

$$\sum_{i=1}^{n} \rho_i^n \geq \frac{n}{2} \cdot \frac{1}{96} \cdot \frac{2}{3\lambda \log(1/\tau_n)}.$$

□

4.3. **Main Deterministic Result.** For the following simple fact we omit the proof.

Lemma 6. *Let $C > 0$, $\lambda \geq 1$, $n \geq 2$, and $0 < \varepsilon \leq 1/n$ with*

$$\frac{n}{\log(1/\varepsilon)} \leq C \cdot (\log(1/g(\varepsilon)))^4 .$$

Then there exists a constant $\tilde{C} > 0$ that only depends on C and λ such that

$$g(\varepsilon) \leq \tilde{C} \cdot \exp\left(-1/\tilde{C} \cdot n^{1/5}\right).$$

For $n \geq 4$, $\lambda \in [1, \infty[$, and $C > 0$ we define

$$E_{C,n} = \left(F_n \cap J_n^+ \cap J_n^- \cap G_{1/2,n} \cap H_{C,n}\right) \cap \left(\bigcap_{k=\lceil n/2 \rceil}^{n} (G_{1,k} \cap F_k)\right). \tag{17}$$

Note that $E_{C_1,n} \subseteq E_{C_2,n}$, if $0 < C_1 \leq C_2$.

Corollary 2. *Let $\lambda \geq 1$, $n \geq 4$, $C > 0$, and $f \in E_{C,n}$. Then there exists a constant $\tilde{C} > 0$ that only depends on C and λ such that*

$$\Delta_n \leq g(\tau_n) \leq \tilde{C} \cdot \exp\left(-1/\tilde{C} \cdot n^{1/5}\right).$$

Proof. The first inequality holds by definition of $G_{1,n}$. The second inequality is an immediate consequence of Corollary 1, Proposition 1, and Lemma 6. □

5. Probabilistic Arguments

In the previous section we studied the application of the optimization algorithm to an element $f \in F$. In particular, Corollary 2 provides an exponentially small error bound for functions f belonging to subsets $E_{C,n}$ of F. In this section we consider the special case of a Brownian motion $W = (W(t))_{0 \leq t \leq 1}$ and show that the probability of a Brownian path belonging to $E_{C,n}$ tends to 1 at an arbitrarily high polynomial rate, see Corollary 3. It turns out that this probability bound depends on the parameter λ.

Let us stress that all quantities defined in Section 2 (e.g., $M_n, M, \tau_n, \ldots$) are now understood to depend on W instead of f. Hence these quantities are random. Furthermore, for a set of functions $A \subseteq F$ (e.g., $F_n, G_{1/2,n}, \ldots$) we write $\mathrm{P}(A)$ instead of $\mathrm{P}(W \in A)$.

5.1. **Lower Bound for $\mathrm{P}(F_n)$.** The following basic result is well-known, for completeness we add a proof.

Lemma 7. *Let $n \in \mathbb{N}$ and $Z_1, \ldots, Z_n$ be identically distributed with $Z_1 \sim \mathcal{N}(0,1)$. Then we have*

$$\mathrm{P}\left(\max_{1 \leq k \leq n} |Z_k| \leq t\right) \geq 1 - n \cdot \frac{2}{\sqrt{2\pi}} \cdot \frac{1}{t} \cdot \exp(-t^2/2)$$

for all $t > 0$.

Proof. We clearly have

$$\mathrm{P}\left(\max_{1 \leq k \leq n} |Z_k| > t\right) \leq \sum_{k=1}^{n} \mathrm{P}(|Z_k| > t) = n \cdot \mathrm{P}(|Z_1| > t)$$

for $t \in \mathbb{R}$. Combining this with the inequality

$$\mathrm{P}(|Z_1| > t) \leq \frac{2}{\sqrt{2\pi}} \cdot \frac{1}{t} \cdot \exp(-t^2/2)$$

for $t > 0$ yields the claim. □

Lemma 8. *We have*

$$\mathrm{P}(F_n) \geq 1 - 7 \cdot n^{1-\lambda/72}$$

for all $\lambda \geq 1$ and for all $n \geq 2$.

Proof. For $n \geq 2$ we denote by $i_n \in \{1, \dots, n\}$ the index of the interval which will be split in step n, i.e., with $\rho^n_{i_n} = \rho^n$ (note that i_n is random). Moreover, we set $i_1 = 1$. Note that

$$t^{n+1}_{i_n-1} = t^n_{i_n-1}, \qquad t^{n+1}_{i_n+1} = t^n_{i_n}, \qquad t^{n+1}_{i_n} = t_{n+1} = (t^n_{i_n-1} + t^n_{i_n})/2,$$

and

$$t^{n+1}_{i_n} - t^{n+1}_{i_n-1} = t^{n+1}_{i_n+1} - t^{n+1}_{i_n} = (t^n_{i_n} - t^n_{i_n-1})/2.$$

Furthermore, we define $X_1 = \frac{W(t^1_1)-W(t^1_0)}{\sqrt{t^1_1-t^1_0}} = W(1)$ and

$$X_{2n} = \frac{W(t^{n+1}_{i_n}) - W(t^{n+1}_{i_n-1})}{\sqrt{t^{n+1}_{i_n} - t^{n+1}_{i_n-1}}}, \qquad X_{2n+1} = \frac{W(t^{n+1}_{i_n+1}) - W(t^{n+1}_{i_n})}{\sqrt{t^{n+1}_{i_n+1} - t^{n+1}_{i_n}}}$$

for $n \geq 1$. Let us stress that that

$$\{X_1, \dots, X_{2n-1}\} = \left\{ \frac{W(t^k_i) - W(t^k_{i-1})}{\sqrt{t^k_i - t^k_{i-1}}} \,:\, 1 \leq i \leq k \leq n \right\}$$

for $n \geq 1$ and thus

(18) $$\max_{1 \leq k \leq 2n-1} |X_k| = \max_{1 \leq k \leq n} \max_{1 \leq i \leq k} \frac{\left|W(t^k_i) - W(t^k_{i-1})\right|}{\sqrt{t^k_i - t^k_{i-1}}}.$$

For $n \geq 1$ we define

$$Y_n = \frac{W(t^n_{i_n}) - W(t^n_{i_n-1})}{\sqrt{t^n_{i_n} - t^n_{i_n-1}}} = \frac{X_{2n} + X_{2n+1}}{\sqrt{2}}.$$

Note that for every $n \geq 1$ there exists a random index $j_n \in \{2(n-1), 2(n-1)+1\}$ with $Y_n = X_{j_n}$ where we use the convention $X_0 = X_1$. This yields

(19) $$|Y_n| \leq \max\left(\left|X_{2(n-1)}\right|, \left|X_{2(n-1)+1}\right|\right).$$

Finally, we define

$$Z_n = \frac{X_{2n} - X_{2n+1}}{\sqrt{2}}$$

for $n \geq 1$. Since

$$Z_n = \frac{2W(t^{n+1}_{i_n}) - \left(W(t^{n+1}_{i_n-1}) + W(t^{n+1}_{i_n+1})\right)}{\sqrt{t^n_{i_n} - t^n_{i_n-1}}} = \frac{W\left(\frac{t^n_{i_n} + t^n_{i_n-1}}{2}\right) - \frac{W(t^n_{i_n-1}) + W(t^n_{i_n})}{2}}{\sqrt{(t^n_{i_n} - t^n_{i_n-1})/4}},$$

we have $Z_n \sim \mathcal{N}(0,1)$. Furthermore, note that

$$X_{2n} = \frac{Y_n + Z_n}{\sqrt{2}}, \qquad X_{2n+1} = \frac{Y_n - Z_n}{\sqrt{2}}.$$

Hence we get

$$\max\left(|X_{2n}|, |X_{2n+1}|\right) \le \frac{|Y_n| + |Z_n|}{\sqrt{2}} \le \frac{\max\left(\left|X_{2(n-1)}\right|, \left|X_{2(n-1)+1}\right|\right) + |Z_n|}{\sqrt{2}}$$

for $n \ge 1$ due to (19). Combining this with the inequality

$$\frac{\frac{a}{\sqrt{2}-1} + b}{\sqrt{2}} \le \frac{\max(a,b)}{\sqrt{2}-1}$$

for $a, b \in \mathbb{R}$, we obtain by induction

$$\max_{1 \le k \le 2n-1} |X_k| \le \frac{\max\left(|X_1|, |Z_1|, \dots, |Z_{n-1}|\right)}{\sqrt{2}-1} \tag{20}$$

for all $n \ge 1$. Finally, combining (18), (20), and Lemma 7 yields

$$\begin{aligned}
\mathrm{P}(F_n) &= \mathrm{P}\left(\max_{1 \le k \le 2n-1} |X_k| \le \sqrt{\lambda \log(n)/4}\right) \\
&\ge \mathrm{P}\left(\max\left(|X_1|, |Z_1|, \dots, |Z_{n-1}|\right) \le (\sqrt{2}-1) \cdot \sqrt{\lambda \log(n)/4}\right) \\
&\ge \mathrm{P}\left(\max\left(|X_1|, |Z_1|, \dots, |Z_{n-1}|\right) \le \sqrt{\lambda \log(n)/36}\right) \\
&\ge 1 - n \cdot \frac{2}{\sqrt{2\pi}} \cdot 8 \cdot n^{-\lambda/72} \\
&\ge 1 - 7 \cdot n^{1-\lambda/72}
\end{aligned}$$

for $\lambda \ge 1$ and $n \ge 2$. □

Remark 4. Let us comment on the distribution of the random variables $X_1, X_2, \dots$ defined in the proof of Lemma 8. Obviously, the random variables X_1, X_2, X_3 are standard normally distributed and jointly Gaussian. In contrast to that, the random variables X_1, X_2, X_3, X_4, X_5 are not jointly Gaussian, but still X_4 and X_5 are standard normally distributed. However, computer simulations strongly suggest that X_6 is not standard normally distributed. Since the evaluation points $t_0, t_1, \dots$ are computed adaptively, we conjecture that X_n is not standard normally distributed for all $n \ge 6$.

5.2. Lower Bound for $\mathrm{P}\left(G_{1/2,n}\right)$ and $\mathrm{P}\left(G_{1,n}\right)$.

Lemma 9. *We have*

$$\mathrm{P}\left(G_{1,n}\right) \ge \mathrm{P}\left(G_{1/2,n}\right) \ge 1 - 8 \cdot n^{1-\lambda/72}$$

for all $\lambda \ge 1$ and for all $n \ge 2$.

Proof. For $n \in \mathbb{N}$ we denote by

$$\mathfrak{A}_n = \sigma(W(t_1), \dots, W(t_n)) = \sigma(t_1, W(t_1), \dots, t_n, W(t_n))$$

the σ-algebra generated by $(W(t_1), \dots, W(t_n))$. Note that 1_{F_n} is measurable w.r.t. $\mathfrak{A}_n$ for all $\lambda \ge 1$ and $n \ge 2$.

Conditional on $\mathfrak{A}_n$, the minimizers over all subintervals $[t_0^n, t_1^n], \dots, [t_{n-1}^n, t_n^n]$ are independent with distribution (independent Brownian bridges)

$$\mathrm{P}\left(\min_{t_{i-1}^n \le s \le t_i^n} W(s) < y\right) = \exp\left(-\frac{2}{t_i^n - t_{i-1}^n}\left(W(t_{i-1}^n) - y\right)\left(W(t_i^n) - y\right)\right)$$

for $y < \min(W(t^n_{i-1}), W(t^n_i))$, see [2, IV.4, p. 67] or [18]. For $\beta \in [0,1]$, we hence get

$$\begin{aligned}
&\mathrm{P}(\Delta_n \leq \beta g(\tau_n) \mid \mathfrak{A}_n) \\
&\quad = \prod_{i=1}^n \left(1 - \exp\left(-\frac{2}{t^n_i - t^n_{i-1}} \left(W(t^n_{i-1}) - M_n + \beta g(\tau_n)\right)\left(W(t^n_i) - M_n + \beta g(\tau_n)\right)\right)\right) \\
&\quad \geq \prod_{i=1}^n \left(1 - \exp\left(-2\beta^2/\rho^n_i\right)\right) \\
&\quad \geq \left(1 - \exp\left(-2\beta^2/\rho^n\right)\right)^n .
\end{aligned}$$

Then, Lemma 1 implies

$$\begin{aligned}
\mathrm{P}(\Delta_n \leq \beta g(\tau_n) \mid \mathfrak{A}_n) &\geq \left(1 - \exp\left(-\beta^2 \lambda \log(n)\right)\right)^n \\
&\geq 1 - n^{1-\beta^2\lambda}
\end{aligned}$$

on F_n. Setting $B_n = \{\Delta_n \leq \beta g(\tau_n)\}$, we thus obtain

$$\mathrm{P}(B_n) \geq \mathrm{E}\left(\mathrm{E}\left(1_{B_n \cap F_n} \mid \mathfrak{A}_n\right)\right) = \mathrm{E}\left(1_{F_n} \cdot \mathrm{E}\left(1_{B_n} \mid \mathfrak{A}_n\right)\right) \geq \mathrm{P}(F_n) \cdot \left(1 - n^{1-\beta^2\lambda}\right).$$

Set $\beta = 1/2$. Finally, Lemma 8 shows

$$\mathrm{P}(B_n) \geq \left(1 - 7 \cdot n^{1-\lambda/72}\right)^+ \cdot \left(1 - n^{1-\lambda/4}\right)^+ \geq 1 - 8 \cdot n^{1-\lambda/72}$$

for $\lambda \geq 1$ and $n \geq 2$. □

5.3. **Lower Bound for** $\mathrm{P}(J^+_n)$ **and** $\mathrm{P}(J^-_n)$**.**

Lemma 10. *We have*

$$\mathrm{P}(J^+_n) = \mathrm{P}(J^-_n) \geq 1 - n^{1-\lambda/8}$$

for all $\lambda \geq 1$ *and for all* $n \geq 2$.

Proof. Let

$$Y_i = \max_{t^n_{i-1} \leq s \leq t^n_i} \frac{W(s) - L_n(s)}{\sqrt{t^n_i - t^n_{i-1}}},$$

which is the maximum of a standard Brownian bridge, and thus $\mathrm{P}(Y_i > y) = \exp(-2y^2)$ for $y \geq 0$, see [2, IV.4, p. 67] or [18]. Moreover, the family $(Y_1, \ldots, Y_n)$ is independent and so

$$\mathrm{P}\left(\max_{1 \leq i \leq n} Y_i \leq \frac{1}{2}\sqrt{\lambda \log(n)/4}\right) = (1 - \exp(-\lambda \log(n)/8))^n \geq 1 - n^{1-\lambda/8}.$$

By symmetry, we obtain the same bound for J^-_n. □

5.4. **Lower Bound for** $\mathrm{P}(H_{C,n})$**.** For $T > 0$ and $z \geq 0$ let $R^{T,z} = (R^{T,z}(t))_{0 \leq t \leq T}$ denote a 3-dimensional Bessel bridge from 0 to z on $[0,T]$, that is a 3-dimensional Bessel process started at 0 conditioned to have value z at time T. In other words, for independent Brownian bridges B^T_1, B^T_2, and B^T_3 from 0 to 0 on $[0,T]$, see, e.g., [15, p. 274], we have

$$(R^{T,z}(t))_{0 \leq t \leq T} \overset{\mathrm{d}}{=} \left(\sqrt{\left(\frac{z \cdot t}{T} + B^T_1(t)\right)^2 + (B^T_2(t))^2 + (B^T_3(t))^2}\right)_{0 \leq t \leq T}, \tag{21}$$

where $\overset{\mathrm{d}}{=}$ denotes equality in distribution. A consequence of (21) is the following scaling property

$$(R^{T,z}(t))_{0\leq t\leq T} \overset{\mathrm{d}}{=} \left(\frac{1}{\sqrt{c}}\, R^{c\cdot T,\sqrt{c}\cdot z}(c\cdot t)\right)_{0\leq t\leq T} \tag{22}$$

for all $c > 0$. Moreover, if $z_1 \leq z_2$, there exist 3-dimensional Bessel bridges R^{T,z_1} and R^{T,z_2} (on a common probability space) such that

$$R^{T,z_1}(t) \leq R^{T,z_2}(t) \tag{23}$$

for all $0 \leq t \leq T$. We refer to [16, Chap. XI] for a detailed discussion of Bessel processes and Bessel bridges.

Lemma 11. *For all $r \geq 1$ there exists a constant $C > 0$ such that for all $0 < T \leq 1$ and $z \geq 0$ we have*

$$\mathrm{P}\left(\int_0^T \frac{1}{(R^{T,z}(t))^2+\varepsilon}\,\mathrm{d}t \geq C\cdot(\log(1/\varepsilon))^4\right) \leq C\cdot\varepsilon^r$$

for all $\varepsilon > 0$.

Proof. We may assume $T = 1$ and $z = 0$ due to (22) and (23), respectively. In this case

$$R^{1,0} = B^{\mathrm{ex}} = (B^{\mathrm{ex}}(t))_{0\leq t\leq 1}$$

is a Brownian excursion of length 1, see, e.g., [15, Lem. 15] or [20].

Let $B = (B(t))_{t\geq 0}$ be a Brownian motion. Consider the stochastic differential equation

$$\mathrm{d}X(t) = \left(4 - \frac{(X(t))^2}{1-\int_0^t X(s)\,\mathrm{d}s}\right)\mathrm{d}t + 2\sqrt{X(t)}\,\mathrm{d}B(t), \qquad X(0) = 0,$$

to be solved on $[0, V(X)[$, where $V(X) = \inf\{t \geq 0 : \int_0^t X(s)\,\mathrm{d}s = 1\}$ and $X(t) = 0$ for $t \geq V(X)$. For properties of this SDE and its solution, see [15]. In particular, there it is shown that this SDE has a unique continuous nonnegative strong solution $X = (X(t))_{t\geq 0}$. Moreover, this solution satisfies

$$(L_1(x))_{x\geq 0} \overset{\mathrm{d}}{=} (X(t))_{t\geq 0}, \tag{24}$$

where $L_1 = (L_1(x))_{x\in\mathbb{R}}$ denotes the local time of B^{ex} up to time $t = 1$. More precisely, L_1 is the continuous density with respect to the Lebesgue measure on $\mathbb{R}$ of the push-forward measure of the Lebesgue measure on $[0, 1]$ under the mapping B^{ex}, i.e.,

$$\int_0^1 h(B^{\mathrm{ex}}(t))\,\mathrm{d}t = \int_{-\infty}^{\infty} h(x)\cdot L_1(x)\,\mathrm{d}x \tag{25}$$

for all nonnegative Borel measurable $h\colon \mathbb{R} \to \mathbb{R}$. Note that $L_1(x) = 0$ for $x \leq 0$.

Consider the stochastic differential equation

$$\mathrm{d}Y(t) = 4\,\mathrm{d}t + 2\sqrt{Y(t)}\,\mathrm{d}B(t), \qquad Y(0) = 0.$$

It is known that this SDE has a unique strong solution $Y = (Y(t))_{t\geq 0}$, which is a 4-dimensional squared Bessel process started at 0, i.e.,

$$(Y(t))_{t\geq 0} \overset{\mathrm{d}}{=} \left(\sum_{k=1}^4 (W_k(t))^2\right)_{t\geq 0} \tag{26}$$

with independent Brownian motions $W_k = (W_k(t))_{t\geq 0}$ for $k = 1,\dots,4$, see, e.g., [16, Chap. XI]. Using a slight modification of the comparison principle [9, Prop. V.2.18] we obtain

$$X(t) \leq Y(t) \tag{27}$$

for $t \geq 0$. Combining (25), (24), (27), and (26) yields

$$\begin{aligned}
\int_0^1 \frac{1}{(B^{\mathrm{ex}}(t))^2+\varepsilon}\,\mathrm{d}t &\leq 1 + \int_0^1 \frac{1\,(B^{\mathrm{ex}}(t)\leq 1)}{(B^{\mathrm{ex}}(t))^2+\varepsilon}\,\mathrm{d}t \\
&= 1 + \int_0^1 \frac{1}{x^2+\varepsilon}\cdot L_1(x)\,\mathrm{d}x \\
&\stackrel{\mathrm{d}}{=} 1 + \int_0^1 \frac{1}{t^2+\varepsilon}\cdot X(t)\,\mathrm{d}t \\
&\leq 1 + \int_0^1 \frac{1}{t^2+\varepsilon}\cdot Y(t)\,\mathrm{d}t \\
&\stackrel{\mathrm{d}}{=} 1 + \int_0^1 \left(\sum_{k=1}^4 \frac{(W_k(t))^2}{t^2+\varepsilon}\right)\mathrm{d}t
\end{aligned} \tag{28}$$

for all $\varepsilon > 0$.

Consider the Gaussian random function $\xi = (\xi(t))_{0\leq t\leq 1}$ given by $\xi(0) = 1$ and

$$\xi(t) = \frac{W(t)}{\sqrt{t}\cdot\sqrt{\log\left(1+\frac{1}{t}\right)}}$$

for $0 < t \leq 1$, which is bounded due to the law of the iterated logarithm. Using [11, Thm. 12.1] we get the existence of constants $C > 0$ and $a > 1$ such that

$$\mathrm{P}\left(\sup_{0\leq t\leq 1}|\xi(t)| > q\right) \leq \frac{1}{a^{q^2}} \tag{29}$$

for all $q \geq C$. A small computation shows that there exist positive constants $C_1, C_2 > 0$ such that

$$\int_0^1 \frac{t\cdot\log\left(1+\frac{1}{t}\right)}{t^2+\varepsilon}\,\mathrm{d}t \leq C_1\cdot(\log(1/\varepsilon))^2$$

for all $0 < \varepsilon < C_2$, and consequently

$$\int_0^1 \frac{(W(t))^2}{t^2+\varepsilon}\,\mathrm{d}t = \int_0^1 (\xi(t))^2\cdot\frac{t\cdot\log\left(1+\frac{1}{t}\right)}{t^2+\varepsilon}\,\mathrm{d}t \leq C_1\cdot\left(\sup_{0\leq t\leq 1}|\xi(t)|\right)^2\cdot(\log(1/\varepsilon))^2.$$

Let $r \geq 1$. Using (29) with $q = \log(1/\varepsilon)$, there exists a constant $\tilde{C}_2 > 0$ such that

$$\mathrm{P}\left(\int_0^1 \frac{(W(t))^2}{t^2+\varepsilon}\,\mathrm{d}t \leq C_1\cdot(\log(1/\varepsilon))^4\right) \geq 1-\varepsilon^r$$

for all $0 < \varepsilon < \tilde{C}_2$, and hence

$$\mathrm{P}\left(\int_0^1 \left(\sum_{k=1}^4 \frac{(W_k(t))^2}{t^2+\varepsilon}\right)\mathrm{d}t \leq 4C_1\cdot(\log(1/\varepsilon))^4\right) \geq 1-4\cdot\varepsilon^r \tag{30}$$

for all $0 < \varepsilon < \tilde{C}_2$. Combining (28) and (30) completes the proof. □

Lemma 12. *For all $r \geq 1$ there exists a constant $C > 0$ such that*

$$\mathrm{P}\left(\int_0^1 \frac{1}{(W(t)-M)^2+\varepsilon}\,\mathrm{d}t \geq C \cdot (\log(1/\varepsilon))^4\right) \leq C \cdot \varepsilon^r$$

for all $\varepsilon > 0$.

Proof. Let $t^* \in \,]0,1[$ be the (a.s. unique) minimizer of the Brownian motion W, i.e., $M = W(t^*)$. Conditionally on $t^* = s$, $M = m$, and $W(1) = w$, the process

$$(W(s-t) - m)_{0 \leq t \leq s}$$

is a 3-dimensional Bessel bridge from $(0,0)$ to $(s,-m)$, and the process

$$(W(s+t) - m)_{0 \leq t \leq 1-s}$$

is a 3-dimensional Bessel bridge from $(0,0)$ to $(1-s, w-m)$, see, e.g., [1, Prop. 2]. Let $r \geq 1$ and $C > 0$ be according to Lemma 11. Then we have

$$\begin{aligned}
&\mathrm{P}\left(\int_0^1 \frac{1}{(W(t)-M)^2+\varepsilon}\,\mathrm{d}t \geq C \cdot (\log(1/\varepsilon))^4 \mid t^* = s,\ M = m,\ W(1) = w\right)\\
&\leq \mathrm{P}\left(\int_0^s \frac{1}{(W(t)-m)^2+\varepsilon}\,\mathrm{d}t \geq \frac{C}{2} \cdot (\log(1/\varepsilon))^4 \mid t^* = s,\ M = m,\ W(1) = w\right)\\
&\qquad + \mathrm{P}\left(\int_s^1 \frac{1}{(W(t)-m)^2+\varepsilon}\,\mathrm{d}t \geq \frac{C}{2} \cdot (\log(1/\varepsilon))^4 \mid t^* = s,\ M = m,\ W(1) = w\right)\\
&\leq \frac{C}{2} \cdot \varepsilon^r + \frac{C}{2} \cdot \varepsilon^r
\end{aligned}$$

by Lemma 11, which establishes the claim. □

Lemma 13. *For all $r \geq 1$ and for all $\lambda \geq 1$ there exists a constant $C > 0$ such that*

$$\mathrm{P}\left(H_{C,n}\right) \geq 1 - C \cdot n^{-r}$$

for all $n \geq 2$.

Proof. Fix $\lambda \geq 1$. Since $\tau_n \in \mathcal{A}$ and $\tau_n \geq 1/2^{n-1}$, we get that $(g(\tau_n))^2$ takes at most n different values, which we denote by A_n (note that A_n depends on λ). Hence we get

$$\begin{aligned}
\mathrm{P}(H^c_{C,n}) &= \mathrm{P}\left(\int_0^1 \frac{1}{(W(t)-M+g(\tau_n))^2}\,\mathrm{d}t \geq C \cdot (\log(1/g(\tau_n)))^4\right)\\
&\leq \mathrm{P}\left(\int_0^1 \frac{1}{(W(t)-M)^2+g(\tau_n)^2}\,\mathrm{d}t \geq C \cdot (\log(1/g(\tau_n)))^4\right)\\
&= \sum_{\varepsilon \in A_n} \mathrm{P}\left(\int_0^1 \frac{1}{(W(t)-M)^2+\varepsilon}\,\mathrm{d}t \geq \frac{C}{16} \cdot (\log(1/\varepsilon))^4,\ (g(\tau_n))^2 = \varepsilon\right).
\end{aligned}$$

Moreover, due to $\tau_n \leq 1/n$ we have

$$(g(\tau_n))^2 \leq \lambda \log(n)/n,$$

and thus there exists $N \geq 2$ such that $g(\tau_n)^2 \leq 1/\sqrt{n}$ for all $n \geq N$. Now, let $r \geq 1$ and $C/16 > 0$ be according to Lemma 12. Then we get

$$\mathrm{P}(H^c_{C,n}) \leq n \cdot \frac{C}{16} \cdot (1/\sqrt{n})^r = \frac{C}{16} \cdot n^{1-\frac{r}{2}}$$

for all $n \geq N$. □

5.5. **Main Probabilistic Result.**

Corollary 3. *For all $r \geq 1$ and all $\lambda \geq 72 \cdot (2+r)$ there exists a constant $C > 0$ such that*

$$\mathrm{P}\left(E_{C,n}\right) \geq 1 - C \cdot n^{-r}$$

for all $n \geq 4$.

Proof. Let $r \geq 1$, $\lambda \geq 72 \cdot (2+r) \geq 1$, and $C > 0$ be according to Lemma 13. Combining Lemma 8, Lemma 9, Lemma 10, and Lemma 13 yields

$$\begin{aligned}
\mathrm{P}(E_{C,n}^c) &\leq \big(\mathrm{P}(F_n^c) + \mathrm{P}(J_n^{+c}) + \mathrm{P}(J_n^{-c}) + \mathrm{P}(G_{1/2,n}^c) + \mathrm{P}(H_{C,n}^c)\big) \\
&\quad + \left(\sum_{k=\lceil n/2\rceil}^{n} \mathrm{P}(G_{1,k}^c) + \sum_{k=\lceil n/2\rceil}^{n} \mathrm{P}(F_k^c)\right) \\
&\leq \big(7n^{-r} + n^{-r} + n^{-r} + 8n^{-r} + Cn^{-r}\big) \\
&\quad + \big(n \cdot 8(n/2)^{1-\lambda/72} + n \cdot 7(n/2)^{1-\lambda/72}\big)
\end{aligned}$$

for all $n \geq 4$. □

6. Proof of Theorem 1

Let $r \geq 1$ and $p \geq 1$. Moreover, we fix

$$\lambda \geq 144 \cdot (1 + p \cdot r) = 72 \cdot (2 + 2pr).$$

According to Corollary 3 there exists a constant $C > 0$ such that

$$\mathrm{P}\left(E_{C,n}\right) \geq 1 - C \cdot n^{-2pr}$$

for all $n \geq 4$. Furthermore, due to Corollary 2 there exists a constant $\tilde{C} > 0$ such that for all $n \geq 4$

$$\Delta_n = \Delta_{n,\lambda}(W) \leq \tilde{C} \cdot \exp(-1/\tilde{C} \cdot n^{1/5})$$

if $W \in E_{C,n}$. Noting that

$$\Delta_n \leq -\inf_{0 \leq t \leq 1} W(t) \overset{\mathrm{d}}{=} |Z|$$

with $Z \sim \mathcal{N}(0,1)$, we obtain using the Cauchy-Schwarz inequality

$$\begin{aligned}
\mathrm{E}\left(|\Delta_n|^p\right) &= \mathrm{E}\left(|\Delta_n|^p \cdot 1_{E_{C,n}}\right) + \mathrm{E}\left(|\Delta_n|^p \cdot 1_{E_{C,n}^c}\right) \\
&\leq \tilde{C}^p \cdot \exp(-p/\tilde{C} \cdot n^{1/5}) + \sqrt{\mathrm{E}\left(|\Delta_n|^{2p}\right)} \cdot \sqrt{1 - \mathrm{P}(E_{C,n})} \\
&\leq \tilde{C}^p \cdot \exp(-p/\tilde{C} \cdot n^{1/5}) + \sqrt{\mathrm{E}\left(|Z|^{2p}\right)} \cdot \sqrt{C} \cdot n^{-pr},
\end{aligned}$$

for all $n \geq 4$. This completes the proof of Theorem 1.

Acknowledgement

We thank Klaus Ritter for valuable discussions and comments.

References

[1] Søren Asmussen, Peter Glynn, and Jim Pitman. Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. Appl. Probab.*, 5(4):875–896, 1995.

[2] Andrei N. Borodin and Paavo Salminen. *Handbook of Brownian motion – facts and formulae*. Probability and its Applications. Birkhäuser Verlag, Basel, second edition, 2002.

[3] James M. Calvin. Average performance of a class of adaptive algorithms for global optimization. *Ann. Appl. Probab.*, 7(3):711–730, 1997.

[4] James M. Calvin. A one-dimensional optimization algorithm and its convergence rate under the Wiener measure. *J. Complexity*, 17(2):306–344, 2001.

[5] James M. Calvin. A lower bound on complexity of optimization on the Wiener space. *Theoretical Computer Science*, 383:132–139, 2007.

[6] James M. Calvin. An adaptive univariate global optimization algorithm and its convergence rate under the Wiener measure. *Informatica (Vilnius)*, 22(4):471–488, 2011.

[7] J. Michael Harrison. *Brownian motion and stochastic flow systems*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1985.

[8] Mario Hefter and André Herzwurm. Optimal strong approximation of the one-dimensional squared Bessel process. *ArXiv e-prints*, 2016.

[9] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[10] Harold J. Kushner. A versatile stochastic model of a function of unknown and time varying form. *J. Math. Anal. Appl.*, 5:150–167, 1962.

[11] Mikhail A. Lifshits. *Gaussian random functions*, volume 322 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 1995.

[12] Jonas Mockus. Bayesian methods for extremum search. *Avtomat. i Vyčisl. Tehn.*, (3):53–62, 1972.

[13] Jonas Mockus. On Bayesian methods for seeking the extremum. In *Optimization Techniques, IFIP Technical Conference, Novosibirsk, USSR, July 1-7, 1974*, pages 400–404, 1974.

[14] Erich Novak. *Deterministic and stochastic error bounds in numerical analysis*, volume 1349 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1988.

[15] Jim Pitman. The SDE solved by local times of a Brownian excursion or bridge derived from the height profile of a random tree or forest. *Ann. Probab.*, 27(1):261–283, 1999.

[16] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, third edition, 1999.

[17] Klaus Ritter. Approximation and optimization on the Wiener space. *J. Complexity*, 6(4):337–364, 1990.

[18] Lawrence A. Shepp. The joint density of the maximum and its location for a Wiener process with drift. *J. Appl. Probab.*, 16(2):423–427, 1979.

[19] Aimo Törn and Antanas Žilinskas. *Global Optimization*. Springer, Berlin, 1989.

[20] David Williams. Decomposing the Brownian path. *Bull. Amer. Math. Soc.*, 76: 871–873, 1970.

[21] Antanas Žilinskas. Axiomatic characterization of a global optimization algorithm and investigation of its search strategy. *Oper. Res. Lett.*, 4(1):35–39, 1985.

Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102-1982, USA

E-mail address: calvin@njit.edu

Department of Mathematics, Technische Universität Kaiserslautern, 67653 Kaiserslautern, Germany

E-mail address: hefter@mathematik.uni-kl.de

Department of Mathematics, Technische Universität Kaiserslautern, 67653 Kaiserslautern, Germany

E-mail address: herzwurm@mathematik.uni-kl.de