J Glob Optim (2010) 48:17–27
DOI 10.1007/s10898-010-9530-z

A lower bound on convergence rates of nonadaptive algorithms for univariate optimization with noise

James M. Calvin

Received: 10 January 2010 / Accepted: 29 January 2010 / Published online: 14 February 2010

Abstract This paper considers complexity bounds for the problem of approximating the global minimum of a univariate function when the function evaluations are corrupted by random noise. We take an average-case point of view, where the objective function is taken to be a sample function of a Wiener process and the noise is independent Gaussian. Previous papers have bounded the convergence rates of some nonadaptive algorithms. We establish a lower bound on the convergence rate of any nonadaptive algorithm.

Keywords Global optimization · Wiener process · Noisy information

1 Introduction

Stochastic models of objective functions have been used to motivate the construction of efficient global optimization algorithms (see [6,12,13,15]), and also as a basis for studying average-case complexity bounds [16]. These models are well-suited for inclusion of random noise in the function evaluations; early studies of different variations of this problem appeared in [7,11]. In both cases the error must be analyzed probabilistically, and it is interesting to compare the magnitude of the error for the cases of exact function evaluation and evaluation with random error.

The most common (and tractable) probability model for objective functions is the Wiener process model; see [6,14]. In the case where the function can be evaluated exactly, it was proved in [10] that any nonadaptive algorithm has expected error of order at least $n^{-1/2}$ after n function evaluations. It is much more difficult to optimize when the function evaluations are corrupted by noise. The most common model is white noise; that is, the function value is observed with Gaussian noise that is independent for different evaluations. In [5] it was shown that with equispaced observations the expected error is of order $n^{-1/4}$ when the function values are corrupted by independent Gaussian noise. A natural question is whether any

J. M. Calvin (✉)
Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102-1982, USA
e-mail: calvin@njit.edu

nonadaptive algorithm can have a better convergence rate. In the noisy case considered in this paper, no method (adaptive or nonadaptive) can achieve a convergence rate better than $n^{-1/2}$. This is because even if the location of the minimizer is known to the searcher and all observations are made at that point, with independent noise the central limit theorem implies a root mean-squared error of order $n^{-1/2}$ just to estimate the function value at that point. Therefore, $n^{-1/2}$ is a lower bound for all algorithms.

In this paper, we show that there is a positive constant c such that any nonadaptive algorithm has root mean-squared error $\geq c\, n^{-1/4}$ after n function evaluations corrupted by independent Gaussian noise.

In the next section, we describe the problem. In Sect. 3, we summarize some background on the conditional distribution of the Wiener process given a set of noise-corrupted evaluations. The main result is stated and proved in Sect. 4.

2 Problem description

Assume that f is an element of the class $\mathcal{C}_0 = \mathcal{C}_0([0, 1])$ of continuous real-valued functions f on [0, 1], with $f(0) = 0$. We consider the problem of approximating the global minimum of f using a fixed number of evaluations of the function corrupted by random error. We adopt a Gaussian model both for the error and for the function f. We take f to be a sample path of a Wiener process. Let P denote the Wiener measure on $\mathcal{C}_0$, which is the centered Gaussian measure with covariance function

$$R(s, t) \equiv \int_{\mathcal{C}_0} f(s) f(t)\, \mathrm{d}P(f) = \min(s, t).$$

If $\{f(t); 0 \leq t \leq 1\}$ is a Wiener process, then the increments are independent and $f(t) - f(s)$ has a normal distribution with mean zero and variance $t - s$. With probability one, f is nowhere differentiable and the set of local minimizers is dense in [0, 1], so this model is particularly challenging from the point of view of optimization. We suppose that if we choose to evaluate at $t \in [0, 1]$, then we observe $f(t) + \xi(t)$, where $\{\xi(t) : t \in [0, 1]\}$ are independent normal random variables with mean 0 and variance σ^2.

A nonadaptive algorithm using $n \geq 1$ evaluations specifies a set of n points $\{t_i : 1 \leq i \leq n\}$ which we label in increasing order $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \leq 1$. Let $y_i = f(t_i) + \xi(t_i)$, $1 \leq i \leq n$ be the observed values. Based on the n observations, the algorithm uses a function $t_n^* : \mathbb{R}^n \to [0, 1]$ to produce an estimate $t_n^*(y)$ of the global minimizer t^*. Setting $t_n^*(y)$ to the (first) minimizer of the conditional mean minimizes the expectation of $\Delta_n \equiv f(t_n^*(y)) - f(t^*)$; see Eq. 2. Several papers have studied the average error $E\Delta_n$ for nonadaptive algorithms, both for the noisy case and for the case of no noise ($\sigma = 0$).

Consider first the case of randomized algorithms with $\sigma = 0$. The minimizer of the Wiener process has the arcsine density; if observation points are chosen randomly according to the arcsine density $(\pi\sqrt{t(1-t)})^{-1}$ on (0, 1), then $\sqrt{n}E(\Delta_n) \to \mathcal{B}(3/4, 3/4)/\sqrt{2\pi} \approx 0.675978$, where $\mathcal{B}$ is the beta function. This is not the best choice of sampling distribution; if the t_i are instead chosen independently according to the $Beta(2/3, 2/3)$ distribution, then $\sqrt{n}E(\Delta_n) \to \mathcal{B}(2/3, 2/3)^{3/2}/(\pi\sqrt{2}) \approx 0.662281$, a slight improvement. Both distributions are better than the uniform distribution, for which $\sqrt{n}E(\Delta_n) \to 2^{-1/2}$; see [1].

The deterministic versions of these nonadaptive algorithms have slightly better convergence rates. A sequence of knots $\{t_i^n; 1 \leq i \leq n\}$ is a *regular sequence* for a density ψ if the knots form quantiles with respect to ψ; i.e.,

$$\int_{t=0}^{t_i^n} \psi(t)\,\mathrm{d}t = \frac{i-1}{n-1}$$

for $1 \le i \le n$. If we take ψ to be the uniform distribution and construct an algorithm based on the corresponding regular sequence, then $\sqrt{n}E(\Delta_n) \to c \approx 0.5826$. With ψ the arcsine density, the convergence is to a value of approximately $0.956c$, and with ψ the $Beta(2/3, 2/3)$ density, this improves to approximately $0.937c$; see [3]. These deterministic algorithms are noncomposite in that the number of evaluations n must be specified in advance, and this can be seen as a disadvantage relative to the randomized algorithms that tends to offset their better convergence rates; see [2].

In the case of noisy evaluations, it was shown in [5] that with equispaced points, $E\Delta_n = \Omega\left(n^{-1/4}\right)$. Regular sequences from beta distributions were considered in [4]. There it was shown that the regular sequence from a beta distribution with parameters (3/5, 3/5) results in a smaller error than equi-spaced points. For equi-spaced points,

$$\frac{n^{1/4}}{\sigma} E\Delta_n \to \gamma$$

for a constant $\gamma \in [1/2, 3/2]$. For the regular sequence from a beta distribution with parameters (3/5, 3/5),

$$\frac{n^{1/4}}{\sigma} E\Delta_n$$

converges to approximately 0.958γ.

In the exact-evaluation case ($\sigma = 0$), $\Delta_n \ge 0$ and $E\Delta_n$ is a reasonable error criterion. With noise corrupted evaluations, Δ_n can be negative or positive, and so we consider the error

$$e_n \equiv \left(E\Delta_n^2\right)^{1/2}.$$

In this paper we will show that there is a positive constant c such that $e_n \ge cn^{-1/4}$ after $n \ge 4$ noisy function evaluations, no matter how the $\{t_i\}$ are chosen.

3 The conditional distribution

In order to analyze the error e_n we need formulas for the conditional distribution of f given n function evaluations with error. Since both function and noise are Gaussian, so is the conditional measure. Denote by $\mathcal{F}_n$ the σ-field generated by the observations:

$$\mathcal{F}_n = \sigma\{f(t_i) + \xi(t_i) : 1 \le i \le n\}.$$

Let us recall some facts about the conditional distribution; see [8,5]. Let R_n be the covariance function of the conditional distribution given $\{f(t_i) + \xi(t_i) : 1 \le i \le n\}$;

$$R_n(s,t) = E\left(f(s)f(t) \,|\, \mathcal{F}_n\right), \quad 0 \le s,t \le 1.$$

We will only need the diagonal $R_n(s,s)$, which gives the conditional variance at $s \in [0, 1]$. For $t_{i-1} \le s \le t_i$,

$$R_n(s,s) = \frac{(s-a_{i-1})(b_i - s)}{b_i - a_{i-1}}$$

and for $t_n \le s \le 1$,

$$R_n(s,s) = s - a_n,$$

where the sequences $\{a_i\}$, $\{c_i\}$, $\{d_i\}$, $0 \le i \le n$, and $\{b_i\}$, $1 \le i \le n$ are defined as follows:

$$\begin{aligned} a_0 &= c_0 = 0 \\ c_i &= \frac{\sigma^2(t_i - a_{i-1})}{\sigma^2 + (t_i - a_{i-1})} \\ a_i &= t_i - c_i \end{aligned}$$

for $i = 1, \ldots, n$, and

$$\begin{aligned} d_n &= c_n \\ b_i &= a_{i-1} + \frac{(t_i - a_{i-1})^2}{(t_i - a_{i-1}) - d_i} \\ d_{i-1} &= \frac{(t_{i-1} - a_{i-1})(b_i - t_{i-1})}{b_i - a_{i-1}} \end{aligned}$$

for $i = n, \ldots, 1$ (see [8]). The covariance function R_n does not depend on the values observed.

The conditional mean m_n of f given the observations is constructed as follows. Define the covariance matrix Σ_n by $\Sigma_n(i,i) = t_i + \sigma^2$ and $\Sigma_{ij} = t_{\min(i,j)}$ if $i \ne j$ for $1 \le i, j \le n$. Then for $t_i \le s \le t_{i+1}$,

$$\begin{aligned} m_n(s,y) &\equiv E\left(f(s) \mid f(t_i) + \xi_i = y_i, 1 \le i \le n\right) \\ &= (y_1, y_2, \ldots, y_n)\left(\Sigma_n\right)^{-1}(t_1, \ldots, t_i, s, s, \ldots, s); \end{aligned}$$

see [5], where it is shown how to compute Σ_n^{-1} in time $\mathcal{O}(n^2)$.

4 Lower bound

This section contains the statement and proof of the main result. Two technical lemmas used in the proof of Theorem 1 follow at the end of this section.

Consider a fixed nonadaptive algorithm that makes $n \ge 4$ function evaluations. Let $y \in \mathbb{R}^n$ be the observed information, where $y_i = f(t_i) + \xi_i$. Let μ be the a priori distribution of the information y, and let $w(df \mid y)$ be the conditional distribution of f given the information y.

In general we approximate the minimizer t^* by some function $A(y)$ of the observed information y. We take the approximation error to be

$$E\left(f(A(y)) - f(t^*)\right),$$

where t^* is the (first) global minimizer of f. Let

$$m_n^*(y) = \min_{0 \le s \le 1} m_n(s,y),$$

and set

$$t_n^*(y) = \inf\{t \in [0,1] : m_n(t,y) = m_n^*(y)\}.$$

Then

$$E\left(f(A(y)) - f(t^*)\right) = \int\limits_{\mathbb{R}^n}\int\limits_{C_0} f(A(y))w(\mathrm{d}f \mid y)\mu(\mathrm{d}y) - Ef(t^*)$$
$$= \int\limits_{\mathbb{R}^n} m_n(A(y), y)\mu(\mathrm{d}y) - Ef(t^*) \quad (1)$$
$$\geq \int\limits_{\mathbb{R}^n} m_n(t_n^*(y), y)\mu(\mathrm{d}y) - Ef(t^*)$$
$$= E\left(f(t_n^*(y)) - f(t^*)\right). \quad (2)$$

Therefore, $t_n^*(y)$ is optimal and we consider only it from now on in establishing our lower complexity bound. We will examine the mean-squared error

$$e_n^2 \equiv E\left(m_n(t_n^*(y), y) - f(t^*)\right)^2$$

of the estimator $m_n(t_n^*(y), y)$, obtaining a lower bound. Our point of view is that the algorithm that chooses the minimum of the conditional mean is reasonable, and optimal in the sense that it has minimal bias among all estimators. Our goal is to bound the mean-squared error of this estimator over all choices of evaluation points. Our main result is

Theorem 1 *For any nonadaptive algorithm using $n \geq 4$ noisy function evaluations,*

$$e_n \geq \left(\frac{4\alpha^2\sqrt{13}}{75\pi\sqrt{3}}\left(1 - \alpha^2\frac{129}{80}\right)\right)^{1/2}\sqrt{\sigma}\, n^{-1/4}, \quad (3)$$

where

$$\alpha = \frac{2\sqrt{2} - 2}{\sqrt{\pi}}.$$

Fix $2 \leq k_n < n$ with $n \geq 4$ and for $i = 1, 2, \ldots, k_n$ set

$$K_i = \left[\frac{i-1}{k_n}, \frac{i}{k_n}\right],$$
$$J_i = \left[\frac{5i-4}{5k_n}, \frac{5i-1}{5k_n}\right],$$
$$I_i = \left[\frac{5i-3}{5k_n}, \frac{5i-2}{5k_n}\right].$$

That is, we partition the unit interval into k_n equal-length subintervals, and for the ith subinterval K_i, I_i is the central one-fifth and J_i is the central three-fifths. Since t^* has density

$$h(t) = \frac{1}{\pi\sqrt{t(1-t)}}$$

which is bounded below by $2/\pi$ on $[0, 1]$, we have

$$P(t^* \in I_i) = \int\limits_{t=\frac{5i-3}{5k_n}}^{t=\frac{5i-2}{5k_n}} \frac{1}{\pi\sqrt{t(1-t)}}\mathrm{d}t \geq \frac{2}{5\pi k_n}. \quad (4)$$

The main idea of the proof is as follows. There is probability at least $2/5\pi$ (for any n) that the global minimizer t^* of f lies in one of the I_i's. If $t^* \in I_i$, we consider two possibilities (at least one must hold, and both can hold). If $t_n^*(y) \notin J_i$, then because $\left|t_n^*(y) - t^*\right|$ is at least $(5k_n)^{-1}$, we obtain a lower bound on the mean-squared error based on the growth of the deviations of the Wiener path from the global minimum with distance from the minimizer. On the other hand, if $t_n^*(y) \in K_i$, then we use a lower bound on the variance of the conditional distribution in K_i to bound the mean-squared error (with also the noise magnitude σ appearing in the bound). Choosing k_n of order $n^{1/2}$ gives the required lower bound of $n^{-1/4}$ for the root mean-squared error e_n.

Let

$$Z_n = \{i \in \{1, 2, \ldots, k_n\} : J_i \cap \{t_1, t_2, \ldots, t_n\} = \emptyset\}$$

and set $z_n = |Z_n|$. Then

$$\begin{aligned} e_n^2 &= E\left((m_n(t_n^*, y) - f(t^*))^2\right) \\ &\geq \sum_{i \in Z_n} \int_{t \in I_i} E\left((m_n(t_n^*, y) - f(t^*))^2 \,|\, t^* = t\right) P(t^* \in \mathrm{d}t) \\ &\quad + \sum_{i \notin Z_n} \int_{t \in I_i} E\left((m_n(t_n^*, y) - f(t^*))^2 \,|\, t^* = t\right) P(t^* \in \mathrm{d}t). \end{aligned} \tag{5}$$

Let us consider the first sum. For $i \in Z_n, t_n^* \notin J_i$, since t_n^* is the first minimum of the conditional mean, which is piecewise linear between the t_i's. Therefore, for $t_n^* \notin J_i$,

$$E\left(m_n(t_n^*, y) - f(t^*) \,|\, t^* = t \in I_i\right) = E\left(f(t_n^*) - f(t^*) \,|\, t^* = t \in I_i\right) \geq \frac{\alpha}{\sqrt{5k_n}} \tag{6}$$

by Lemma 1. Then,

$$\begin{aligned} &E\left((m_n(t_n^*, y) - f(t^*))^2 \,|\, t^* = t \in I_i\right) \\ &\quad \geq \left(E\left(m_n(t_n^*, y) - f(t^*) \,|\, t^* = t \in I_i\right)\right)^2 \geq \frac{\alpha^2}{5k_n} \end{aligned} \tag{7}$$

and

$$\begin{aligned} &\sum_{i \in Z_n} \int_{t \in I_i} E\left((m_n(t_n^*, y) - f(t^*))^2 \,|\, t^* = t\right) P(t^* \in \mathrm{d}t) \\ &\quad \geq \sum_{i \in Z_n} \frac{\alpha^2}{5k_n} P(t^* \in I_i) \geq z_n \frac{\alpha^2}{5k_n} \frac{2}{5\pi k_n} \end{aligned} \tag{8}$$

by Eq. 4. Consider next the second sum in Eq. 5. Define the indicator variable

$$I_{\{f(t) < m_n(t,y)\}} = \begin{cases} 1 & \text{if } f(t) < m_n(t, y), \\ 0 & \text{otherwise.} \end{cases}$$

Observe that

$$\begin{aligned} E(m_n(t, y) - f(t^*))^2 &\geq E(m_n(t, y) - f(t^*))^2 I_{\{f(t) < m_n(t,y)\}} \\ &\geq E(m_n(t, y) - f(t))^2 I_{\{f(t) < m_n(t,y)\}} \\ &= \frac{1}{2} R_n(t, t) \end{aligned} \tag{9}$$

for any $t \in [0, 1]$, since $m_n(t, y) - f(t)$ is normally distributed with mean 0 and variance $R_n(t, t)$. Then

$$\begin{aligned}
&\sum_{i \notin Z_n} \int_{t \in I_i} E\left((m_n(t_n^*, y) - f(t^*))^2 \,|\, t^* = t\right) P(t^* \in \mathrm{d}t) \\
&\quad \geq \sum_{i \notin Z_n} \int_{t \in I_i} E\left((m_n(t_n^*, y) - f(t^*))^2 \,|\, t^* = t, t_n^* \in K_i\right) P(t_n^* \in K_i \,|\, t^* \in I_i) P(t^* \in \mathrm{d}t) \\
&\qquad + \sum_{i \notin Z_n} \int_{t \in I_i} E\left((m_n(t_n^*, y) - f(t^*))^2 \,|\, t^* = t, t_n^* \notin K_i\right) P(t_n^* \notin K_i \,|\, t^* \in I_i) P(t^* \in \mathrm{d}t) \\
&\quad \geq \sum_{i \notin Z_n} \int_{t \in I_i} \frac{1}{2} \min_{s \in K_i} R_n(s, s) P(t_n^* \in K_i \,|\, t^* \in I_i) P(t^* \in \mathrm{d}t) \text{ by (9)} \\
&\qquad + \sum_{i \notin Z_n} \int_{t \in I_i} \frac{\alpha^2}{5k_n} P(t_n^* \notin K_i \,|\, t^* \in I_i) P(t^* \in \mathrm{d}t) \text{ by (7)} \\
&\quad \geq \frac{2}{5\pi k_n} \left(\frac{1}{2} \frac{8}{27n/\sigma^2 + 12k_n^2} \left(\sum_{i \notin Z_n} \sqrt{q_i} \right)^2 + \frac{\alpha^2}{5k_n} \sum_{i \notin Z_n} (1 - q_i) \right) \text{ by Lemma 2,}
\end{aligned}$$

where

$$q_i = P(t_n^* \in K_i \,|\, t^* \in I_i).$$

Combining the last inequality with Eqs. 5 and 8 gives

$$e_n^2 \geq \frac{2}{5\pi k_n} \left(z_n \frac{\alpha^2}{5k_n} + \frac{4}{27n/\sigma^2 + 12k_n^2} \left(\sum_{i \notin Z_n} \sqrt{q_i} \right)^2 + \frac{\alpha^2}{5k_n} \sum_{i \notin Z_n} (1 - q_i) \right). \tag{10}$$

If

$$\frac{4}{27n/\sigma^2 + 12k_n^2} < \frac{\alpha^2}{5k_n} \tag{11}$$

then the right-hand side of Eq. 10 is a concave function of each $q_i \in [0, 1]$ and the minimum is attained at $q_i \in \{0, 1\}$. If also

$$k_n < \frac{\alpha^2}{5k_n} \frac{27n/\sigma^2 + 12k_n^2}{4}, \tag{12}$$

then the minimum is attained with $q_i = 1 \forall i$. The condition Eq. 12 implies Eq. 11, and Eq. 12 holds if we choose

$$k_n = \sqrt{\frac{3}{13}} \frac{1}{\sigma} \lceil \sqrt{n} \rceil.$$

Then

$$e_n^2 \geq \frac{2}{5\pi k_n}\left(z_n \frac{\alpha^2}{5k_n} + \frac{4}{27n/\sigma^2 + 12k_n^2}(k_n - z_n)^2\right)$$
$$= \frac{2}{5\pi k_n}\left(\frac{\alpha^2}{5}\frac{z_n}{k_n} + \frac{4}{27n/k_n^2\sigma^2 + 12}\left(1 - \frac{z_n}{k_n}\right)^2\right). \quad (13)$$

Considering this last expression as a function of the real variable $w = z_n/k_n \in [0, 1]$, the minimum value is attained at

$$w^* = 1 - \frac{\alpha^2}{10}\frac{27n/k_n^2\sigma^2 + 12}{4}$$

giving the lower bound

$$e_n^2 \geq \frac{2\alpha^2}{25\pi k_n}\left(1 - \frac{\alpha^2}{80}\left(27\frac{n}{k_n^2\sigma^2} + 12\right)\right) \geq \frac{4\alpha^2\sigma}{75\pi\sqrt{\frac{3}{13}}\sqrt{n}}\left(1 - \alpha^2\frac{129}{80}\right),$$

where in the last inequality we used the fact that for $n \geq 4$,

$$\lceil\sqrt{n}\rceil \leq \frac{3}{2}\sqrt{n}.$$

This completes the proof of Theorem 1.

Lemma 1 *For* $0 < \delta < 1/2$ *and* $2\delta \leq t^* = t \leq 1 - 2\delta$,

$$E\left(\min_{|s-t|\geq\delta} f(s) - f(t) \,|\, t^* = t\right) \geq \sqrt{\delta}\,\alpha,$$

where

$$\alpha = \frac{2\sqrt{2} - 2}{\sqrt{\pi}}.$$

Proof Conditional on t^*, $\{f(t^* + s) - f(t^*), -t^* \leq s \leq 1 - t^*\}$ is a "two-sided" Brownian meander of duration t^* to the left and $1 - t^*$ to the right. That is, conditional on t^*, $\{f(t^* + s) - f(t^*), 0 \leq s \leq 1 - t^*\}$ is a Brownian meander on $[0, 1 - t^*]$ and similarly, $\{f(t^* - s) - f(t^*), 0 \leq s \leq t^*\}$ is an independent Brownian meander on $[0, t^*]$. Roughly speaking, the Brownian meander is a Wiener process conditioned to stay positive over its duration. Let Y be a two-sided three-dimensional Bessel process; that is, $\{Y(t), t \geq 0\}$ and $\{Y(-t), t \geq 0\}$ are independent three-dimensional Bessel processes. This process is a Wiener process conditioned to take positive values after time 0 and to never return to 0, and $Y(t)$ has density

$$P(Y(t) \in \mathrm{d}y)/\mathrm{d}y = \left(\frac{2}{\pi t^3}\right)^{1/2} y^2 \exp\left(-\frac{y^2}{2t}\right), \quad y \geq 0; \quad (14)$$

see [9]. The distribution of a Brownian meander over an interval $[0, t]$ converges to the distribution of a three-dimensional Bessel process over $[0, t]$ as the duration of the Brownian meander tends to infinity. Since the minimum past a point of a Brownian meander is stochastically decreasing in the duration of the meander, the required expectation is bounded below by

$$E\min_{|t|\geq\delta} Y(t) = \sqrt{\delta}E\min_{|t|\geq 1} Y(t).$$

Using Eq. 14 and the fact that the minimum past time t of Y has the distribution $UY(t)$, where U is uniformly distributed between 0 and 1 independent of Y (see [9]), we obtain

$$P\left(\min_{t\geq 1} Y(t) \leq z\right) = \mathrm{Erf}\left(\frac{z}{\sqrt{2}}\right), \quad z \geq 0,$$

and so

$$E \min_{|t|\geq 1} Y(t) = \int_{z=0}^{\infty} \mathrm{Erfc}\left(\frac{z}{\sqrt{2}}\right)^2 \mathrm{d}z = \frac{2\sqrt{2}-2}{\sqrt{\pi}}.$$

□

Lemma 2 *For* $0 \leq q_i \leq 1, 1 \leq i \leq k_n$,

$$\sum_{i\notin Z_n} q_i \min_{s\in K_i} R_n(s,s) \geq \frac{8}{27n/\sigma^2 + 12k_n^2}\left(\sum_{i\notin Z_n}\sqrt{q_i}\right)^2.$$

Proof Set

$$c_j \equiv \#\{t_i \in K_j\}, \quad j = 1, 2, \ldots, k_n.$$

From [8], Lemma 3.8.3, we have that for any $0 \leq a < t < b \leq 1$,

$$R_n(t,t) \geq \frac{\sigma^2 \psi(t)}{\sigma^2 + c(a,b)\psi(t)}, \tag{15}$$

where

$$\psi(t) = \frac{(t-a)(b-t)}{b-a}$$

and $c(a,b)$ is the number of points t_i that lie in the subinterval $(a,b]$. As a consequence,

$$\psi(t) \leq (b-a)/4. \tag{16}$$

Then for $1 < i < k_n$ and $t \in K_i$, applying Eq. 15 to the interval $[(i-2)/k_n, (i+1)/k_n]$ gives the inequality

$$R_n(t,t) \geq \frac{\sigma^2 \Psi(t)}{\sigma^2 + (c_{i-1}+c_i+c_{i+1})\Psi(t)},$$

where

$$\Psi(t) = \frac{(t-(i-2)/k_n)\,((i+1)/k_n - t)}{(i+1)/k_n - (i-2)/k_n} = \frac{(t-(i-2)/k_n)((i+1)/k_n - t)}{3/k_n}.$$

Therefore,

$$\frac{2/3}{k_n} \leq \Psi(t) \leq \frac{3/4}{k_n}$$

and for $1 < i < k_n$ and $t \in K_i$,

$$R_n(t,t) \geq \frac{2/3\sigma^2}{\sigma^2 k_n + (c_{i-1}+c_i+c_{i+1})(3/4)} = \frac{8/9}{4k_n/3 + (c_{i-1}+c_i+c_{i+1})/\sigma^2}.$$

Define

$$p_i = \begin{cases} q_i & \text{if } i \notin Z_n, \\ 0 & \text{otherwise} \end{cases}$$

and let us consider

$$\sum_{i=1}^{k_n} p_i \min_{s \in K_i} R_n(s, s) \geq p_1 \left(\frac{8/9}{4k_n/3 + (c_1 + c_2)/\sigma^2} \right)$$
$$+ \sum_{i=2}^{k_n - 1} p_i \left(\frac{8/9}{4k_n/3 + (c_{i-1} + c_i + c_{i+1})/\sigma^2} \right) + p_{k_n} \left(\frac{8/9}{4k_n/3 + (c_{k_n-1} + c_{k_n})/\sigma^2} \right)$$
$$= \frac{8}{9} \left(p_1 \left(\frac{4k_n}{3} + \frac{c_1 + c_2}{\sigma^2} \right)^{-1} + \sum_{i=2}^{k_n-1} p_i \left(\frac{4k_n}{3} + \frac{c_{i-1} + c_i + c_{i+1}}{\sigma^2} \right)^{-1} \right.$$
$$\left. + p_{k_n} \left(\frac{4k_n}{3} + \frac{c_{k_n-1} + c_{k_n}}{\sigma^2} \right)^{-1} \right). \quad (17)$$

Minimize this last expression over $\{c_i\}$ subject to the constraint

$$\sum_{i=1}^{k_n} c_i = n. \quad (18)$$

Let $x_1 = (c_1 + c_2)/\sigma^2$, $x_m = (c_{m-1} + c_m)/\sigma^2$, and for $1 < i < k_n$ set $x_i = (c_{i-1} + c_i + c_{i+1})/\sigma^2$. Then minimizing Eq. 17 subject to Eq. 18 is equivalent to minimizing

$$\frac{8}{9} \sum_{i=1}^{k_n} p_i \left(\frac{4k_n}{3} + x_i \right)^{-1}$$

subject to

$$\sum_{i=1}^{k_n} x_i = (3n - c_1 - c_m)/\sigma^2 \quad (19)$$

and also linear constraints on the x_i to ensure nonnegativity of the c_i. To get a lower bound, ignore the latter constraints, and also replace Eq. 19 with

$$\sum_{i=1}^{k_n} x_i = 3n/\sigma^2 \quad (20)$$

which gives a lower value since the new constraint allows the x_i to be larger and thus the objective function to be smaller. The first-order optimality conditions (equating derivatives with respect to each x_i) require that

$$x_i = \lambda \sqrt{p_i} - \frac{4}{3} k_n$$

for some constant λ. The constraint Eq. 20 implies that

$$\lambda = \frac{3n/\sigma^2 + \frac{4}{3} k_n^2}{\sum_{j=1}^{k_n} \sqrt{p_j}}.$$

Therefore, the solution is

$$x_i = \left(3n/\sigma^2 + \frac{4}{3}k_n^2\right) \frac{\sqrt{p_i}}{\sum_{j=1}^{k_n} \sqrt{p_j}} - \frac{4}{3}k_n,$$

giving the lower bound

$$\frac{8/9}{3n/\sigma^2 + \frac{4}{3}k_n^2} \left(\sum_{i=1}^{k_n} \sqrt{p_i}\right)^2 = \frac{8}{27n/\sigma^2 + 12k_n^2} \left(\sum_{i \notin Z_n} \sqrt{q_i}\right)^2 .$$

□

Acknowledgments This material is based upon work supported by the National Science Foundation under grant CMMI-0825381.

References

1. Al-Mharmah, H., Calvin, J.: Optimal random nonadaptive algorithm for global optimization of Brownian motion. J. Glob. Optim. **8**, 81–90 (1996)
2. Al-Mharmah, H., Calvin, J.: Comparison of one-dimensional composite and non-composite passive algorithms. J. Glob. Optim. **15**, 169–180 (1999)
3. Calvin, J.M.: Average performance of passive algorithms for global optimization of Brownian motion. J. Math. Anal. Appl. **191**, 608–617 (1995)
4. Calvin, J.M.: Nonadaptive univariate optimization for observations with noise. In: Törn A., Zilinskas, J. (eds.) Models and Algorithms for Global Optimization, pp. 185–192. Springer, New York (2007)
5. Calvin, J.M., Žilinskas, A.: One-dimensional global optimization for observations with noise. Comput. Math. Appl. **50**, 157–169 (2005)
6. Kushner, H.: A versatile stochastic model of a function of unknown and time-varying form. J. Math. Anal. Appl. **5**, 150–167 (1962)
7. Kushner, H.: A new method of locating the maximum point of a multipeak curve in the presence of noise. J. Basic Eng. **86**, 97–106 (1964)
8. Plaskota, L.: Noisy Information and Computational Complexity. Cambridge University Press, Cambridge (1996)
9. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion. Springer-Verlag, Berlin Heidelberg (1991)
10. Ritter, K.: Approximation and optimization on the Wiener space. J. Complex. **6**, 337–364 (1990)
11. Strongin, R.: Information method for multiextremal minimization with noisy observations. Trans. USSR Acad. Sci. Eng. Cybern. **6**, 116–126 (1969)
12. Strongin, R., Sergeyev, Y.: Global Optimization with Non-Convex Constraints. Kluwer, Dordrecht (2000)
13. Törn, A., Žilinskas, A.: Global Optimization. Lecture Notes in Computer Science, Vol. 350. Springer-Verlag, Berlin (1989)
14. Žilinskas, A.: Two algorithms for one-dimensional multimodal minimization. Math. Oper. Stat. Ser. Optim. **12**, 53–63 (1981)
15. Žilinskas, A.: Axiomatic characterization of a global optimization algorithm and investigation of its search strategies. Oper. Res. Lett. **4**, 35–39 (1985)
16. Zhigljavsky, A., Žilinskas, A.: Stochastic Global Optimization. Springer, New york (2008)