Journal of Complexity 27 (2011) 404–416

Contents lists available at ScienceDirect

Journal of Complexity

journal homepage: www.elsevier.com/locate/jco

A lower bound on complexity of optimization under the r-fold integrated Wiener measure☆

James M. Calvin

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

ARTICLE INFO

Article history:
Received 1 February 2010
Accepted 20 October 2010
Available online 31 October 2010

Keywords:
Optimization
Statistical models
Convergence

ABSTRACT

We consider the problem of approximating the global minimum of an r-times continuously differentiable function on the unit interval, based on sequentially chosen function and derivative evaluations. Using a probability model based on the r-fold integrated Wiener measure, we establish a lower bound on the expected number of function evaluations required to approximate the minimum to within ϵ on average.

1. Introduction

The global optimization problem is to approximate the minimum of a function that may have more than one local minimum. In this paper we are interested in how many function or derivative evaluations are required in order to obtain an ϵ approximation to the global minimum. We consider function classes that are subsets of $C^r([0, 1])$ for some $r \geq 1$. In the worst-case setting this problem is intractable without further restriction on the class of functions; for any approximation to the minimum based on a finite set of function and derivative values, there are elements of $C^r([0, 1])$ with those values and with arbitrarily large error. If we assume a bound on $\|f^{(r)}\|$, then $\Theta(\epsilon^{-1/r})$ function evaluations are required to obtain an error of at most ϵ; see [4]. An alternative to the worst-case analysis is an average-case analysis. We present such an analysis for random univariate functions distributed according to the r-fold integrated Wiener measure. This measure has been used in several studies of the efficiency of numerical methods, including global optimization [4] and zero-finding [2]. Additional problems are treated in [3]. We use the probability model and notation of [2], where the authors describe the average-case complexity of zero-finding. The global optimization problem considered here can be thought of as the problem of approximating a particular zero of the first derivative.

☆ This material is based upon work supported by the National Science Foundation under Grant No. CMMI-0825381.
E-mail address: calvin@njit.edu.

0885-064X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jco.2010.10.006

For the probability model that we consider, an average-case analysis is presented in [4] that shows that an ϵ approximation can be obtained with at most

$$C\left((1/\epsilon)^{1/(r+1/2)}\log(1/\epsilon)^{1/(2r+1)}\right) \tag{1.1}$$

evaluations (see Eq. (8) in [4]), where C depends on r but not on ϵ. In this paper we establish a lower bound on the complexity, showing that for $r \geq 2$ any algorithm needs at least a constant times

$$\log(1/\epsilon)^{1/((2r-3)(2r+1))}$$

evaluations. For $r = 1$, at least a constant times

$$\log(1/\epsilon)^{1/3}$$

evaluations are required.

The gap between the upper bound given by (1.1) and the lower bound of this paper is very large. The upper bound (1.1) was derived from on upper bound for L_∞ approximation, and it was stated that it seemed unlikely that the bound was sharp for optimization. We conjecture that for optimization it is possible to obtain an ϵ approximation with a number of function evaluations that is bounded by a multiple of $\log(1/\epsilon)^c$ for some $c > 0$.

The question of a lower complexity bound in the case of $r = 0$ was treated in [1], where it was shown that to obtain an error of at most ϵ, a constant times $\log\log(1/\epsilon)\log(1/\epsilon)$ function evaluations are necessary.

The complexity of zero-finding under the probability model considered here for $r \geq 2$ was examined in [2] (there the boundary conditions were chosen to ensure the existence of a zero). The average-case complexity of computing an ϵ-approximation to a zero is of order $\log(1/\epsilon)$ if the number of function evaluations is nonadaptive. However, if the number of evaluations can be chosen adaptively, then the average-case complexity is of order $\log\log(1/\epsilon)$. (The constant factors giving upper and lower bounds depend on r and the boundary values.)

2. Problem formulation

We are interested in approximating the global minimum of a member of a class F of functions $f : [0, 1] \to \mathbb{R}$ that are at least once continuously differentiable. After making some function evaluations, we construct an approximation $A(f)$ to the minimizer of f, and our goal is to make $f(A(f)) - \min_{0\leq t\leq 1} f(t)$ small.

We undertake an average-case analysis based on a variant of the Wiener measure. The r-fold integrated Wiener measure is obtained by r-fold integration of the Brownian motion paths. We translate each path by a suitable polynomial so that prescribed boundary conditions are satisfied, thereby obtaining a conditional r-folded Wiener measure [2]. Let $r \geq 1$ denote the smoothness, and $\mathbf{a} = (a_0, \ldots, a_r)$ and $\mathbf{b} = (b_0, \ldots, b_r)$ the boundary conditions at 0 and 1, respectively. We consider a class of functions

$$F = \{f \in C^r([0, 1]) : f^{(i)}(0) = a_i, f^{(i)}(1) = b_i, \text{ for } i = 0, 1, \ldots, r\}.$$

In order to ensure that the global minimum occurs in the interior of the interval, we assume that $a_1 < 0$ and $b_1 > 0$. Equip the space $C^r([0, 1])$ with the norm

$$\|f\| = \max\left\{\|f\|_\infty, \ldots, \left\|f^{(r)}\right\|_\infty\right\},$$

where

$$\|f\|_\infty = \sup_{0\leq s\leq 1} |f(s)|.$$

Denote the Borel σ-algebra of F by $\mathcal{F}$. Let P_r denote the distribution of the r-fold integrated Wiener process.

We now describe the class of algorithms for which our lower complexity bound holds. For the average-case analysis, certain Borel measurability restrictions are required so that the integrals that appear below are well defined. The formal description of an algorithm follows. The initial evaluation point $t_1 \in [0,1]$ and order of derivative $j_1 \in \{0,\ldots,r\}$ are fixed, and $f^{(j_1)}(t_1)$ is evaluated. Termination is determined by a sequence of measurable maps

$$h_i : \mathbb{R}^i \to \{0,1\}.$$

If $h_1(f^{(j_1)}(t_1)) = 1$ the algorithm terminates, otherwise a new point t_2 and order of derivative j_2 are chosen and the process continues. Suppose that the algorithm has performed k evaluations. If

$$h_k\left(f^{(j_1)}(t_1),\ldots,f^{(j_k)}(t_k)\right) = 0$$

then we compute

$$t_{k+1}\left(f^{(j_1)}(t_1),\ldots,f^{(j_k)}(t_k)\right), \qquad j_{k+1}\left(f^{(j_1)}(t_1),\ldots,f^{(j_k)}(t_k)\right),$$

where

$$t_{k+1} : \mathbb{R}^k \to [0,1], \qquad j_{k+1} : \mathbb{R}^k \to \{0,\ldots,r\}$$

are Borel measurable. Then $f^{(j_{k+1})}(t_{k+1})$ is evaluated. The total number of evaluations is

$$n(f) = \inf\left\{k : h_k\left(f^{(j_1)}(t_1),\ldots,f^{(j_k)}(t_k)\right) = 0\right\}.$$

The approximation to the minimizer is given by

$$A(f) = \phi_{n(f)}\left(f^{(j_1)}(t_1),\ldots,f^{(j_{n(f)})}(t_{n(f)})\right)$$

for measurable maps $\phi_d : \mathbb{R}^d \to [0,1]$, $d \geq 1$. Then $A : F \to [0,1]$ is measurable.

Let $N : F \to Y = \cup_{d=1}^{\infty}([0,1]\times\{0,\ldots,r\}\times\mathbb{R})^d$ denote the information operator, where

$$N(f) = \left((t_i, j_i, f^{(j_i)}(t_i)),\ 1 \leq i \leq n(f)\right).$$

This operator is measurable $\mathcal{F}/\mathcal{Y}$, where $\mathcal{Y}$ is the smallest σ-field on Y that contains sets of the form

$$\prod_{i=1}^{d}(A_i, j_i, B_i)$$

for $d \geq 1$, $j_i \in \{0,\ldots,r\}$, and Borel sets A_i, B_i. Let $Q_r = N \cdot P_r$ denote the image of P_r under N. Since $(F, \|\cdot\|)$ is a complete separable metric space, there exists a regular conditional probability $R_r : \mathcal{F} \times Y \to [0,1]$, so that

$$P_r(A \cap N^{-1}(B)) = \int_B R_r(A,y)Q_r(\mathrm{d}y)$$

for every $B \in \mathcal{Y}$, $A \in \mathcal{F}$.

We refer to $n(f)$ as the cardinality of information, and take the cost of an algorithm, when applied to the function f, to be $n(f)$. Since we are interested in a lower bound, we assume that $n(f)$ is finite and ignore the cost of the computations involved in computing the evaluation points and constructing the approximation. The error, when applied to f, is

$$\Delta(f) \equiv f(A(f)) - \min_{0\leq t\leq 1} f(t).$$

Given a probability measure P_r defined on $\mathcal{F}$, we define the average cost and average error

$$\mathrm{cost}(A,N) = \int_F n(f)P_r(\mathrm{d}f),$$

and

$$\mathrm{error}(A,N) = \int_F [f(A(f)) - \min_{0\leq t\leq 1} f(t)]P_r(\mathrm{d}f) = \int_F \Delta(f)P_r(\mathrm{d}f).$$

We define the local error given the information $y \in Y$ by

$$\int_F [f(A(f)) - \min_{0\leq t\leq 1} f(t)]R_r(\mathrm{d}f, y) = \int_F \Delta(f)R_r(\mathrm{d}f, y).$$

Since

$$\int_F f(A(f))R_r(\mathrm{d}f, y)$$

is the conditional mean of f given the information y evaluated at the point $A(f)$, the local error is minimized by choosing $A(f)$ to be the first minimizer of the conditional mean; then

$$\int_F f(A(f))R_r(\mathrm{d}f, y) = \min_{0\le t\le 1}\int_F g(t)R_r(\mathrm{d}g, y).$$

Since this is the optimal choice, this is the only one we consider in the rest of the paper, and the local error is

$$\min_{0\le t\le 1}\int_F f(t)R_r(\mathrm{d}f, y) - \int_F \min_{0\le t\le 1} f(t)R_r(\mathrm{d}f, y). \tag{2.1}$$

The form of the statement and proof of the lower bound are slightly different for $r = 1$ and for $r \ge 2$. In order to treat both cases together, we introduce the notation $q = \min\{r, 2\}$. Our result is

Theorem 2.1. *Let $0 < \epsilon < 1/16$ and $r \ge 1$ and consider an arbitrary algorithm (A, N) that has average error at most ϵ:*

$$\int_F \Delta(f)P_r(\mathrm{d}f) \le \epsilon.$$

Then there exists a positive number $C(r, \mathbf{a}, \mathbf{b})$ such that the average cost of the algorithm is at least

$$\int_F n(f)P_r(\mathrm{d}f) \ge C(r, \mathbf{a}, \mathbf{b}) \cdot \log(1/\epsilon)^{\frac{1}{(2r+1-2q)(2r+1)}}.$$

The proof of Theorem 2.1 appears in the next section. In the remainder of this section we justify two simplifying assumptions that will be used in the proof. The first is to base our approximation on *all* function and derivative values at each of the selected points. Because the vector process

$$\left(f(t), f'(t), \ldots, f^{(r)}(t)\right)$$

has the Markov property, this will simplify the calculations since the conditional processes between evaluation points are independent. For $f \in F$, let

$$f[t] = \left(f(t), f'(t), \ldots, f^{(r)}(t)\right) \tag{2.2}$$

denote the Hermite information of order r. We assume that at each step a point $t \in (0, 1)$ is chosen and $f[t]$ is evaluated. Since we seek a lower bound, we count the cost of each such evaluation as 1 instead of $r + 1$.

For convenience, let us renumber the evaluation points in increasing order. Given Hermite information at $0 \le t_1 < t_2 < \cdots < t_n \le 1$, let $\mu_n(t)$ denote the conditional mean and $\sigma_n^2(t)$ the conditional variance at t. Conditional on $(t_i, j_i, f^{(j_i)}(t_i))$, $1 \le i \le n$, the processes

$$\{f(s) : t_{i-1} \le s \le t_i\}$$

are independent Gaussian processes. The conditional mean μ_n is the polynomial of degree at most $2r+1$ interpolating the observed function and derivative values at the endpoints of each subinterval; that is, for each i,

$$\mu_n[t_i] = f[t_i].$$

The conditional variance at $t \in [t_{i-1}, t_i]$ is given by

$$\sigma_n^2(t) = \frac{1}{(2r+1)(r!)^2}\left(\frac{(t-t_{i-1})(t_i-t)}{t_i-t_{i-1}}\right)^{2r+1}; \tag{2.3}$$

see Eq. (4.2) in [2].

The second simplification is to base the computation of the local error not on the conditional mean, but on a simple function that majorizes the conditional mean with high probability. The majorizing functions are given in the following

Lemma 2.2. *There exists a number $\beta \geq 1$ and a measurable set $F_\beta \subset F$ with the following properties:*

(1) $P_r(F_\beta) \geq \frac{1}{2}$;
(2) *For any $n \geq 1$ and any set of points $0 \leq t_1 < t_2 < \cdots < t_n \leq 1$, if we construct the conditional mean μ_n with respect to $f[t_1], f[t_2], \ldots, f[t_n]$ for $f \in F_\beta$, then for all $s \in [-t_n^*, 1 - t_n^*]$,*

$$\mu_n(t_n^* + s) - \mu_n(t_n^*) \leq \begin{cases} \beta\,|s| & \text{if } r = 1, \\ \beta s^2 & \text{if } r \geq 2, \end{cases}$$

where

$$t_n^* = \inf\{t \in [0, 1] : \mu_n(t) = \min_{0 \leq s \leq 1} \mu_n(s)\}$$

is the first minimizer of μ_n.

Proof. According to Lemma 4.1 in [2], for $r \geq 1$ there exists a constant $L_1 = L_1(r)$ such that

$$\sup_{t_{i-1} \leq s \leq t_i} \left|f'(s) - \mu_n'(s)\right| \leq L_1 \cdot (t_i - t_{i-1})^{r-1} \cdot \sup_{t_{i-1} \leq s \leq t_i} \left|f^{(r)}(s)\right|. \tag{2.4}$$

If $r \geq 2$, then there exists a constant $L_2 = L_2(r)$ such that

$$\sup_{t_{i-1} \leq s \leq t_i} \left|f''(s) - \mu_n''(s)\right| \leq L_2 \cdot (t_i - t_{i-1})^{r-2} \cdot \sup_{t_{i-1} \leq s \leq t_i} \left|f^{(r)}(s)\right|. \tag{2.5}$$

Therefore

$$\left\|\mu_n'\right\|_\infty \leq \left\|f'\right\|_\infty + L_1 \cdot \left\|f^{(r)}\right\|_\infty \leq (1 + L_1)\,\|f\|\,,$$

and so

$$\mu_n(t_n^* + s) - \mu_n(t_n^*) \leq (1 + L_1)\,\|f\|\,|s|\,.$$

Then we can choose $\beta > \max\{1, |a_i|\,, |b_i|\,, 0 \leq i \leq r\}$ large enough so that

$$P_r\left(\|f\| \leq \frac{\beta}{1 + L_1}\right) \geq \frac{1}{2}$$

and take

$$F_\beta = \left\{f \in F : \|f\| \leq \frac{\beta}{1 + L_1}\right\}.$$

Similarly, if $r \geq 2$, then

$$\left\|\mu_n''\right\|_\infty \leq \left\|f''\right\|_\infty + L_2 \cdot \left\|f^r\right\|_\infty \leq (1 + L_2)\,\|f\|\,,$$

and so

$$\begin{aligned} \mu_n(t_n^* + s) - \mu_n(t_n^*) &= s\mu_n'(t_n^*) + \int_{u=0}^{s} \int_{v=0}^{u} \mu_n''(t_n^* + v)\mathrm{d}v\,\mathrm{d}u \\ &\leq \frac{1}{2}s^2 \left\|\mu_n''\right\|_\infty \\ &\leq \frac{1}{2}s^2(1 + L_2)\,\|f\|\,. \end{aligned}$$

We used the fact that t_n^* is a minimizer of μ_n and so $\mu_n'(t_n^*) = 0$. Then we can choose $\beta > \max\{1, |a_i|\,, |b_i|\,, 0 \leq i \leq r\}$ large enough so that

$$P_r\left(\|f\| \leq \frac{2\beta}{1 + L_2}\right) \geq \frac{1}{2}$$

and set

$$F_\beta = \left\{ f \in F : \|f\| \leq \frac{2\beta}{1+L_2} \right\}. \quad \square$$

3. Proof of Theorem 2.1

Denote the normal cumulative distribution function by

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{y} e^{-x^2/2}\, dx,$$

and set $\Psi(y) = 1 - \Phi(y)$ and $\varphi(y) = \Phi'(y)$. Repeated integration by parts, using $\varphi'(y) = -y\varphi(y)$, yields

$$\begin{aligned}\Psi(z) &= 1 - \Phi(z) \\ &= \frac{1}{z}\varphi(z) - \int_{x=z}^{\infty} \frac{1}{x^2}\varphi(x)dx \\ &= \varphi(z)\left(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5} - 15\int_{x=z}^{\infty}\frac{1}{x^6}\varphi(x)dx\right).\end{aligned}$$

Therefore,

$$\Psi(z) \leq \varphi(z)\left(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5}\right),$$

and so

$$\varphi(z) - z\Psi(z) \geq \varphi(z)\left(\frac{1}{z^2} - \frac{3}{z^4}\right).$$

If $z \geq \sqrt{6}$, then

$$\frac{1}{z^2} - \frac{3}{z^4} \geq \frac{1}{2z^2},$$

and so we have the inequality

$$\varphi(z) - z\Psi(z) \geq \frac{1}{2z^2}\varphi(z), \quad z \geq \sqrt{6}. \tag{3.1}$$

Let us suppose an algorithm has made n evaluations, obtaining information

$$y = \left((t_i, j_i, f^{(j_i)}(t_i)), 1 \leq i \leq n\right).$$

In terms of the conditional mean μ_n and its minimizer t_n^*, the local error is given by

$$\mu_n(t_n^*) - \int_F \min_{0\leq t\leq 1} f(t) R_r(df, y).$$

The conditional distribution of the minimum of f over a subinterval $[t_{i-1}, t_i]$ is not known except for the case $r = 0$, which we exclude. We will approximate the conditional error in two steps. First we use the expected undershoot at a fixed point in a subinterval to give an upper bound on the minimum over the subinterval. Then instead of computing the expected minimum using the conditional mean μ_n, we replace μ_n with the majorizing function given in Lemma 2.2. Further, we consider only the error to the right or left of t_n^* (whichever subinterval is larger).

First, let us consider the expected absolute undershoot of $\mu_n(t_n^*)$ by $f(t)$, which we denote by $\gamma_n(t)$. Since conditional on y, for $t_{i-1} < t < t_i$,

$$f(t) \sim N(\mu_n(t), \sigma_n^2(t)),$$

we have

$$\begin{aligned}\gamma_n(t) &= \int_F (\mu_n(t_n^*) - f(t))^+ R_r(\mathrm{d}f, y)\\ &= \int_{x=-\infty}^{\infty} \left(\mu_n(t_n^*) - \mu_n(t) - \sigma_n(t)x\right)^+ \varphi(x)\mathrm{d}x\\ &= \sigma_n(t)\left(\varphi(\theta_n(t)) - \theta_n(t)\Psi(\theta_n(t))\right),\end{aligned}$$

where

$$\theta_n(t) = \frac{\mu_n(t) - \mu_n(t_n^*)}{\sigma_n(t)}. \tag{3.2}$$

Let $\{s_i\}$ be arbitrary points with $t_{i-1} \le s_i \le t_i$, $1 \le i \le n$. Then for every f, the local error satisfies

$$\mu_n(t_n^*) - \int_F \min_{0\le t\le 1} f(t) R_r(\mathrm{d}f, y) \ge \mu_n(t_n^*) - \int_F \min_{1\le i\le n} f(s_i) R_r(\mathrm{d}f, y). \tag{3.3}$$

Let us consider the local error for a function $f \in F_\beta$ (recall Lemma 2.2). If we require that the local error does not exceed δ, then for each i we require that $\gamma_n(s_i) \le \delta$. Since $\gamma_n(s_i)$ is decreasing in $\theta_n(s_i)$, if we replace μ_n by a function g that satisfies $g(t_n^*) = \mu_n(t_n^*)$ and $g(s) \ge \mu_n(s)\forall s$, then $\gamma_n(s_i)$ will be no larger and so must also be at most δ. Thus by Lemma 2.2, we can replace $\mu_n(t) - \mu_n(t_n^*)$ in the numerator of θ_n, defined at (3.2), by $\beta\left|t - t_n^*\right|$ if $r = 1$ and by $\beta(t - t_n^*)^2$ if $r \ge 2$.

Without loss of generality, assume that $t_n^* \le 1/2$ and that $t_j \le t_n^* < t_{j+1}$ for some j. (If not, then we consider the subinterval to the left of t_n^* below.) Then the expected error to the right of t_n^* is at least as large as in the case where $t_j = t_n^* \le 1/2$. Since we seek a lower bound, we can consider only the error to the right of t_n^*. Then we have $1/2 \ge t_n^* = t_j < t_{j+1} < \cdots < t_{j+k} = 1$, where k is the number of evaluations to the right of t_n^*. To simplify some of the subsequent expressions, set $\tau_i = t_{j+i} - t_j$ and $T_i = t_{j+i} - t_{j+i-1}$, $1 \le i \le k$.

We apply the lower bound (3.3) with the choice of points

$$s_{j+i} = t_{j+i-1} + \frac{t_{j+i} - t_{j+i-1}}{2r+1}, \quad 1 \le i \le k.$$

Set

$$\zeta_r = \frac{r!\sqrt{2r+1}(2r+1)^{2r+1}}{(2r)^{(2r+1)/2}},$$

so that

$$\sigma_n(s_{j+i}) = \zeta_r^{-1} T_i^{r+1/2}$$

using (2.3).

We will use the following fact several times in the sequel.

Lemma 3.1. *The expression*

$$\frac{r!(2r+1)^{2r-1/2}}{(2r)^{r+1/2}} \tag{3.4}$$

is increasing in $r \ge 2$.

Proof. We will show that the function of a real variable

$$g(x) = \frac{\Gamma(x+1)(2x+1)^{2x-1/2}}{(2x)^{x+1/2}}$$

is increasing in $x \ge 2$ by showing that the derivative of its logarithm is positive. We have

$$\frac{\mathrm{d}}{\mathrm{d}x}\log(g(x)) = \frac{\Gamma'(x+1)}{\Gamma(x+1)} + \log\left(\frac{(2x+1)^2}{2x}\right) + \frac{4x^2 - 6x - 1}{2x(2x+1)}.$$

$\Gamma(x+1)$ is positive and increasing for $x \ge 2$, and $4x^2 - 6x - 1 > 0$ for $x \ge 2$, and so $\log(g)$, and therefore g, is increasing on $[2, \infty)$. For integer arguments $g(r)$ is equal to the expression in (3.4). □

Recalling the notation $q \equiv \min\{r, 2\} \in \{1, 2\}$ and applying Lemma 2.2,

$$\begin{aligned} \theta_n(s_{j+i}) &= \frac{\mu_n(s_{j+i}) - \mu_n(t_j)}{\sigma_n(s_{j+i})} \\ &\le \frac{\beta\left(t_{j+i-1} - t_j + T_i/(2r+1)\right)^q}{\frac{1}{\zeta_r} T_i^{r+1/2}} \\ &= \frac{\beta\left(\tau_{i-1} + T_i/(2r+1)\right)^q}{\frac{1}{\zeta_r} T_i^{r+1/2}}. \end{aligned}$$

In order to apply inequality (3.1) we need to show that for each i,

$$\theta_n(s_{j+i}) > \sqrt{6}.$$

Note that $T_i \le 1,\ \beta \ge 1$, and $r - q \ge 0$ implies that

$$\frac{\zeta_r \beta(\tau_{i-1} + T_i/(2r+1))^q}{T_i^{(2r+1)/2}} \ge \frac{\zeta_r}{(2r+1)^q T_i^{r+1/2-q}} \ge \frac{r!(2r+1)^{2r+3/2-q}}{(2r)^{r+1/2}}.$$

For $r = 1 = q$ the last expression has value

$$3 \cdot (3/2)^{3/2} > \sqrt{6}.$$

For $r \ge 2 = q$ the expression is

$$\frac{r!(2r+1)^{2r-1/2}}{(2r)^{r+1/2}},$$

which is increasing in $r \ge 2$ by Lemma 3.1, and with $r = 2$ takes the value

$$10(5/4)^{5/2} > \sqrt{6}.$$

Therefore $\theta(s_{j+i}) > \sqrt{6}$ and we can apply inequality (3.1) to obtain

$$\begin{aligned} \gamma_n\left(s_{j+i}\right) &= \sigma_n(s_{j+i})\left(\varphi(\theta_n(s_{j+i})) - \theta_n(s_{j+i})\Psi(\theta_n(s_{j+i}))\right) \\ &\ge \sigma_n(s_{j+i}) \frac{1}{2\theta_n(s_{j+i})^2} \varphi(\theta_n(s_{j+i})) \\ &= \frac{1}{2\zeta_r} T_i^{(2r+1)/2} \varphi\left(\frac{\zeta_r \beta(\tau_{i-1} + T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right) \left(\frac{\zeta_r \beta(\tau_{i-1} + T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right)^{-2}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\log\left(\gamma_n\left(t_{j+i-1} + (t_{j+i} - t_{j+i-1})/(2r+1)\right)\right) \ge -\log(2\zeta_r) - \log(\sqrt{2\pi}) + \frac{(2r+1)}{2}\log(T_i) \\ &\quad - \frac{1}{2}\left(\frac{\zeta_r \beta(\tau_{i-1} + T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right)^2 - 2\log\left(\frac{\zeta_r \beta(\tau_{i-1} + T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right). \end{aligned}$$

In order for the error to be at most δ over the subinterval $[t_{j+i-1}, t_{j+i}]$, we must have that

$$\begin{aligned} \log(\delta) \ge\ & -\log(2\zeta_r) - \log(\sqrt{2\pi}) + \frac{(2r+1)}{2}\log(T_i) \\ & - \frac{1}{2}\left(\frac{\zeta_r \beta(\tau_{i-1} + T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right)^2 - 2\log\left(\frac{\zeta_r \beta(\tau_{i-1} + T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right), \end{aligned}$$

or

$$\log(1/\delta) \le \log(2\zeta_r\sqrt{2\pi}) - \frac{(2r+1)}{2}\log(T_i) + \frac{1}{2}\left(\frac{\zeta_r\beta(\tau_{i-1}+T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right)^2 + 2\log\left(\frac{\zeta_r\beta(\tau_{i-1}+T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right). \quad (3.5)$$

We claim that the last expression is at most

$$\left(\frac{\zeta_r\beta(\tau_{i-1}+T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right)^2,$$

or, equivalently, that

$$\frac{1}{2}\left(\frac{\zeta_r\beta(\tau_{i-1}+T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right)^2 - \log(2\zeta_r\sqrt{2\pi}) + \frac{(2r+1)}{2}\log(T_i) - 2\log\left(\frac{\zeta_r\beta(\tau_{i-1}+T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right) \ge 0. \quad (3.6)$$

The function $x \mapsto x^2/2 - 2\log(x)$ is increasing for $x \ge \sqrt{2}$, and

$$\begin{aligned}\frac{\zeta_r\beta(\tau_{i-1}+T_i/(2r+1))^q}{T_i^{(2r+1)/2}} &\ge \frac{\zeta_r}{(2r+1)^q T_i^{r+1/2-q}} \\ &\ge \frac{\zeta_r}{(2r+1)^q} \\ &= \frac{r!(2r+1)^{2r+3/2-q}}{(2r)^{r+1/2}} \\ &\ge \frac{r!(2r+1)^{2r-1/2}}{(2r)^{r+1/2}} \\ &\ge r!\left(\frac{2r+1}{2r}\right)^r (2r+1)^{r-1}.\end{aligned}$$

For $r = 1$ the last expression is equal to $3/2 > \sqrt{2}$, and for $r \ge 2$, the expression is at least $r!(2r+1)^{r-1} \ge 2$. Therefore, the expression on the left in (3.6) is increasing in τ_{i-1}, so we need only establish the inequality for $\tau_{i-1} = 0$, which is equivalent to showing that

$$\frac{\zeta_r^2\beta^2}{2(2r+1)^{2q}T_i^{2r+1-2q}} - \log\left(\zeta_r^2\beta^2/(2r+1)^{2q}\right) - \log(2\zeta_r\sqrt{2\pi}) - \frac{6r+3-4q}{2(2r+1-2q)}\log\left(\left(\frac{1}{T_i}\right)^{2r+1-2q}\right) \ge 0.$$

Now

$$\log(2\zeta_r\sqrt{2\pi}) \le \log\left(10\zeta_r^2/(2r+1)^{2q}\right) \le \log\left(10\zeta_r^2\beta^2/(2r+1)^{2q}\right).$$

The last inequality is from $\beta \ge 1$ and the first is because

$$\frac{\zeta_r}{(2r+1)^{2q}} = \frac{r!(2r+1)^{2r+3/2-2q}}{(2r)^{r+1/2}}$$

is increasing in $r \ge 2$ by Lemma 3.1, and at $r = 2 = q$ has the value

$$\frac{5\sqrt{5}}{16} > \frac{5}{8},$$

and at $r = 1 = q$ has the value $(3/2)^{3/2} > 5/8$, and so in either case

$$10\zeta_r^2/(2r+1)^{2q} \geq 10\zeta_r \frac{5}{8} > 2\sqrt{2\pi}\zeta_r.$$

Therefore,

$$\begin{aligned}
&\frac{\zeta_r^2\beta^2}{2(2r+1)^{2q}T_i^{2r-1-2q}} - \log\left(\zeta_r^2\beta^2/(2r+1)^{2q}\right) - \log(2\zeta_r\sqrt{2\pi}) \\
&\quad - \frac{6r+3-4q}{2(2r+1-2q)}\log\left(\left(\frac{1}{T_i}\right)^{2r+1-2q}\right) \\
&\geq \frac{\zeta_r^2\beta^2}{2(2r+1)^{2q}T_i^{2r-1-2q}} - 2\log\left(\zeta_r^2\beta^2/(2r+1)^{2q}\right) - \log(10) \\
&\quad - \frac{6r+3-4q}{2(2r+1-2q)}\log\left(\left(\frac{1}{T_i}\right)^{2r+1-2q}\right) \\
&\equiv \frac{1}{2}xy - 2\log(x) - \frac{6r+3-4q}{2(2r+1-2q)}\log(y) - \log(10) \equiv h(x,y),
\end{aligned}$$

where we made the substitutions

$$x \equiv \left(\frac{\zeta_r\beta}{(2r+1)^q}\right)^2 \geq \min\left\{\frac{3^7}{2^3}, \frac{5^7}{4^4}\right\} = \frac{3^7}{2^3}$$

(using again the fact that $\zeta_r(2r+1)^{-q}$ is increasing in $r \geq 2$ by Lemma 3.1) and

$$y \equiv T_i^{-2r-1+2q} \geq 1.$$

Now

$$\frac{\partial h}{\partial x} = \frac{y}{2} - \frac{2}{x} > 0$$

and for $x \geq 3^7 2^{-3}$ and $y \geq 1$,

$$\begin{aligned}
\frac{\partial h}{\partial y} &= \frac{x}{2} - \frac{6r+3-4q}{2(2r+1-2q)}\frac{1}{y} \\
&\geq \frac{3^7 2^{-3}}{2} - \frac{6r+3-4q}{2(2r+1-2q)} \\
&\geq \begin{cases} 3^7 2^{-4} - 7/2 > 0 & \text{if } r \geq q = 2, \\ 3^7 2^{-4} - 5/2 > 0 & \text{if } r = q = 1. \end{cases}
\end{aligned}$$

Therefore we need only show that

$$h(3^7 \cdot 2^{-3}, 1) \geq 0.$$

But

$$\begin{aligned}
h(3^7 \cdot 2^{-3}, 1) &= \frac{1}{2}3^7 2^{-3} - 2\log(3^7/2^3) - \log(10) \\
&> 136 - 12 - 3 > 0.
\end{aligned}$$

By (3.5), we must have that for $i \geq 1$,

$$\log(1/\delta) \leq \left(\frac{\zeta_r\beta(\tau_{i-1} + T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right)^2. \tag{3.7}$$

For $i = 1$, $\tau_0 = 0$ and (3.7) implies that

$$\log(1/\delta) \le \frac{\zeta_r^2 \beta^2}{(2r+1)^{2q} T_1^{2r+1-2q}},$$

and so the first subinterval has width

$$T_1 \le \frac{\left(\zeta_r^2 \beta^2/(2r+1)^{2q}\right)^{1/(2r+1-2q)}}{\log(1/\delta)^{1/(2r+1-2q)}} \equiv c(r, \delta, \beta). \tag{3.8}$$

We will show by induction that for $j \ge 1$,

$$T_j \le c(r, \delta, \beta) j^{2r}. \tag{3.9}$$

We have established this for $j = 1$ in (3.8); assume that $i \ge 2$ and (3.9) holds for $j < i$. Then

$$\tau_{i-1} = \sum_{j=1}^{i-1} T_j \le c(r, \delta, \beta) \sum_{j=1}^{i-1} j^{2r}.$$

To get a contradiction, suppose that

$$T_i > c(r, \delta, \beta) i^{2r}. \tag{3.10}$$

Then,

$$\begin{aligned}
&\left(\frac{\zeta_r \beta (\tau_{i-1} + T_i/(2r+1))^q}{T_i^{(2r+1)/2}}\right)^2 \\
&\quad \le \frac{\zeta_r^2 \beta^2 \left(c(r, \delta, \beta) \sum_{j=1}^{i-1} j^{2r} + c(r, \delta, \beta) \frac{i^{2r}}{(2r+1)}\right)^{2q}}{c(r, \delta, \beta)^{2r+1} i^{2r(2r+1)}} \\
&\quad = \zeta_r^2 \beta^2 c(r, \delta, \beta)^{2q-2r-1} \frac{\left(\sum_{j=1}^{i-1} j^{2r} + \frac{i^{2r}}{(2r+1)}\right)^{2q}}{i^{2r(2r+1)}} \\
&\quad = \zeta_r^2 \beta^2 \left(\frac{\left(\zeta_r^2 \beta^2/(2r+1)^{2q}\right)^{1/(2r+1-2q)}}{\log(1/\delta)^{1/(2r+1-2q)}}\right)^{2q-2r-1} \frac{\left(\sum_{j=1}^{i-1} j^{2r} + \frac{i^{2r}}{(2r+1)}\right)^{2q}}{i^{2r(2r+1)}} \\
&\quad = \zeta_r^2 \beta^2 \frac{\log(1/\delta)}{\zeta_r^2 \beta^2/(2r+1)^{2q}} \frac{\left(\sum_{j=1}^{i-1} j^{2r} + \frac{i^{2r}}{(2r+1)}\right)^{2q}}{i^{2r(2r+1)}} \\
&\quad = \log(1/\delta) \frac{\left((2r+1) \sum_{j=1}^{i-1} j^{2r} + i^{2r}\right)^{2q}}{i^{2r(2r+1)}} \\
&\quad < \log(1/\delta),
\end{aligned}$$

which contradicts (3.7).

The last inequality used the fact that for $i \ge 2$,

$$\left((2r+1) \sum_{j=1}^{i-1} j^{2r} + i^{2r}\right)^q < i^{r(2r+1)}.$$

To see this, note that for $r = 1 = q$,

$$3\sum_{j=1}^{i-1} j^2 + i^2 = i^3 - \frac{1}{2}i(i-1) < i^3$$

for $i \geq 2$. If $r = 2 = q$, then

$$5\sum_{j=1}^{i-1} j^4 + i^4 = i^5 - \frac{1}{6}i(9i^3 - 10i^2 + 1) < i^5, \quad i \geq 2.$$

It remains to consider the case $r \geq 3$, $q = 2$, for which we use the bound

$$\begin{aligned}(2r+1)\sum_{j=1}^{i-1} j^{2r} + i^{2r} &\leq (2r+1)\int_1^i x^{2r}\,\mathrm{d}x + i^{2r} \\ &= i^{2r+1} - 1 + i^{2r}.\end{aligned}$$

Therefore, it suffices to show that

$$i^{2r+1} - 1 + i^{2r} < \left(i^{r(2r+1)}\right)^{1/2} = i^{r^2+r/2}.$$

But

$$i^{2r+1} - 1 + i^{2r} < 2 \cdot i^{2r+1} \leq i^{2r+2} < i^{r^2+r/2},$$

since

$$2r + 2 < r^2 + r/2$$

for $r \geq 3$.

Therefore, we have established that

$$T_j \leq c(r, \delta, \beta) j^{2r}, \quad j \geq 1. \tag{3.11}$$

Since we must have

$$\sum_{j=1}^{k} T_j \geq \frac{1}{2},$$

where k is the number of function evaluations to the right of the minimizer, we need k to satisfy

$$\frac{1}{2} \leq c(r, \delta, \beta)\sum_{j=1}^{k} j^{2r} \leq c(r, \delta, \beta) k^{2r+1},$$

which implies that

$$k \geq \left(\frac{1}{2c(r, \delta, \beta)}\right)^{1/(2r+1)} \tag{3.12}$$

$$= \left(\frac{1}{2}\left(\frac{(2r+1)^q}{\zeta_r \beta}\right)^{2/(2r+1-2q)}\right)^{1/(2r+1)} \log(1/\delta)^{\frac{1}{(2r+1-2q)(2r+1)}} \equiv \nu(\delta). \tag{3.13}$$

We have now established that for $\delta > 0$,

$$\int_{F_\beta \cap \{n(f) < \nu(\delta)\}} \Delta(f) P_r(\mathrm{d}f) \geq \delta P_r\left(F_\beta \cap \{n(f) < \nu(\delta)\}\right).$$

Now suppose that for $\epsilon \in (0, 1/16]$, an algorithm has average error at most ϵ:

$$\epsilon \geq \int_F \Delta(f) P_r(\mathrm{d}f) \geq \int_{F_\beta \cap \{n(f) < \nu(\delta)\}} \Delta(f) P_r(\mathrm{d}f) \geq \delta P_r\left(F_\beta \cap \{n(f) < \nu(\delta)\}\right). \tag{3.14}$$

Taking $\delta = 4\epsilon$, (3.14) implies that

$$P_r\left(F_\beta \cap \{n(f) < \nu(4\epsilon)\}\right) \le 1/4.$$

Therefore,

$$\begin{aligned} 1/4 &\ge P_r\left(F_\beta \cap \{n(f) < \nu(4\epsilon)\}\right) \\ &= P_r\left(F_\beta\right) + P_r\left(\{n(f) < \nu(4\epsilon)\}\right) - P_r\left(F_\beta \cup \{n(f) < \nu(4\epsilon)\}\right) \\ &\ge 1/2 + P_r\left(\{n(f) < \nu(4\epsilon)\}\right) - 1, \end{aligned}$$

and so

$$P_r\left(\{n(f) \ge \nu(4\epsilon)\}\right) \ge 1/4,$$

which implies that

$$\int_F n(f) P_r(\mathrm{d}f) \ge \frac{1}{4}\nu(4\epsilon).$$

For $\epsilon \le 1/16$,

$$\log(1/(4\epsilon)) \ge \frac{1}{2}\log(1/\epsilon),$$

and so

$$\begin{aligned} \nu(4\epsilon) &= \left(\frac{1}{2}\left(\frac{(2r+1)^q}{\zeta_r\beta}\right)^{2/(2r+1-2q)}\right)^{1/(2r+1)} \log(1/(4\epsilon))^{\frac{1}{(2r+1-2q)(2r+1)}} \\ &\ge \left(\frac{1}{2}\left(\frac{(2r+1)^q}{\zeta_r\beta}\right)^{2/(2r+1-2q)}\right)^{1/(2r+1)} \left(\frac{1}{2}\right)^{\frac{1}{(2r+1-2q)(2r+1)}} \log(1/\epsilon)^{\frac{1}{(2r+1-2q)(2r+1)}}. \end{aligned}$$

We conclude that if an algorithm has average error at most ϵ, then the average cost is at least

$$\begin{aligned} \int_{F_\beta} n(f) P_r(\mathrm{d}f) &\ge \frac{1}{4}\nu(4\epsilon) \\ &\ge \frac{1}{4}\left(\frac{1}{2}\left(\frac{(2r+1)^q}{\zeta_r\beta}\right)^{2/(2r+1-2q)}\right)^{1/(2r+1)} \left(\frac{1}{2}\right)^{\frac{1}{(2r+1-2q)(2r+1)}} \log(1/\epsilon)^{\frac{1}{(2r+1-2q)(2r+1)}} \\ &\equiv C(r, \mathbf{a}, \mathbf{b}) \cdot \log(1/\epsilon)^{\frac{1}{(2r+1-2q)(2r+1)}}. \end{aligned}$$

This completes the proof of Theorem 2.1.

Acknowledgment

A suggestion from an anonymous referee resulted in the improvement (for $r \ge 2$) of the exponent in the lower bound from $1/(2r+1)$.

References

[1] J.M. Calvin, A lower bound on complexity of optimization on the Wiener space, Theoret. Comput. Sci. 383 (2007) 132–139.
[2] E. Novak, K. Ritter, H. Woźniakowski, Average case optimality of a hybrid secant-bisection method, Math. Comp. 64 (1995) 1517–1539.
[3] K. Ritter, Average-Case Analysis of Numerical Problems, Springer, Berlin, 2000.
[4] G. Wasilkowski, On average complexity of global optimization problems, Math. Program. 57 (1992) 313–324.