


On convergence rate of a rectangular partition based global optimization algorithm

James Calvin¹ · Gražina Gimbutienė² ·
William O. Phillips¹ · Antanas Žilinskas² 

Received: 4 October 2017 / Accepted: 28 February 2018 / Published online: 7 March 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract The convergence rate of a rectangular partition based algorithm is considered. A hyper-rectangle for the subdivision is selected at each step according to a criterion rooted in the statistical models based theory of global optimization; only the objective function values are used to compute the criterion of selection. The convergence rate is analyzed assuming that the objective functions are twice- continuously differentiable and defined on the unit cube in d -dimensional Euclidean space. An asymptotic bound on the convergence rate is established. The results of numerical experiments are included.

Keywords Global optimization · Convergence rate · Rectangular partition · Statistical models for global optimization · Bayesian approach · P-algorithm

1 Introduction

Problems of global optimization of continuous multimodal functions are attacked using deterministic [1–7], stochastic [8–12], and heuristic approaches [13–15]. A promising research direction is the hybridization of the advantages of some of the successful methods [16, 17]. Although the proposed algorithms are thoughtfully investigated there remains an important issue that has received relatively little attention, namely the convergence rate problem. In the present paper we focus on this problem continuing the research in [18–21] where the convergence rate was established for algorithms rooted in the approach based on statistical models of global optimization. The convergence rate of global optimization algorithms in a stochastic setting has been considered for certain Gaussian probabilities; see for example

✉ Antanas Žilinskas
antanas.zilinskas@mii.vu.lt

James Calvin
james.m.calvin@njit.edu

¹ Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA

² Institute of Mathematics and Informatics, Vilnius University, Akademijos 4, 08663 Vilnius, Lithuania

[19]. The general question of average-case tractability of multivariate global optimization is mentioned (Open Problem 87) in [22].

We consider the problem of approximating the global minimum f^* of a twice-continuously differentiable function f defined on the d -dimensional unit cube $A = [0, 1]^d$, for $d \geq 2$. We assume that the global minimizer, denoted by x^* , is unique and lies in the interior of the cube. While we assume differentiability, we are interested in algorithms that only evaluate the function and not derivatives. Our interest centers on how the number of function values required to obtain a given quality approximation grows with the dimension of the domain.

The algorithm presented here is similar to a one-dimensional algorithm presented in [20] and a two-dimensional algorithm based on Delaunay triangulations presented in [21]. Those papers presented algorithms with the property that for large enough N (depending on f), the residual error (difference between the smallest function value observed and the global minimum) after N function evaluations is at most

$$c_1 \exp(-c_2 \sqrt{N})$$

for c_1, c_2 depending on dimension (1 or 2) and f . While the algorithm presented in [21] was based on a Delaunay triangulation of the domain, the algorithm considered in this paper is based on a rectangular decomposition. Rectangular decompositions have been employed in previous studies of global optimization, for example in [4, 23–25]. The rectangular partition of the feasible region is inherent also for the interval arithmetic-based global optimization algorithms, see e.g. [26]. One of the advantages of the algorithms using rectangular partition of the feasible region is their suitability for parallelization; however the investigation of parallel versions of the algorithm proposed here is postponed for further research. For a discussion of the alternative simplicial partition of the feasible region in the development of global optimization algorithms we refer to [27].

The algorithm in [21] works for arbitrary domains, but the error bound, which required a certain quality triangulation, could only be proved for bivariate functions. In the present paper we are able to prove an asymptotic error bound for arbitrary dimension. Our main result is that eventually the error is at most

$$c_1(f, d) \exp\left(-c_2(f, d) \sqrt{N}\right),$$

where $c_2(f, d)$ decreases exponentially in dimension d . Recall that we assume that f has continuous partial derivatives up to order 2, and that the global minimizer is unique and is in the interior of the domain.

In the worst-case setting any nonadaptive optimization algorithm has an error of order at least $N^{-2/d}$. It is known that for a convex and symmetric function class such as $C^2([0, 1]^d)$ adaptive algorithms are essentially no more powerful than nonadaptive algorithms for linear problems such as function approximation with error measured using the L_∞ norm ([28]). Global optimization is as difficult as L_∞ approximation (see Theorem 18.24 in [22]). Therefore, $N^{-2/d}$ is an optimal worst-case error bound for optimization on $C^2([0, 1]^d)$. To obtain an error of at most ϵ , we then require on the order of $(1/\epsilon)^{d/2}$ function values. For the algorithm proposed in this paper, we consider the asymptotic decay of the error, showing that for small enough ϵ on the order of $d^{d+3} \log(1/\epsilon)^2$ function values suffice. Because of the asymptotic setting, this does not contradict the worst-case results mentioned above.

With the algorithm defined in this paper, the error converges to zero for a much broader class of functions than the one we consider. For example, the error converges to zero for any continuous function defined on $[0, 1]^d$, though it could be difficult to characterize the rate of convergence for such a broad class.

At the present stage of investigation we have not focused on the details of efficient software development, and have tested a rather simple implementation (without sophisticated data structures) of the algorithm with well known test functions. Our initial guess was that Delaunay triangulation-based algorithm would be considerably more efficient than the algorithm proposed in this paper. This guess was based on the observation that hyper-rectangles have more volume far from the vertices than simplexes. Also, in our Delaunay-based algorithm, each step placed a single function evaluation in a simplex, while subdividing a hyper-rectangle takes a number of evaluations that grows exponentially with dimension. However, at least for the low dimensional problems in our experiments, this guess was not confirmed. The results are presented in the last section of the paper where the testing results of the proposed algorithm and the algorithm based on Delaunay triangulation [21] are presented for comparison.

2 The algorithm

The algorithm operates by decomposing $[0, 1]^d$ into hyper-rectangles as follows. Given a current decomposition, choose one of the hyper-rectangles (according to the maximal value of a criterion to be defined below) and bisect it along the longest axis by evaluating the function at up to 2^{d-1} midpoints of the longest hyper-rectangle edges (some of the midpoints may have been evaluated in previous decompositions).

The criterion of choice of a rectangle is rooted in definition of the P-algorithm described in detail in [29]. The P-algorithm chooses the point for the current computing the an objective function value according to the criterion of maximum of improvement probability

$$\max_{x \in A} \mathbf{P}\{\xi(x) \leq y_{0N} - \varepsilon | \xi(x_i) = y_i, i = 1, \dots, N\},$$

where $\xi(x)$ is a Gaussian random field accepted as a statistical model of objective functions, and y_i values of the objective function computed at previous iterations, and $y_{0N} = \min_{i=1, \dots, N} y_i$, $\varepsilon > 0$. The maximization of the improvement probability can be simplified because of the equality

$$\begin{aligned} & \arg \max_{x \in A} \mathbf{P}\{\xi(x) \leq y_{0N} - \varepsilon | \xi(x_i) = y_i, i = 1, \dots, N\} \\ &= \arg \max_{x \in A} \frac{y_{0N} - \varepsilon - \mu(x|(x_i, y_i), i = 1, \dots, N)}{\sigma(x|(x_i, y_i), i = 1, \dots, N)}, \end{aligned} \tag{1}$$

where $\mu(x|(x_i, y_i), i = 1, \dots, N)$ and $\sigma^2(x|(x_i, y_i), i = 1, \dots, N)$ denote conditional mean and variance correspondingly. The criterion for the selection of a rectangle by the algorithm considered in the present paper is a modification of (1) where conditional mean and variance are replaced by the relevant parameters of the rectangle, and ε is replaced by a parameter $g(\cdot)$ which is defined below; its form of the dependence on the number of iteration is determined aiming at fast asymptotic convergence of the algorithm,

Suppose that the algorithm has evaluated f at N points. Let $s(N)$ denote the number of hyper-rectangles in our partition. Each hyper-rectangle subdivision (iteration) requires up to 2^{d-1} function evaluations; let us note that at the current iteration some function values can be known since computed at previous subdivisions. Therefore, the following inequality is valid

$$s(N) \geq \frac{N}{2^{d-1}}. \tag{2}$$

We next define some constants and functions that are used in the definition of the algorithm. Define

$$q \equiv \frac{3 \cdot 2^{2/3} e^{-1}}{2 \log(2)} \approx 1.27, \quad \lambda \equiv (qd^2/4)^{d/2},$$

and set

$$g(x) \equiv (\lambda x \log(1/x))^{2/d},$$

for $0 < x \leq 1/2$ and $g(1) = qd^2/4$. Note that g is increasing and $g(x) \downarrow 0$ as $x \downarrow 0$.

It will be convenient to index quantities to be defined by the iteration number of the algorithm, which we denote by n , instead of the number of function evaluations, N , which is generally larger. Later we will also describe the error bound in terms of the number of function evaluations N .

After n iterations of the algorithm, let $M_n = \min_{1 \leq i \leq n} f(x_i)$ and denote the error by $\Delta_n = M_n - f(x^*)$. Define the function $L_n : [0, 1]^d \rightarrow \mathbb{R}$ as follows. In the interior of hyper-rectangle R , L_n coincides with the multilinear interpolant of the function values at the vertices of R . On the boundaries of hyper-rectangles set $L_n \equiv M_n$, so that M_n is equal to the global minimum of L_n . Let v_n denote the smallest volume of a hyper-rectangle after n iterations.

Our algorithm will assign a numerical value to each hyper-rectangle in the subdivision. After n iterations of the algorithm, for each hyper-rectangle R_i , $1 \leq i \leq n$, set

$$\rho_i^n \equiv \int_{R_i} \frac{ds}{(L_n(s) - M_n + g(v_n))^{d/2}}, \tag{3}$$

and

$$\bar{\rho}_i^n \equiv \frac{|R_i|}{(\inf_{t \in \text{int}(R_i)} L_n(t) - M_n + g(v_n))^{d/2}}. \tag{4}$$

The criteria (3) and (4) are analogues of reciprocal (1) with opposite sign where $g(v_n)$ is a generalisation of ε . Since the hyper-rectangle boundaries have measure zero, the way we defined L_n on the boundaries is immaterial for the values of the ρ_i^n . Note that the \inf in the denominator of (4) is equal to the minimum of f at the vertices of the hyper-rectangle.

If we are about to subdivide the smallest hyper-rectangle, then

$$\rho_i^n \leq \bar{\rho}_i^n \leq \frac{|R_i|}{g(v_n)^{d/2}} = \frac{v_n}{\lambda v_n \log(1/v_n)} = \frac{1}{\lambda \log(1/v_n)} \leq \frac{1}{\lambda \log(n)}, \tag{5}$$

with the last inequality following from the fact that the smallest hyper-rectangle in the subdivision has volume $v_n \leq 1/n$.

A more formal description of the algorithm follows. In the description, N is the total number of function evaluations to make, i is the current number of hyper-rectangles, and j is the number of function evaluations that have been made. Let v , the volume of the smallest hyper-rectangle, take initial value 1 and let M , the minimum of the observed function values, take initial value the minimum of the function values at the vertices of $[0, 1]^d$.

The algorithm comprises the following steps.

1. We start with the hyper-rectangle $[0, 1]^d$ and with the function values at all 2^d vertices. Set i , the number of hyper-rectangles, to 1 and set j , the number of function evaluations, to 2^d .

2. For each hyper-rectangle R_k in the current collection $\{R_1, R_2, \dots, R_i\}$, compute ρ_k^i , keeping track of the rectangle with the lowest index γ that has the maximal value of ρ_γ^i .
3. From the hyper-rectangle R_γ with maximal value ρ_γ^i , form two new hyper-rectangles as follows. Suppose that

$$R_\gamma = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_d, b_d],$$

and that j is the smallest index with $b_j - a_j \geq b_i - a_i$ for $1 \leq i \leq d$. The two new hyper-rectangles are

$$R'_\gamma \equiv [a_1, b_1] \times \dots \times [a_j, (a_j + b_j)/2] \times \dots \times [a_d, b_d]$$

and

$$R''_\gamma \equiv [a_1, b_1] \times \dots \times [(a_j + b_j)/2, b_j] \times \dots \times [a_d, b_d].$$

Evaluate f at the (at most) 2^{d-1} points

$$\left(c_1, \dots, \frac{a_j + b_j}{2}, \dots, c_d \right), \quad c_i \in \{a_i, b_i\}, \quad 1 \leq i \leq d, \quad i \neq j,$$

which have not been evaluated previously.

If $|R_\gamma| = v$, then set $v \leftarrow v/2$, and if the smallest of the new function values is less than the previously observed minimum M , then set M to that new smallest value. Update the number of hyper-rectangles $i \leftarrow i + 1$ and increment j by the number of new function evaluations.

4. If $j < N$, return to step 2.

Typical subdivisions for the two-dimensional case are depicted in Fig. 1 in Sect. 5. In the graphs, each vertex of the mesh corresponds to a point where the function has been evaluated. Each hyper-rectangle has been evaluated at its four vertices, as well as possibly other points on the perimeter. The function L_n agrees on the interior of a hyper-rectangle with the multilinear interpolation of the four vertices, disregarding additional vertices that may lie on the perimeter. The aspect ratio of each hyper-rectangle is either 1 or 2.

Let $\hat{F} \subset C^2([0, 1]^d)$ denote the set of functions that have a unique global minimizer $x^* \in]0, 1[^d$. Our error bound will hold for functions in \hat{F} . It is essential for our result that the function has only a finite number of isolated global minimizers; the uniqueness assumption is to simplify the presentation.

Denote the matrix of second-order partial derivatives at the minimizer by $D^2 f(x^*)$, assumed positive definite.

The error bound will depend on the regularity of the function as measured by a seminorm that we now define. For a compact set $K \subset [0, 1]^d$ and $f \in C^2(K)$, define the seminorm

$$\|D^2 f\|_{\infty, K} \equiv \sup_{x \in K} \sup_{\substack{u_1, u_2 \in \mathbb{R}^d \\ \|u_j\|=1}} |D_{u_1} D_{u_2} f(x)|,$$

where $D_y f$ is the directional derivative of f in the direction y . This is a measure of the maximum size of the second derivative of f over K .

Our main result on the convergence rate of the error for this algorithm follows. Let note, that the convergence of the original version of the P-algorithm for continuous objective functions we refer to ([29]) where the improvement threshold ε , analogous to $g(\cdot)$ in (3), (4), can be any positive constant. With such a weaker assumption, however, the convergence rate remains not known.

Theorem 1 Suppose that $f \in \hat{F}$. There is a number $n_0(f)$ such that for $n \geq n_0(f)$, the residual error after n iterations of the algorithm satisfies

$$\Delta_n(f) \equiv \min_{1 \leq i \leq n} f(x_i) - f(x^*) \leq \frac{4^{2/d}}{8} \|D^2 f\|_{\infty, [0,1]^d} (q \cdot d) \exp(-\sqrt{n}\beta(f, d)),$$

where

$$\beta(f, d) = \left(\frac{2\sqrt{\pi} (\det(D^2 f(x^*)))^{1/2}}{9ed^{(d+3)/2}(\pi eq)^{d/2}} \exp\left(-\frac{1}{2} - \frac{16}{q}\right) \right)^{1/2}.$$

The number $n_0(f)$ after which the bound applies will be the maximum of several other $n_i(f)$'s that will be defined in the proof of the theorem. We note here that $n_0(f)$ depends (exponentially) on the value of the seminorm $\|D^2 f\|_{\infty, [0,1]^d}$.

Note that the limiting error decreases when the determinant of the second derivative at the minimizer increases; a larger second derivative allows the search effort to concentrate more around the minimizer. The convergence rate in terms of the number of function evaluations n is quite fast, but the term $\beta(f, d)$ decreases exponentially fast as d increases: as $d \rightarrow \infty$, $\beta(f, d) = O(d^{-(d+3)/2})$. Therefore, this algorithm does not escape the curse of dimension.

We focus on the information complexity; that is, the cost is the number of function evaluations. The worst-case combinatorial cost of the algorithm grows quadratically with the number of iterations.

3 Interpolation error bounds

We will use results about interpolation from [30].

Lemma 1 Consider a hyper-rectangle $R = \{x \in [0, 1]^d : a_i \leq x_i \leq b_i, 1 \leq i \leq d\}$, and let L denote the multilinear function that interpolates f at the vertices of the hyper-rectangle. We have the error bound

$$\begin{aligned} |f(x) - L(x)| &\leq \frac{1}{2} \sum_{i=1}^d (x_i - a_i)(b_i - x_i) \|D^2 f\|_{\infty, R} \\ &= \frac{1}{2} \frac{\sum_{i=1}^d (x_i - a_i)(b_i - x_i)}{|R|^{2/d}} \|D^2 f\|_{\infty, R} |R|^{2/d}, \end{aligned}$$

which implies the bound

$$\begin{aligned} \max_{x \in R} |f(x) - L(x)| &\leq \frac{1}{8} \sum_{i=1}^d (b_i - a_i)^2 \|D^2 f\|_{\infty, R} \\ &= \frac{1}{8} \frac{\sum_{i=1}^d (b_i - a_i)^2}{|R|^{2/d}} \|D^2 f\|_{\infty, R} |R|^{2/d}. \end{aligned}$$

Proof See [30]. □

The quantity

$$\frac{\sum_{i=1}^d (b_i - a_i)^2}{|R|^{2/d}} = \frac{\sum_{i=1}^d (b_i - a_i)^2}{\left(\prod_{i=1}^d (b_i - a_i)\right)^{2/d}}$$

is a measure of how “bad” the hyper-rectangle is from the point of view of interpolation error; that is, how far it is from a cube. Since by construction of the algorithm the widths are within a factor of 2 of each other, there is some number $1 \leq k \leq d$ and a positive number w such that k of the dimensions have width w and $d - k$ have width $2w$, and

$$\frac{\sum_{i=1}^d (b_i - a_i)^2}{\left(\prod_{i=1}^d (b_i - a_i)\right)^{2/d}} = \frac{kw^2 + (d - k)(2w)^2}{(2^{d-k}w^d)^{2/d}} = \frac{(4d - 3k)}{2^{2(d-k)/d}} = \left(d - \frac{3}{4}k\right) 2^{2k/d}.$$

For $k = d$ we get the value d and for $k = 1$ we get the value $(d - 3/4)2^{2/d}$. Maximizing

$$d(1 - 3x/4)2^{2x}$$

over $x \in]0, 1[$ gives a maximum value of

$$\frac{3 \cdot 2^{2/3} e^{-1}}{2 \log(2)} d \equiv q \cdot d \approx 1.27d.$$

Therefore, the quality metric ranges from d up to about $1.27d$, and the second bound in Lemma 1 implies that

$$\max_{x \in R} |f(x) - L(x)| \leq \frac{1}{8} q \cdot d \|D^2 f\|_{\infty, R} |R|^{2/d}. \tag{6}$$

Also note that

$$d \leq \frac{\sum_{i=1}^d (b_i - a_i)^2}{\left(\prod_{i=1}^d (b_i - a_i)\right)^{2/d}} \leq qd. \tag{7}$$

In addition to $\|D^2 f\|_{\infty, [0,1]^d}$, we introduce a functional that is in a sense a measure of how difficult it is to obtain an ϵ approximation of f . For $\epsilon > 0$, define

$$\mathcal{I}_f(\epsilon) \equiv \int_{[0,1]^d} \frac{dx}{(f(x) - f^* + \epsilon)^{d/2}}.$$

The following lemma, proved in “Appendix A”, generalizes Lemma 3.2 of [21] to arbitrary dimension.

Lemma 2

$$\lim_{\epsilon \downarrow 0} \frac{\mathcal{I}_f(\epsilon)}{\log(1/\epsilon)} = \frac{d(2\pi)^{d/2}}{2\Gamma(1 + d/2)} \cdot (\det(D^2 f(x^*)))^{-1/2} \equiv \alpha(f, d).$$

4 Proof of Theorem 1

We begin with a rough outline the steps of the proof of Theorem 1, which will be based on a series of lemmas. The proofs of the lemmas appear in “Appendix B”.

Lemma 4 establishes that the hyper-rectangle that contains the global minimizer cannot be much larger than the smallest rectangle at that iteration; this fact allows us to bound the error in terms of v_n , the volume of the smallest rectangle. Recalling that ρ_i^n is the metric used to choose the next rectangle to split, Lemmas 5 and 6 establish that

$$\frac{1}{n} \sum_{i=1}^n \rho_i^n \sim \frac{1}{\log(1/v_n)},$$

while Lemma 7 shows that

$$\sum_{i=1}^n \rho_i^n \sim \mathcal{I}_f(g(v_n)).$$

With this, Lemma 2 allows us to bound $g(v_n)$, and thereby the error. All of the analysis relies on Lemma 3, which gives an upper bound for ρ^n .

We consider a fixed function $f \in \hat{F}$. The subscript n will indicate the number of iterations of the algorithm (not the number of function evaluations). We assume that $n \geq n_2(f)$, where $n_2(f)$ is the first time after $n_1(f)$ that the smallest rectangle is about to be split and

$$n_1(f) = \inf \left\{ n : n \geq \exp \left(\left(d \|D^2 f\|_{\infty, [0,1]^d} \right)^{d/2} \right) \right\}. \tag{8}$$

(Later we will further restrict n .)

Lemma 3 For all $n \geq n_2(f)$,

$$\rho^n \equiv \max\{\rho_1^n, \rho_2^n, \dots, \rho_n^n\} \leq \frac{4}{\lambda \log(1/v_n)} \leq \frac{4}{\lambda \log(n)}.$$

Lemma 4 Let v_n^* denote the volume of the hyper-rectangle containing the minimizer x^* . For $n \geq n_2$,

$$v_n^* \leq 4v_n \tag{9}$$

and

$$\Delta_n \equiv M_n - f(x^*) \leq 2d^{-2}g(v_n). \tag{10}$$

The following lemma gives a lower bound for ρ^n that complements the upper bound provided by Lemma 3.

Lemma 5 For $n \geq n_2(f)$,

$$\rho^n \equiv \max_{i \leq n} \rho_i^n \geq \frac{\exp(-16/q)}{\lambda \log(1/v_n)}.$$

Lemma 6 For $n \geq n_3(f) \equiv 2n_2(f)$, we have the bounds

$$\frac{1}{n} \sum_{i=1}^n \rho_i^n \leq \rho^n \leq \frac{4}{\lambda \log(1/v_n)} \tag{11}$$

and

$$\frac{1}{n} \sum_{i=1}^n \rho_i^n \geq \frac{1}{2} \exp \left(-\frac{1}{2} - \frac{16}{q} \right) \frac{1}{\lambda \log(1/v_n)}. \tag{12}$$

The following lemma allows us to approximate the average of the $\{\rho_i^n\}$ in terms of $\mathcal{I}_f(\cdot)$, for which we have the approximation given by Lemma 2.

Lemma 7 For $n \geq n_2(f)$,

$$\frac{1}{n} \mathcal{I}_f(g(v_n)) \leq \sum_{i=1}^n \rho_i^n \leq \frac{9}{4} \mathcal{I}_f(g(v_n)). \tag{13}$$

We are now ready to prove Theorem 1. From Lemma 6 we have the bounds

$$\frac{1}{n} \sum_{i=1}^n \rho_i^n \leq \frac{4}{\lambda \log(1/v_n)} \tag{14}$$

and

$$\frac{1}{n} \sum_{i=1}^n \rho_i^n \geq \frac{1 \exp(-1/2 - 16/q)}{2 \lambda \log(1/v_n)}. \tag{15}$$

Therefore

$$\begin{aligned} \frac{1}{8} \exp(-1/2 - 16/q) \frac{1/2}{n} \mathcal{I}_f(g(v_n)) &\leq \frac{1}{8} \exp(-1/2 - 16/q) \frac{1}{n} \sum_{i=1}^n \rho_i^n \\ &\leq \frac{1}{2} \exp(-1/2 - 16/q) \frac{1}{\lambda \log(1/v_n)} \leq \frac{1}{n} \sum_{i=1}^n \rho_i^n \leq \frac{9/4}{n} \mathcal{I}_f(g(v_n)), \end{aligned}$$

using the first inequality in (13), (14), (15), and the second inequality in (13), respectively. We focus on the inequalities

$$\begin{aligned} \frac{1}{16} \exp(-1/2 - 16/q) \frac{1}{n} \mathcal{I}_f(g(v_n)) &\leq \frac{1}{4} \exp(-1/2 - 16/q) \frac{1}{\lambda \log(1/v_n)} \\ &\leq \frac{9/4}{n} \mathcal{I}_f(g(v_n)). \end{aligned}$$

There is a number $n_4(f)$ such that, by Lemma 2, $n \geq n_4(f)$ implies that

$$\frac{1}{2} \log(1/g(v_n))\alpha(d) \leq \mathcal{I}_f(g(v_n)) \leq 2 \log(1/g(v_n))\alpha(d).$$

Replace $\mathcal{I}_f(g(v_n))$ by $\log(1/g(v_n))$ (times appropriate factor) in the extreme terms of (??) to get

$$\begin{aligned} \frac{1}{32} \exp(-1/2 - 16/q) \frac{1}{n} \log(1/g(v_n))\alpha(d) \\ \leq \frac{1}{4} \exp(-1/2 - 16/q) \frac{1}{\lambda \log(1/v_n)} \leq \frac{9/4}{n} 2 \log(1/g(v_n))\alpha(d). \end{aligned}$$

Now substitute the value

$$\log(1/g(v_n)) = -\log(qd^2/4) + \frac{2}{d} (\log(1/v_n) - \log \log(1/v_n))$$

to obtain the bounds

$$\begin{aligned} \frac{1}{32} \exp(1/2 - 16/q) \frac{1}{n} \left(-\log(qd^2/4) + \frac{2}{d} (\log(1/v_n) - \log \log(1/v_n)) \right) \alpha(d) \\ \leq \frac{1}{4} \exp(-1/2 - 16/q) \frac{1}{\lambda \log(1/v_n)} \\ \leq \frac{9/4}{n} 2 \left(-\log(qd^2/4) + \frac{2}{d} (\log(1/v_n) - \log \log(1/v_n)) \right) \alpha(d). \end{aligned}$$

This implies that eventually, say for $n \geq n_5(f)$,

$$\log(1/v_n) \geq \left(\frac{dn}{36\alpha(d)\lambda} \exp\left(-\frac{1}{2} - \frac{16}{q}\right) \right)^{1/2},$$

which entails that

$$v_n \leq \exp \left(- \left(\frac{dn}{36\alpha(d)\lambda} \exp \left(-\frac{1}{2} - \frac{16}{q} \right) \right)^{1/2} \right).$$

Then

$$\begin{aligned} M_n - f^* &\leq \frac{4^{2/d}}{8} \|D^2 f\|_{\infty, [0,1]^d} (q \cdot d) v_n^{2/d} \\ &\leq \frac{4^{2/d}}{8} \|D^2 f\|_{\infty, [0,1]^d} (q \cdot d) \exp \left(- \left(\frac{dn}{36\alpha(d)\lambda} \exp \left(-\frac{1}{2} - \frac{16}{q} \right) \right)^{1/2} \frac{2}{d} \right). \end{aligned}$$

For $n \geq n_0(f) \equiv \max\{n_2(f), n_3(f), n_4(f), n_5(f)\}$,

$$\begin{aligned} \Delta_n &\leq \frac{4^{2/d}}{8} \|D^2 f\|_{\infty, [0,1]^d} (q \cdot d) \exp \left(- \left(\frac{dn}{36\alpha(d)\lambda} \exp \left(-\frac{1}{2} - \frac{16}{q} \right) \right)^{1/2} \frac{2}{d} \right) \\ &= \frac{4^{2/d}}{8} \|D^2 f\|_{\infty, [0,1]^d} (q \cdot d) \exp \left(-\sqrt{n} \left(\frac{d}{36\alpha(d)\lambda} \frac{4}{d^2} \exp \left(-\frac{1}{2} - \frac{16}{q} \right) \right)^{1/2} \right). \end{aligned} \tag{16}$$

Substituting the value of α from Lemma 2, and using the bound (from Stirling’s approximation)

$$\Gamma(1 + d/2) \geq \sqrt{2\pi} (d/2)^{(1+d)/2} e^{-(1+d/2)},$$

we obtain

$$\begin{aligned} \frac{d}{36\alpha(d)\lambda} \frac{4}{d^2} &= \frac{1}{9\alpha(d)\lambda d} = \frac{1}{9 \frac{d(2\pi)^{d/2}}{2\Gamma(1+d/2)} \cdot (\det(D^2 f(x^*)))^{-1/2} \lambda d} \\ &= \frac{2\Gamma(1 + d/2) (\det(D^2 f(x^*)))^{1/2}}{9d(2\pi)^{d/2} (qd^2/4)^{d/2} d} \\ &\geq \frac{2\sqrt{2\pi} (d/2)^{(1+d)/2} e^{-(1+d/2)} (\det(D^2 f(x^*)))^{1/2}}{9d(2\pi)^{d/2} (qd^2/4)^{d/2} d} \\ &= \frac{2\sqrt{\pi} (\det(D^2 f(x^*)))^{1/2}}{9ed^{3/2} (e\pi qd)^{d/2}}. \end{aligned}$$

Combined with (16) this yields the bound

$$\begin{aligned} \Delta_n &\leq \frac{4^{2/d}}{8} \|D^2 f\|_{\infty, [0,1]^d} (q \cdot d) \\ &\quad \times \exp \left(-\sqrt{n} \left(\frac{2\sqrt{\pi} (\det(D^2 f(x^*)))^{1/2}}{9ed^{3/2} (e\pi qd)^{d/2}} \exp \left(-\frac{1}{2} - \frac{16}{q} \right) \right)^{1/2} \right) \\ &= \frac{4^{2/d}}{8} \|D^2 f\|_{\infty, [0,1]^d} (q \cdot d) \exp \left(-\sqrt{n} \beta(f, d) \right) \end{aligned}$$

where

$$\beta(f, d) = \left(\frac{2\sqrt{\pi} (\det(D^2 f(x^*)))^{1/2}}{9ed^{(d+3)/2} (\pi eq)^{d/2}} \exp \left(-\frac{1}{2} - \frac{16}{q} \right) \right)^{1/2}.$$

Finally, note that this is the expression for the error after n iterations of the algorithm, which corresponds to as many as $2^{d-1}n$ function evaluations. After a total of N function evaluations, an upper bound on the error is given by

$$\frac{4^{2/d}}{8} \|D^2 f\|_{\infty, [0,1]^d} (q \cdot d) \exp\left(-\sqrt{N}2^{-(d-1)/2} \beta(f, d)\right).$$

This completes the proof of Theorem 1.

5 Numerical examples

The main goal of the research presented in the paper was the evaluation of convergence rate of the algorithm which generalizes the previously investigated algorithms for the case of arbitrary d . However, we do not expect the upper bound to be tight for typical test functions since the worst case objective function is considered. Therefore, including of some results of preliminary numerical experiments seemed reasonable to illustrate relation of the asymptotic convergence rate to the practical performance. The theoretical overestimate of the actual error is commented after the presentation of the experimental results.

A simple C++ implementation of the algorithm, applicable to the problems of low dimensionality, was used. Further, we plan to implement parallel version of the algorithm and perform the so called competitive software testing according to the widely accepted methodology [13, 31, 32].

The algorithm consists of the following main parts: computation of the selection criterion (3), selection and subdivision of the hyper rectangle, and storing of the new hyper rectangles produced. The computational burden needed for the implementation of the current iteration of algorithm depends on the dimensionality d and on the number of iteration. The dimensionality influences computing time of the criterion (3). The amount of stored information grows with the number of iteration, and correspondingly grows computer time needed for its management. Therefore, for the problems of higher dimensionality, the parallel implementation of the algorithm is planned. Meanwhile, the numerical results below are obtained using the straightforward C++ implementation.

The computation of criterion (3) reduces to the numerical integration over the hyper rectangle. In the two-dimensional case the integral can be computed using a combination of analytic and numerical techniques. Let $\mathbf{p}^i = (p_0^i, p_1^i), i = 0, \dots, 3$, be the rectangle vertices, numbered such that $p_0^0 = p_0^1 < p_0^2 = p_0^3$ and $p_1^0 = p_1^2 < p_1^1 = p_1^3$. The corresponding objective function values are then $f_i = f(\mathbf{p}^i), i = 0, \dots, 3$. The bi-linear interpolation of the objective value at an arbitrary point $\mathbf{x} = (x_0, x_1), x_j \in [a_j, b_j], a_j = p_j^0, b_j = p_j^3, j = 0, 1$, can then be expressed as

$$\begin{aligned} L(\mathbf{x}) &= L(x_0, x_1) = A(x_1)x_0 + B(x_1), \\ A(x_1) &= \frac{\frac{x_1-a_1}{b_1-a_1}(f_0 - f_1 - f_2 + f_3) + (f_2 - f_0)}{b_0 - a_0}, \\ B(x_1) &= \frac{\frac{x_1-a_1}{b_1-a_1}(-f_0p_0^3 + f_1p_0^2 + f_2p_0^1 - f_3p_0^0) + (f_0p_0^3 - f_2p_0^1)}{b_0 - a_0}. \end{aligned} \tag{17}$$

Then

$$\rho_i^n = \int_{R_i} \frac{d\mathbf{x}}{L(\mathbf{x}) - M_n + g(v_n)} = \int_{a_1}^{b_1} \int_{a_0}^{b_0} \frac{dx_0}{L(x_0, x_1) - M_n + g(v_n)} dx_1$$

$$= \int_{a_1}^{b_1} \int_{a_0}^{b_0} \frac{dx_0}{A(x_1)x_0 + C(x_1)} dx_1 = \int_{a_1}^{b_1} u(x_1) dx_1, \tag{18}$$

where

$$u(x_1) = \begin{cases} l \frac{1}{A(x_1)} (\ln(A(x_1)b_0 + C(x_1)) - \ln(A(x_1)a_0 + C(x_1))), & \text{if } A(x_1) \neq 0, \\ \frac{1}{C(x_1)} (b_0 - a_0), & \text{if } A(x_1) = 0, \end{cases}$$

$$C(x_1) = B(x_1) - M_n + g(v_n). \tag{19}$$

Note that the function $u(x_1)$ is continuous at the point $\hat{x}_1 = \frac{f_0 - f_2}{f_0 - f_1 - f_2 + f_3}$, $\hat{x}_1 \in [a_1, b_1]$, that gives $A(\hat{x}_1) = 0$.

In the experiments for the two dimensional case we compute ρ_i^n as a single dimensional integral using Eqs. 17–19. The single dimensional integral is computed numerically using the GNU Scientific Library [33] routine *gsl_integration_cquad* requiring a relative error smaller than 10^{-7} . We call the resulting implementation *Rect-A*. For comparison we also include the integration over the rectangle R_i using the classical Monte Carlo method (GNU Scientific Library routine *gsl_monte_plain_integrate*) with $M = 10,000$ evaluations of the two dimensional integrand (referred to as *Rect-MC*). Finally, we present a simplified version of computing ρ_i^n for dimension $d = 2$ as

$$\rho_i^n \approx \frac{|R_i|}{L(\mathbf{x}) - M_n + g(v_n)}, x_j = (a_j + b_j)/2, j = 0, \dots, d. \tag{20}$$

We call this algorithm version *Rect-1*. In all implementations the function evaluation database is employed, so that the objective is never evaluated twice at the same location.

For comparison, two following algorithms were used. A recent algorithm [21], based on the Delaunay triangulation of the optimization region and referred to as *Simpl* in this section, is also implemented in the C++ programming language and uses the 2D and 3D Delaunay triangulation code from the CGAL library [34]. The comparison with a non-adaptive algorithm presents interest especially because of mentioned in the introduction different convergence rates. The non-adaptive algorithm is implemented using the so called LP_τ sequences [35], and is worst case quasi optimal since it generates points for computing objective function values nearly uniformly in the feasible region. The results of that algorithm are referred to as *Sobol*.

We start by illustrating the partition of the feasible regions by the algorithms *Rect-A* during minimization of the following objective functions:

- The six-hump camel back function [36]:

$$f(\mathbf{x}) = (4 - 2.1x_1^2 + x_1^4/3)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2, \mathbf{x} \in [-3, 3] \times [-2, 2]. \tag{21}$$

- The Branin function [37]:

$$f(\mathbf{x}) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + 5\frac{x_1}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10, \mathbf{x} \in [-5, 10] \times [0, 15]. \tag{22}$$

Figure 1 shows the partition after $n = 3000$ objective function evaluations. It can be seen that the space is roughly explored everywhere, however, the trial points are positioned more densely in the regions of potential global minima.

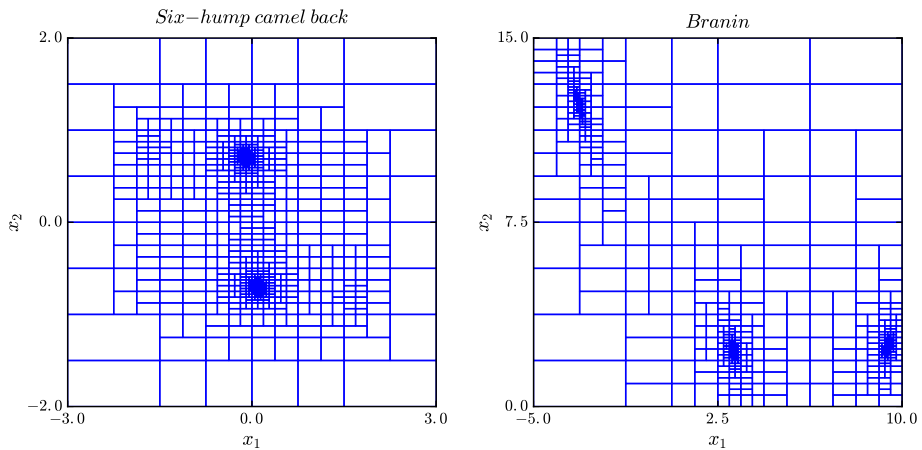


Fig. 1 Examples of the partition of feasible region for the six-hump camel back function (left) and Branin (right) test functions after $n = 3000$ function evaluations

Next, we present the numerical experiments aimed at the comparison of the proposed algorithm with the algorithm of similar destination which implements the partition based on the Delaunay triangulation of the feasible region [21]. The latter was tested using a set of 2-dimensional objective functions from [38], that were used as well as in numerous other papers investigating Lipschitz optimization algorithms.

Table 1 presents results of minimization of test functions from [38] using the same numeration. The optimization error is presented (on a logarithmic scale) as $\delta_n = -\log_{10}(M_n - f_{min})$, after $n = 1000, 3000$ function evaluations for each of the algorithms *Rect-A*, *Rect-I*, *Rect-MC*, *Simpl* [21] and *Sobol*. Symbol ∞ means that zero error has been achieved within machine precision. The numbers in the left column have been copied from the function definitions in [38] on p. 468. The values for the algorithm [21] have been copied from the Table 1 of the paper [21].

It can be seen that the various versions of the proposed algorithm perform comparably with the algorithm [21]. Moreover, there appears to be no significant difference between implementations *Rect-A* and *Rect-MC*, suggesting that the Monte Carlo technique could be used to approximate the integral in the higher dimensions. As could be expected, the results of the non-adaptive *Sobol* algorithm are the worst. The minimization of the function No. 13 was similarly problematic to all algorithms, as the global minimum of this function is hidden in a small basin behind a sharp spike.

The minimization results for objective functions defined over 3-dimensional regions are shown in Table 2. The functions are also taken from [38]. Since the analytic expression of the integral is not available in the 3-dimensional case, we included only versions *Rect-I* and *Rect-MC* ($M = 100000$ evaluations of the integrand per hyper-rectangle) of the proposed algorithm. The table shows that the proposed algorithm performs better than other algorithms.

The number of iteration needed to find an appropriate approximation of minimum can depend on a coincidental match of properties of an algorithm and objective function. To reduce that factor, the performance profiles below are evaluated using a sample of randomly generated problems where the same objective function is minimized over a feasible region which is randomly (with uniform distribution) scrolled for 10% of its original extend. The average and maximum errors of 100 random problems were recorded and plotted against

Table 1 Testing results with functions from [38] defined over a 2-dimensional region

f(x)	Rect-A		Rect-1		Rect-MC		Simpl [21]		Sobol	
	n		n		n		n		n	
	1000	3000	1000	3000	1000	3000	1000	3000	1000	3000
1	5.80	6.84	5.30	5.80	5.80	6.84	5.30	6.86	0.97	1.95
2	6.93	11.00	6.93	6.93	6.93	11.00	6.95	∞	1.82	2.01
3	11.88	11.88	13.80	14.30	12.91	14.30	∞	∞	1.36	4.54
3.1	11.88	11.88	14.16	14.16	12.31	14.30	∞	∞	3.44	3.44
3.2	11.88	11.88	14.16	14.16	12.31	14.30	∞	∞	1.80	1.80
3.3	11.87	11.87	13.43	14.30	13.37	14.30	∞	∞	1.55	2.33
4	∞	∞	∞	∞	∞	∞	2.97	3.29	∞	∞
5	6.01	6.01	4.14	6.01	6.01	6.01	6.01	6.01	0.76	1.37
6	4.49	6.21	4.49	4.49	4.49	6.21	4.25	4.25	-0.48	-0.48
7	2.81	12.38	11.83	12.38	2.88	12.38	25.56	29.61	1.31	1.31
8	6.23	8.53	6.12	6.23	8.53	8.53	6.23	8.53	0.99	0.99
9	∞	∞	∞	∞	∞	∞	∞	∞	0.50	1.28
9.1	∞	∞	-1.91	-1.91	∞	∞	∞	∞	-1.33	0.47
9.2	10.61	10.61	-1.91	-1.91	∞	∞	∞	∞	-0.23	-0.23
9.3	∞	∞	∞	∞	∞	∞	∞	∞	2.15	2.15
10	12.27	12.27	∞	∞	∞	∞	∞	∞	3.39	3.39
11	9.17	14.10	9.17	∞	9.17	∞	8.04	8.04	2.86	3.00
12	6.02	9.04	6.02	9.04	6.02	9.04	2.38	2.38	3.22	4.24
13	1.51	1.59	1.51	1.56	1.51	1.59	1.37	1.58	1.15	1.38

Optimization stops after a predefined number of function evaluations n

Table 2 Testing results with functions from [38] defined over a 3-dimensional region

f(x)	Rect-1		Rect-MC		Simpl [21]		Sobol	
	n		n		n		n	
	3000	5000	3000	5000	3000	5000	3000	5000
20	∞	∞	∞	∞	∞	∞	1.86	1.86
21	7.83	11.06	7.83	11.05	4.45	4.61	1.82	1.82
22	8.49	12.19	9.79	13.30	4.11	4.52	1.52	2.13
23	0.84	1.20	0.88	1.20	0.94	0.94	1.44	1.44
24	10.53	13.57	10.53	12.21	3.88	3.88	1.61	1.61
25	4.00	4.00	4.00	4.00	1.18	1.28	1.20	1.20
26	∞	∞	∞	∞	-0.69	-0.60	-0.81	-0.81

Optimization stops after a predefined number of function evaluations n

number of function evaluations n . The logarithmic scale for the presentation of graphs was used: $\delta_n = -\log_{10}(M_n - f_{min})$, where M_n denotes the best function value found up to the n th function evaluation and f_{min} is the actual global minimum value. In Fig. 2 are presented graphs for (2-dimensional) six-hump camel back and Branin functions; the version of algorithm *Rect-A* was used, partition of the feasible region is shown in Fig. 1.

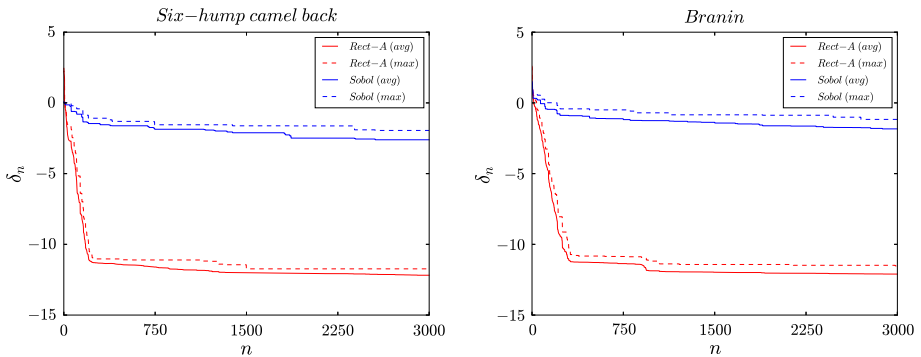


Fig. 2 The maximum and average error after n function evaluations of algorithms *Rect-A* and *Sobol* with 100 perturbed cases of the 2-dimensional test problems: the six-hump camel back function (left) and the Branin function (right)

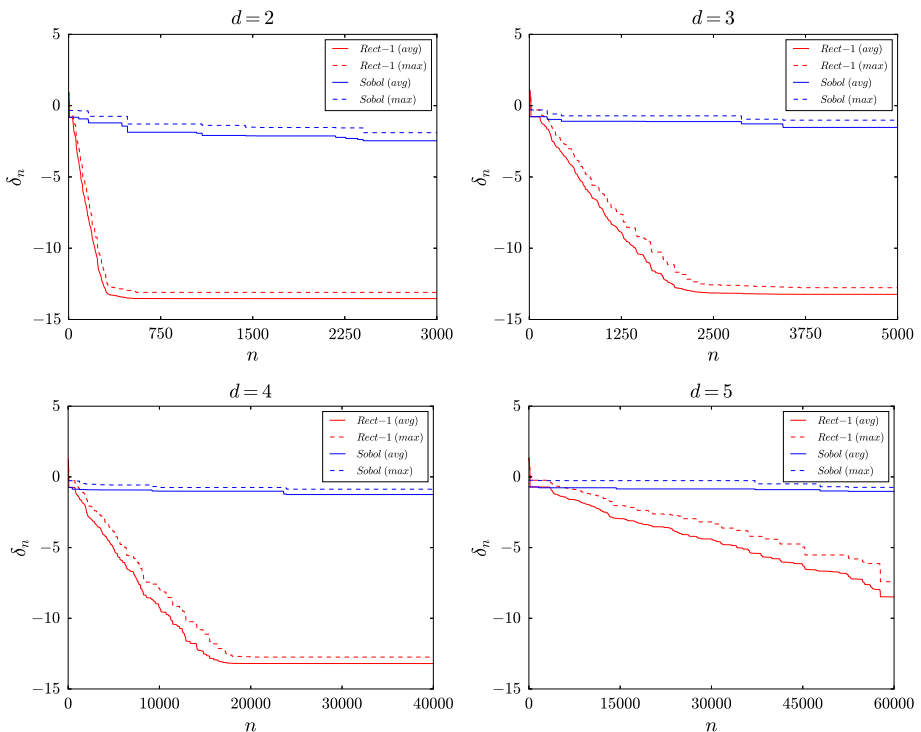


Fig. 3 The maximum and average error after n function evaluations of algorithms *Rect-1* and *Sobol* with 100 perturbed cases of the Rastrigin test problem (Eq. 23) for $d = 2, 3, 4, 5$

To illustrate the influence of the dimension similar experiments were performed with test problems up to dimension 5. Let us note, that contemporary publications on algorithms of similar destination, i.e supposed for expensive black box type problems, typically present experiments with objective functions up to dimensionality five, six [16,39]. The test problems were constructed extending the two dimensional Rastrigin function [15] for dimension d :

Table 3 The average/maximum number of objective function evaluations, performed by algorithms *Rect-1* and *Sobol*, before $-\delta_n$ drops below ι , for the 100 perturbed cases of the Rastrigin test problem (Eq. 23) in dimensions $d = 2, 3, 4, 5$

Alg.	t	Rastrigin function			
		d = 2	d = 3	d = 4	d = 5
Rect-1	− 2	51/63	287/356	1838/2231	10,017/13,725
	− 5	101/117	668/827	4935/6350	34,186/45,297
	− 10	215/239	1443/1658	11,212/12,892	> 60,000/> 60,000
Sobol	− 2	1089/> 3000	> 5000/> 5000	> 40,000/> 40,000	> 60,000/> 60,000
	− 5	> 3000/> 3000	> 5000/> 5000	> 40,000/> 40,000	> 60,000/> 60,000
	− 10	> 3000/> 3000	> 5000/> 5000	> 40,000/> 40,000	> 60,000/> 60,000

$$f(\mathbf{x}) = \sum_{i=1}^d x_i^2 - \cos(\pi x_1) - \cos(\pi x_2), \quad \mathbf{x} \in [-2, 2]^d. \tag{23}$$

Such an extension keeps fixed number of local minima equal to 5. The global minimum point is the d dimensional vector with zero components. The results for $d = 2, 3, 4, 5$ are presented in Fig. 3, and are also quantified in the Table 3.

Now we can compare the upper bound given by Theorem 1 with the actual error estimated experimentally. Our analysis uses the inequality $v_n \leq 1/n$ (v_n is the volume of the smallest rectangle after n iterations), for example in the last inequalities in the proof of Lemma 4. If f is near its minimum over much of the domain, then $v_n \approx 1/n$ (think of a nearly constant function with a tiny, locally quadratic valley in the center of the domain). For our test functions, v_n is much smaller than $1/n$. We computed the characteristics of the test function at (23), and arrived at a value of $\beta = 2.05 \times 10^{-4}$ for dimension $d = 5$. For $n = 60,000$, this gives an upper bound of -0.02 , while the corresponding number is less than -5 in the last graph of Fig. 3 (the difference is even larger if we take into account the factor in front of the exponential term).

Acknowledgements We thank the reviewers for valuable hints and remarks which facilitated the improvement of presentation of our results. The work of J. Calvin was supported by the National Science Foundation under Grant No. CMMI-0926949 and the work of G. Gimbutienė and A. Žilinskas was supported by the Research Council of Lithuania under Grant No. P-MIP-17-61.

Appendix A

This appendix contains the proof of Lemma 2.

Denote the eigenvalues of $D^2 f(x^*)$ by

$$\theta_1 \geq \theta_2 \geq \dots \geq \theta_d > 0.$$

Then by Taylor’s theorem (x a column vector),

$$f(x^* + x) - f(x^*) = \frac{1}{2}x^T D^2 f(x^*)x + o(\|x\|^2).$$

Since $D^2 f(x^*)$ is symmetric positive definite, we can express it as

$$D^2 f(x^*) = V \Theta V^T$$

for an orthogonal matrix V and diagonal matrix $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_d)$. Then

$$D^2 f(x^*) = V \Theta^{1/2} \Theta^{1/2} V^T$$

and

$$f(x^* + x) - f^* = \frac{1}{2} x^T V \Theta^{1/2} \Theta^{1/2} V^T x + o(\|x\|^2) = \frac{1}{2} \|Tx\|^2 + o(\|x\|^2), \tag{24}$$

where $Tx \equiv \Theta^{1/2} V^T x$.

Let $B_c^d(0)$ denote the ball of radius c , centered at 0, in \mathbb{R}^d . For $c > 0$, let

$$f_c \equiv \inf\{f(x^* + x) : x \notin B_c^d(0)\},$$

which is positive since the minimizer is assumed unique and f is continuous. Then

$$\int_{[0,1]^d \setminus B_c^d(x^*)} \frac{dx}{(f(x) - f(x^*) + \epsilon)^{d/2}} \leq f_c^{-d/2}$$

and

$$\begin{aligned} \int_{[0,1]^d} \frac{dx}{(f(x) - f(x^*) + \epsilon)^{d/2}} &\leq f_c^{-d/2} + \int_{B_c^d(0)} \frac{dx}{(f(x^* + x) - f(x^*) + \epsilon)^{d/2}} \\ &= f_c^{-d/2} + \int_{B_c^d(0)} \frac{dx}{\left(\frac{1}{2} \|Tx\|^2 + \epsilon + o(\|x\|^2)\right)^{d/2}} \end{aligned}$$

by (24). For any $\eta \in]0, 1/2]$, we can choose $c > 0$ small enough so that

$$\frac{1}{2} \|Tx\|^2 + \epsilon + o(\|x\|^2) \geq \left(\frac{1}{2} - \eta\right) \|Tx\|^2 + \epsilon.$$

Therefore,

$$\int_{[0,1]^d} \frac{dx}{(f(x) - f(x^*) + \epsilon)^{d/2}} \leq \int_{B_c^d(0)} \frac{dx}{\left((1/2 - \eta) \|Tx\|^2 + \epsilon\right)^{d/2}} + O(1) \tag{25}$$

as $\epsilon \downarrow 0$, and similarly

$$\int_{[0,1]^d} \frac{dx}{(f(x) - f(x^*) + \epsilon)^{d/2}} \geq \int_{B_c^d(0)} \frac{dx}{\left(\left(\frac{1}{2} + \eta\right) \|Tx\|^2 + \epsilon\right)^{d/2}} + O(1). \tag{26}$$

Using the orthogonality of V , for any $b > 0$,

$$\begin{aligned} \int_{B_c^d(0)} \frac{dx}{(b \|Tx\|^2 + \epsilon)^{d/2}} &= \int_{V(B_c^d(0))} \frac{dx}{\left(b \|\Theta^{1/2} x\|^2 + \epsilon\right)^{d/2}} \\ &= \int_{B_c^d(0)} \frac{dx}{\left(b \|\Theta^{1/2} x\|^2 + \epsilon\right)^{d/2}} \end{aligned}$$

$$= \int_{E(c,b,\epsilon)} \frac{dy}{(\|y\|^2 + 1)^{d/2}} \frac{1}{b^{d/2} \left(\prod_{i=1}^d \theta_i\right)^{1/2}}, \tag{27}$$

where $E(c, b, \epsilon)$ is the image of the ball of radius c under the map $x_i \mapsto x_i(\theta_i b/\epsilon)^{1/2}$, and we used the substitution $y_i \leftarrow x_i(\theta_i b/\epsilon)^{1/2}$ in the last equation. Therefore,

$$B_{c\sqrt{b\theta_d/\epsilon}}^d(0) \subset E(c, b, \epsilon) \subset B_{c\sqrt{b\theta_1/\epsilon}}^d(0),$$

and (27) gives the bounds

$$\begin{aligned} \int_{B_{c\sqrt{b\theta_d/\epsilon}}^d(0)} \frac{dy}{(\|y\|^2 + 1)^{d/2}} \frac{1}{b^{d/2} \left(\prod_{i=1}^d \theta_i\right)^{1/2}} &\leq \int_{B_c^d(0)} \frac{dx}{(b\|Tx\|^2 + \epsilon)^{d/2}} \\ &\leq \int_{B_{c\sqrt{b\theta_1/\epsilon}}^d(0)} \frac{dy}{(\|y\|^2 + 1)^{d/2}} \frac{1}{b^{d/2} \left(\prod_{i=1}^d \theta_i\right)^{1/2}}. \end{aligned}$$

Let $\mathcal{V}(x) = \pi^{d/2} \Gamma(d/2 + 1)^{-1} x^d$ denote the volume of the d -dimensional ball of radius x . For $z > 0$,

$$\begin{aligned} \int_{B_z^d(0)} \frac{dx}{(\|x\|^2 + 1)^{d/2}} &= \int_{r=0}^z \frac{d\mathcal{V}(r)}{(r^2 + 1)^{d/2}} = \frac{d\pi^{d/2}}{\Gamma(d/2 + 1)} \int_{r=0}^z \frac{r^{d-1} dr}{(r^2 + 1)^{d/2}} \\ &= \frac{d\pi^{d/2}}{\Gamma(d/2 + 1)} \int_{r=0}^z \frac{1}{r} \left(\frac{r^2}{r^2 + 1}\right)^{d/2} dr. \end{aligned}$$

For $z > 1$,

$$\int_{r=0}^z \frac{1}{r} \left(\frac{r^2}{r^2 + 1}\right)^{d/2} dr < \int_{r=1}^z \frac{1}{r} dr = \log(z), \tag{28}$$

and for $A < z$,

$$\begin{aligned} \int_{r=0}^z \frac{1}{r} \left(\frac{r^2}{r^2 + 1}\right)^{d/2} dr &> \left(\frac{A^2}{A^2 + 1}\right)^{d/2} \int_{r=A}^z \frac{1}{r} dr \\ &= \left(\frac{A^2}{A^2 + 1}\right)^{d/2} (\log(z) - \log(A)). \end{aligned}$$

Combining this inequality with (28) we conclude that

$$\lim_{z \uparrow \infty} \frac{\int_{r=0}^z \frac{1}{r} \left(\frac{r^2}{r^2 + 1}\right)^{d/2} dr}{\log(z)} = 1. \tag{29}$$

Set

$$C_d = \frac{d\pi^{d/2}}{\Gamma(1 + d/2)}$$

and

$$G_d(c) = \int_{r=0}^c \frac{r^{d-1} dr}{(r^2 + 1)^{d/2}}.$$

By (29), $G_d(c)/\log(c) \rightarrow 1$ as $c \uparrow \infty$. For any $b > 0$, we have the bounds

$$\begin{aligned} b^{-d/2} \left(\prod_{i=1}^d \theta_i \right)^{-1/2} C_d G_d \left(c\sqrt{b\theta_d/\epsilon} \right) &\leq \int_{B_d^c(0)} \frac{dx}{(b \|Tx\|^2 + \epsilon)^{d/2}} \\ &\leq b^{-d/2} \left(\prod_{i=1}^d \theta_i \right)^{-1/2} C_d G_d \left(c\sqrt{b\theta_1/\epsilon} \right), \end{aligned}$$

and

$$\begin{aligned} O(1) + (1/2 + \eta)^{-d/2} \left(\prod_{i=1}^d \theta_i \right)^{-1/2} C_d G_d \left(c\sqrt{(1/2 + \eta)\theta_d/\epsilon} \right) \\ \leq \mathcal{I}_f(\epsilon) \leq O(1) + (1/2 - \eta)^{-d/2} \left(\prod_{i=1}^d \theta_i \right)^{-1/2} C_d G_d \left(c\sqrt{(1/2 - \eta)\theta_1/\epsilon} \right). \end{aligned}$$

Since $\eta > 0$ is arbitrary, the fact that $G_d(c)/\log(c) \rightarrow 1$ as $c \uparrow \infty$ completes the proof.

Appendix B

This appendix contains the proofs of the lemmas upon which the proof of Theorem 1 are based.

Proof of Lemma 3 Let $m \leq n$ be the last time before n that the smallest hyper-rectangle was about to be split, so $v_m = 2v_n$ and by (5),

$$\rho^m \leq \frac{1}{\lambda \log(1/v_m)} = \frac{1}{\lambda \log(1/2v_n)}.$$

Note that $m \geq n_0$ by definition of n_2 .

We will show by induction that

$$\rho^{m+k} \leq \frac{2}{\lambda \log(1/v_{m+k})} \tag{30}$$

for $k \in \{1, 2, \dots, n - m\}$, if the hyper-rectangle that the algorithm is subdividing at iteration $m + k$ has not been previously subdivided since time m . At other times the bound is twice as large.

Let us first consider $\{\rho_i^{m+1} : i \leq m + 1\}$. Suppose that i is the split hyper-rectangle at time m , and let j denote a non-split hyper-rectangle, so $\rho_j^m \leq \rho_i^m$. Then there exists a point $c_j \in R_j$ such that

$$\rho_j^{m+1} = \frac{|R_j|}{(L_m(c_j) - M_{m+1} + g(v_{m+1}))^{d/2}}$$

$$\begin{aligned} &\leq \left(\frac{g(2v_n)}{g(v_n)}\right)^{d/2} \frac{|R_j|}{(L_m(c_j) - M_m + g(2v_n))^{d/2}} \text{ since } g(2v_n) > g(v_n) \\ &= \left(\frac{g(2v_n)}{g(v_n)}\right)^{d/2} \rho_j^m \leq \left(2 \frac{\log(1/2v_n)}{\log(1/v_n)}\right) \frac{1}{\lambda \log(1/v_m)} \text{ by (5)} \\ &= 2 \frac{\log(1/2v_n)}{\log(1/v_n)} \frac{1}{\lambda \log(1/2v_n)} = \frac{2}{\lambda \log(1/v_n)} = \frac{2}{\lambda \log(1/v_{m+1})}. \end{aligned}$$

Next consider a child of the split subrectangle i . Since we are splitting the smallest subrectangle, $v_m = 2v_n$, and there exists a point $c_i \in R_i$ such that (supposing that one child has index i at time $m + 1$)

$$\rho_i^{m+1} = \frac{|R_i|/2}{(L_{m+1}(c_i) - M_{m+1} + g(v_{m+1}))^{d/2}} \leq \frac{v_{m+1}}{g(v_{m+1})^{d/2}} = \frac{1}{\lambda \log(1/v_{m+1})}.$$

We have established the base case for the induction. Now consider iteration $m + k + 1$, $1 \leq k < n - m$. For the induction hypothesis, assume that

$$\rho_i^{m+k} \leq \frac{2}{\lambda \log(1/v_n)}$$

if hyper-rectangle i has not been split since time m and

$$\rho_i^{m+k} \leq \frac{4}{\lambda \log(1/v_n)}$$

if hyper-rectangle i has been split since time m .

Suppose that the most promising hyper-rectangle at iteration $m + k + 1$ is subrectangle $R_j = \prod_{i=1}^d [a_i, b_i]$,

$$\rho_j^{m+k+1} = \rho_j^{m+k+1} = \int_R \frac{ds}{(L_{m+k+1}(s) - M_{m+k+1} + g(v_n))^{d/2}}.$$

Assume that this hyper-rectangle has not been split since time m . Let us suppose that during this interval (between m and the next time that the smallest hyper-rectangle is split) R is split r times, and denote by L' the piecewise multilinear function defined over R after the splitting evaluations. Then

$$\max_{s \in R} |L'(s) - L_{m+k+1}(s)| \leq \max_{s \in R} |f(s) - L_{m+k+1}(s)| \leq \gamma(s),$$

where

$$\gamma(s) = \frac{1}{2} \sum_{i=1}^d (s - a_i)(b_i - s) \|D^2 f\|_{\infty, R} \leq \frac{1}{8} qd |R|^{2/d} \|D^2 f\|_{\infty, R} \equiv \hat{\gamma}.$$

Set

$$\bar{\gamma}(s) = \min\{\gamma(s), L_{m+k+1}(s) - M_{m+k+1}\}.$$

Then, denoting by ρ_1 the sum of the ρ values for the resulting subrectangles of R at time $m + k + q$, we have

$$\rho_1 = \int_R \frac{ds}{(L'(s) - M_{m+k+q} + g(v_n))^{d/2}}$$

$$\begin{aligned} &\leq \int_R \frac{ds}{(L_{m+k+1}(s) - \bar{\gamma}(s) - M_{m+k+1} + g(v_n))^{d/2}} \\ &\leq \int_R \frac{ds}{(L_{m+k+1}(s) - M_{m+k+1} + g(v_n) - \hat{\gamma})^{d/2}}. \end{aligned}$$

Let

$$a + Q(s) = L_{m+k+1}(s) - M_{m+k+1} + g(v_n),$$

where $a = \min_{s \in R} L_{m+k+1}(s) - M_n + g(v_n) > 0$. Then

$$\frac{\rho_1}{\rho_0} \leq \frac{\int_R \frac{ds}{(a+Q(s)-\hat{\gamma})^{d/2}}}{\int_R \frac{ds}{(a+Q(s))^{d/2}}}.$$

The latter ratio is maximized by $Q \equiv 0$, which corresponds to $\rho_0 = |R|/a^{d/2}$, and so

$$\begin{aligned} \frac{\rho_1}{\rho_0} &\leq \frac{\int_R \frac{ds}{(a-\hat{\gamma})^{d/2}}}{\int_R \frac{ds}{(a)^{d/2}}} = \frac{1}{\left(1 - \frac{\hat{\gamma}}{a}\right)^{d/2}} = \frac{1}{\left(1 - \frac{\frac{1}{8}qd|R|^{2/d} \|D^2 f\|_{\infty,R}}{a}\right)^{d/2}} \\ &\leq \frac{1}{\left(1 - \frac{1}{8}qd\rho_0^{2/d} \|D^2 f\|_{\infty,R}\right)^{d/2}} \leq \frac{1}{\left(1 - \frac{1}{8}qd \frac{4 \cdot 4^{2/d}}{qd^3 \|D^2 f\|_{\infty,R}} \|D^2 f\|_{\infty,R}\right)^{d/2}} \\ &= \frac{1}{\left(1 - \frac{4^{2/d}}{2d^2}\right)^{d/2}} \leq 2. \end{aligned}$$

We used the inequalities

$$\begin{aligned} \rho_0^{2/d} &\leq \left(\frac{4}{\lambda \log(n)}\right)^{2/d} \leq \left(\frac{4}{\lambda (d \|D^2 f\|_{\infty,R})^{d/2}}\right)^{2/d} \\ &= \frac{\left(\frac{4}{\lambda}\right)^{2/d}}{d \|D^2 f\|_{\infty,R}} = \frac{4 \cdot 4^{2/d}}{qd^3 \|D^2 f\|_{\infty,R}}, \end{aligned}$$

which follows from the induction hypothesis.

We have shown that whenever the algorithm is about to subdivide a hyper-rectangle for the first time since m , $\rho^{m+k} \leq 2/\lambda \log(m+k)$. After a hyper-rectangle is subdivided, subsequent subdivisions can never result in ρ more than twice as large.

The proof by induction of (30) is complete. □

Proof of Lemma 4 Suppose (to get a contradiction) that $|R_i| = v_n^*$, $|R_j| = v_n$, and $v_n^* = 4v_n$, but yet we are about to split hyper-rectangle R_j , resulting in $v_n^* > 4v_n$; that is, $\rho_j^n \geq \rho_i^n$:

$$\rho_j^n = \frac{|R_j| = v_n}{(L_n(c_j) - M_n + g(v_n))^{d/2}} \geq \frac{|R_i| = v_n^* = 4v_n}{(L_n(c_i) - M_n + g(v_n))^{d/2}} = \rho_i^n$$

for some $c_i \in R_i, c_j \in R_j$. This implies that

$$\frac{L_n(c_i) - M_n + g(v_n)}{L_n(c_j) - M_n + g(v_n)} \geq 4^{2/d}. \tag{31}$$

But $L_n(c_j) \geq M_n$ and

$$L_n(c_i) \leq M_n + \frac{1}{8}qd \|D^2 f\|_{\infty, R_i} |R_i|^{2/d}.$$

Thus

$$L_n(c_i) - L_n(c_j) \leq \frac{1}{8}qd \|D^2 f\|_{\infty, [0,1]^d} v_n^{2/d} 4^{2/d}.$$

This means that

$$\begin{aligned} \frac{L_n(c_i) - M_n + g(v_n)}{L_n(c_j) - M_n + g(v_n)} &= 1 + \frac{L_n(c_i) - L_n(c_j)}{L_n(c_j) - M_n + g(v_n)} \\ &= 1 + \frac{L_n(c_i) - L_n(c_j)}{v_n^{2/d}} \frac{v_n^{2/d}}{L_n(c_j) - M_n + g(v_n)} \\ &= 1 + \frac{L_n(c_i) - L_n(c_j)}{v_n^{2/d}} (\rho_j^n)^{2/d} \\ &\leq 1 + \frac{1}{8}qd \|D^2 f\|_{\infty, [0,1]^d} (4^{2/d}) \left(\frac{4}{\lambda \log(n)}\right)^{2/d} \\ &\leq 1 + \frac{1}{8}qd \|D^2 f\|_{\infty, [0,1]^d} (4^{2/d}) \frac{4^{2/d}}{(qd^2/4) \log(n)^{2/d}} \\ &\leq 1 + \frac{1}{2}qd \|D^2 f\|_{\infty, [0,1]^d} (4^{2/d}) \frac{4^{2/d}}{(qd^2)d \|D^2 f\|_{\infty, [0,1]^d}} \text{ since } n \geq n_2 \\ &= 1 + \frac{1}{2d^2} 4^{4/d} < 4^{2/d} \end{aligned}$$

since $d \geq 2$. But this contradicts (31), and establishes (9).

The proof of (10) follows from

$$\begin{aligned} \frac{\Delta_n}{g(v_n)} &\leq \frac{\frac{1}{8}qd \|D^2 f\|_{\infty, R_i} (4v_n)^{2/d}}{\frac{1}{4}qd^2 (v_n \log(1/v_n))^{2/d}} = \frac{\|D^2 f\|_{\infty, R_i} 4^{2/d}}{2d (\log(1/v_n))^{2/d}} \leq \frac{\|D^2 f\|_{\infty, R_i} 4^{2/d}}{2d (\log(n))^{2/d}} \\ &\leq \frac{\|D^2 f\|_{\infty, R_i} 4^{2/d}}{2d \cdot d \|D^2 f\|_{\infty, [0,1]^d}} \leq \frac{4^{2/d}}{2d^2} \leq \frac{2}{d^2}. \end{aligned}$$

□

Proof of Lemma 5 Recall that R_n^* denotes the hyper-rectangle containing x^* at time n , and that $|R_n^*| = v_n^* \leq 4v_n$ by Lemma 4. If h is the minimal edge length of R_n^* , then $v_n^* = 2^j h^d$ for some $0 \leq j < d$, and the diameter of R_n^* is less than $2h\sqrt{d}$. Also $h = 2^{-j/d} (v_n^*)^{1/d}$. Therefore, for $s \in R_n^*$,

$$\begin{aligned} L_n(s) - M_n &\leq L_n(s) - f(x^*) \\ &\leq \max_{s \in R_n^*} f(s) - f(x^*) \\ &\leq \frac{1}{2} (2h\sqrt{d})^2 \|D^2 f\|_{\infty, R_n^*} \text{ by Taylor's theorem} \\ &= \frac{1}{2} (2 \cdot 2^{-j/d} (v_n^*)^{1/d} \sqrt{d})^2 \|D^2 f\|_{\infty, R_n^*} \\ &\leq 2d (v_n^*)^{2/d} \|D^2 f\|_{\infty, [0,1]^d} \end{aligned}$$

$$\begin{aligned} &\leq 2d(4v_n)^{2/d} \left(\frac{\log(n)}{d}\right)^{2/d} \quad \text{by Lemma 4 and } n \geq n_2 \\ &\leq 2d(4v_n)^{2/d} \left(\frac{\log(1/v_n)}{d}\right)^{2/d} = 2 \cdot 4^{2/d} \frac{d}{d^{2/d}} \lambda^{-2/d} g(v_n) \\ &= 2 \cdot 4^{2/d} \frac{d}{d^{2/d}} \frac{4}{qd^2} g(v_n) < \frac{32}{dq} g(v_n). \end{aligned}$$

Therefore, by the previous inequality

$$\begin{aligned} \rho^n &\geq \int_{R_n^*} \frac{ds}{(L_n(s) - M_n + g(v_n))^{d/2}} \geq \frac{v_n^*}{\left(\frac{32}{dq} g(v_n) + g(v_n)\right)^{d/2}} \\ &\geq \frac{v_n}{\lambda v_n \log(1/v_n) \left(1 + \frac{32}{dq}\right)^{d/2}} = \frac{1}{\lambda \log(1/v_n) \left(1 + \frac{32}{dq}\right)^{d/2}} \\ &\geq \frac{1}{\lambda \log(1/v_n) \exp\left(\frac{16}{q}\right)} = \frac{\exp(-16/q)}{\lambda \log(1/v_n)}. \end{aligned}$$

□

Proof of Lemma 6 Equation (11) follows from Lemma 3.

For $n \geq n_2(f)$, the ρ values for the children of a split hyper-rectangle will not be much smaller than the parent. The largest possible decrease occurs if the values on the boundary of R_i are constant, say with value $M_{m+k} + A$, and the new function values are the largest possible, namely $M_n + A + a$, where

$$a \leq \frac{1}{8} qd \|D^2 f\|_{\infty, R_i} |R_i|^{2/d}.$$

Considering a child of the split hyper-rectangle (say with index i),

$$\begin{aligned} \rho_i^{n+1} &\geq \frac{|R_i|/2}{(A + a + g(v_n))^{d/2}} \geq \frac{|R_i|/2}{(A + g(v_n))^{d/2} \left(1 + \frac{a}{A+g(v_n)}\right)^{d/2}} \\ &\geq \frac{|R_i|/2}{(A + g(v_n))^{d/2} \left(1 + \frac{\frac{1}{8} qd \|D^2 f\|_{\infty, R_i} |R_i|^{2/d}}{|R_i|^{2/d}} \frac{|R_i|^{2/d}}{A+g(v_n)}\right)^{d/2}} \\ &\geq \frac{1}{2} \rho_i^n \frac{1}{\left(1 + \frac{\frac{1}{8} qd \|D^2 f\|_{\infty, R_i} |R_i|^{2/d}}{|R_i|^{2/d}} \frac{|R_i|^{2/d}}{A+g(v_n)}\right)^{d/2}} \\ &= \frac{1}{2} \rho_i^n \frac{1}{\left(1 + \frac{\frac{1}{8} qd \|D^2 f\|_{\infty, R_i} |R_i|^{2/d}}{|R_i|^{2/d}} (\rho_i^n)^{2/d}\right)^{d/2}} \\ &\geq \frac{1}{2} \rho_i^n \frac{1}{\left(1 + \frac{\frac{1}{8} qd \|D^2 f\|_{\infty, R_i} |R_i|^{2/d}}{|R_i|^{2/d}} \left(\frac{4}{\lambda \log(n)}\right)^{2/d}\right)^{d/2}} \quad \text{(by Lemma 3)} \end{aligned}$$

$$\begin{aligned} &\geq \frac{1}{2} \rho_i^n \frac{1}{\left(1 + \frac{\frac{1}{8}qd \|D^2 f\|_{\infty, R_i} |R_i|^{2/d}}{|R_i|^{2/d}} \left(\frac{4}{\lambda d^{d/2} \|D^2 f\|_{\infty, [0,1]^d}^{d/2}}\right)^{2/d}\right)^{d/2}} \quad (\text{since } n \geq n_2) \\ &\geq \frac{1}{2} \rho_i^n \frac{1}{\left(1 + \frac{1}{8}qd \left(\frac{4}{(qd^3/4)^{d/2}}\right)^{2/d}\right)^{d/2}} = \frac{1}{2} \rho_i^n \frac{1}{\left(1 + \frac{1}{8}qd \frac{4^{2/d}}{qd^3/4}\right)^{d/2}} \\ &= \frac{1}{2} \rho_i^n \frac{1}{\left(1 + \frac{1}{2} \frac{4^{2/d}}{d^2}\right)^{d/2}} \geq \frac{1}{2} \rho_i^n \exp\left(-\frac{1}{4d} 4^{2/d}\right) \geq \frac{1}{2} e^{-1/2} \rho_i^n, \end{aligned}$$

using the inequality $(1 + x/k)^k \leq \exp(x)$ and the fact that $d \geq 2$ implies that $4^{2/d}/(4d) \leq 1/2$.

Consider the average of the $\{\rho_i^n\}$ at time n that were the product of splits after time $n/2$. Over this time interval we have

$$\rho^k \geq \frac{\exp(-16/q)}{\lambda \log(1/v_k)} \geq \frac{\exp(-16/q)}{\lambda \log(1/v_n)}$$

by Lemma 5, since $n \geq 2n_2(f)$. Since the children ρ values will not be much smaller than the parent's,

$$\frac{1}{n} \sum_{i=1}^n \rho_i^n \geq \frac{1}{n} \frac{n/2}{n/2} \sum_{i=n/2}^n \rho^i \geq \frac{1}{2} \frac{1}{n/2} \sum_{i=n/2}^n \frac{\exp(-16/q)}{\lambda \log(1/v_i)} \geq \frac{1}{2} \frac{\exp(-16/q)}{\lambda \log(1/v_n)},$$

since $v_i \geq v_n$. □

Proof of Lemma 7 Fix a particular hyper-rectangle R_i with

$$\rho_i^n = \int_{R_i} \frac{ds}{(L_n(s) - f^* + g(v_n))^{d/2}}.$$

We first show that the integrals

$$\int_{R_i} \frac{ds}{(f(s) - f^* + g(v_n))^{d/2}}, \int_{R_i} \frac{ds}{(L_n(s) - f^* + g(v_n))^{d/2}}$$

are close. As in the proof of Lemma 3, let

$$a + Q(s) = L(s) - f^* + g(v_n),$$

where $a = \min_{s \in R_i} L_n(s) - f^* + g(v_n) > 0$. Then, from (6),

$$\max_{x \in R_i} |f(x) - L_n(x)| \leq \frac{1}{8} q \cdot d \|D^2 f\|_{\infty, R_i} |R_i|^{2/d} \equiv b,$$

and

$$\frac{\int_{R_i} \frac{ds}{(f(s) - f^* + g(v_n))^{d/2}}}{\int_{R_i} \frac{ds}{(L_n(s) - f^* + g(v_n))^{d/2}}} \geq \frac{\int_{R_i} \frac{ds}{(L_n(s) - f^* + g(v_n) + b)^{d/2}}}{\int_{R_i} \frac{ds}{(L_n(s) - f^* + g(v_n))^{d/2}}} = \frac{\int_{R_i} \frac{ds}{(a + Q(s) + b)^{d/2}}}{\int_{R_i} \frac{ds}{(a + Q(s))^{d/2}}}.$$

The smallest value of the ratio occurs when $Q(s) \equiv 0$.

The case $Q \equiv 0$ corresponds to $\rho = |R_i|/a^{d/2}$, and so

$$\begin{aligned} \frac{\int_{R_i} \frac{ds}{(a+Q(s)+b)^{d/2}}}{\int_{R_i} \frac{ds}{(a+Q(s))^{d/2}}} &\geq \frac{\int_{R_i} \frac{ds}{(a+b)^{d/2}}}{\int_{R_i} \frac{ds}{a^{d/2}}} = \left(\frac{a}{a+b}\right)^{d/2} \\ &= \frac{1}{\left(1 + \frac{1}{8}q \cdot d \|D^2 f\|_{\infty, R_i} \frac{|R_i|^{2/d}}{a}\right)^{d/2}} = \frac{1}{\left(1 + \frac{1}{8}q \cdot d \|D^2 f\|_{\infty, R_i} \rho^{2/d}\right)^{d/2}}. \end{aligned}$$

Now

$$\begin{aligned} \frac{1}{8}q \cdot d \|D^2 f\|_{\infty, R_i} \rho^{2/d} &\leq \frac{1}{8}q \cdot d \|D^2 f\|_{\infty, R_i} \left(\frac{4}{\lambda \log(n)}\right)^{2/d} \\ &\leq \frac{1}{8}q \cdot d \|D^2 f\|_{\infty, R_i} \frac{4^{2/d}}{\frac{qd^3}{4} \|D^2 f\|_{\infty, R_i}} = \frac{4^{2/d}}{2d^2}. \end{aligned}$$

Therefore,

$$\frac{\int_{R_i} \frac{ds}{(a+Q(s)+b)^{d/2}}}{\int_{R_i} \frac{ds}{(a+Q(s))^{d/2}}} \geq \frac{1}{\left(1 + \frac{4^{2/d}}{2d^2}\right)^{d/2}} \geq \frac{2}{3},$$

since the last expression is increasing in $d \geq 2$. Turning to an upper bound for

$$\frac{\int_{R_i} \frac{ds}{(f(s)-f^*+g(v_n))^{d/2}}}{\int_{R_i} \frac{ds}{(L_n(s)-f^*+g(v_n))^{d/2}}},$$

observe that

$$\begin{aligned} \frac{\int_{R_i} \frac{ds}{(f(s)-f^*+g(v_n))^{d/2}}}{\int_{R_i} \frac{ds}{(L_n(s)-f^*+g(v_n))^{d/2}}} &\leq \frac{\int_{R_i} \frac{ds}{(L_n(s)-f^*+g(v_n)-b)^{d/2}}}{\int_{R_i} \frac{ds}{(L_n(s)-f^*+g(v_n))^{d/2}}} \equiv \frac{\int_{R_i} \frac{ds}{(a+Q(s)-b)^{d/2}}}{\int_{R_i} \frac{ds}{(a+Q(s))^{d/2}}} \\ &\leq \frac{1}{\left(1 - \frac{b}{a}\right)^{d/2}} \leq \frac{1}{\left(1 - \frac{4^{2/d}}{2d^2}\right)^{d/2}} \leq 2. \end{aligned}$$

We have shown that

$$\begin{aligned} \frac{2}{3} \int_{[0,1]^d} \frac{ds}{(L_n(s) - f^* + g(v_n))^{d/2}} &\leq \int_{[0,1]^d} \frac{ds}{(f(s) - f^* + g(v_n))^{d/2}} \\ &\leq 2 \int_{[0,1]^d} \frac{ds}{(L_n(s) - f^* + g(v_n))^{d/2}}. \end{aligned} \tag{32}$$

Recall that v_n^* denotes the volume of the hyper-rectangle containing the minimizer x^* . Since $n \geq n_2$, by Lemma 4 $v_n^* \leq 4v_n$ and

$$M_n - f^* \leq \frac{g(v_n)}{d^2/2}.$$

Therefore,

$$\int_R \frac{ds}{(L_n(s) - M_n + g(v_n))^{d/2}} = \int_R \frac{ds}{(L_n(s) - M_n + g(v_n) + M_n - f^*)^{d/2}}$$

$$\begin{aligned}
&= \int_R \frac{ds}{(L_n(s) - M_n + g(v_n))^{d/2} \left(1 + \frac{\Delta_n}{L_n(s) - M_n + g(v_n)}\right)^{d/2}} \\
&\geq \int_R \frac{ds}{(L_n(s) - M_n + g(v_n)(1 + 2/d^2))^{d/2}} \text{ by (10)} \\
&= \frac{1}{(1 + 2/d^2)^{d/2}} \int_R \frac{ds}{(L_n(s) - M_n + g(v_n))^{d/2}} \\
&\geq \frac{2}{3} \int_R \frac{ds}{(L_n(s) - M_n + g(v_n))^{d/2}}.
\end{aligned}$$

Combining these inequalities with (32) gives

$$\begin{aligned}
\frac{4}{9} \int_{[0,1]^d} \frac{ds}{(L_n(s) - M_n + g(v_n))^{d/2}} &\leq \frac{2}{3} \int_{[0,1]^d} \frac{ds}{(L_n(s) - f^* + g(v_n))^{d/2}} \\
&\leq \int_{[0,1]^d} \frac{ds}{(f(s) - f^* + g(v_n))^{d/2}} \leq 2 \int_{[0,1]^d} \frac{ds}{(L_n(s) - f^* + g(v_n))^{d/2}} \\
&\leq 2 \int_{[0,1]^d} \frac{ds}{(L_n(s) - M_n + g(v_n))^{d/2}}.
\end{aligned}$$

Since

$$\sum_{i=1}^n \rho_i^n = \int_{[0,1]^d} \frac{ds}{(L_n(s) - M_n + g(v_n))^{d/2}},$$

this completes the proof. \square

References

1. Floudas, Ch.: *Deterministic Global Optimization: Theory, Algorithms and Applications*. Kluwer, Dordrecht (2000)
2. Horst, R., Pardalos, A.P., Thoai, N.: *Introduction to Global Optimization*. Kluwer, Dordrecht (1995)
3. Horst, R., Tuy, H.: *Global Optimization: Deterministic Approaches*. Springer, Berlin (1996)
4. Pinter, J.: *Global Optimization in Action*. Kluwer, Dordrecht (1996)
5. Sergeyev, Y.D.: Multidimensional global optimization using the first derivatives. *Comput. Maths. Math. Phys.* **39**(5), 743–752 (1999)
6. Sergeyev, Y.D., Kvasov, D.E.: A deterministic global optimization using smooth diagonal auxiliary functions. *Commun. Nonlinear Sci. Numer. Simul.* **21**, 99–111 (2015)
7. Sergeyev, Y.D., Kvasov, D.E.: *Deterministic Global Optimization, An Introduction to the Diagonal Approach*. Springer, New York (2017)
8. Mockus, J.: *Bayesian Approach to Global Optimization*. Kluwer, Dordrecht (1988)
9. Strongin, R.G., Sergeyev, Y.D.: *Global Optimization with Non-convex Constraints: Sequential and Parallel Algorithms*. Kluwer, Dordrecht (2000)
10. Zhigljavsky, A.: *Theory of Global Random Search*. Kluwer, Dordrecht (1991)
11. Žilinskas, A.: A statistical model-based algorithm for black-box multi-objective optimisation. *Int. J. Syst. Sci.* **45**(1), 82–92 (2014)
12. Žilinskas, A., Zhigljavsky, A.: Stochastic global optimization: a review on the occasion of 25 years of *Informatica*. *Informatica* **27**, 229–256 (2016)
13. Hooker, J.: Testing heuristics: we have it all wrong. *J. Heuristics* **11**, 33–42 (1995)
14. Pardalos, P., Romeijn, H.: *Handbook of Global Optimization*, vol. 2. Springer, Berlin (2002)

15. Rastrigin, L.: Statistical Models of Search. Nauka (2013) (in Russian)
16. Paulavičius, R., Sergejev, Y., Kvasov, D., Žilinskas, J.: Globally-biased DISIMPL algorithm for expensive global optimization. *J. Glob. Optim.* **59**, 545567 (2014)
17. Žilinskas, A., Žilinskas, J.: A hybrid global optimization algorithm for non-linear least squares regression. *J. Glob. Optim.* **56**, 265–277 (2013)
18. Calvin, J., Žilinskas, A.: A one-dimensional P-algorithm with convergence rate $o(n^{-3+\delta})$ for smooth functions. *J. Optim. Theory Appl.* **106**, 297–307 (2000)
19. Calvin, J.M.: An adaptive univariate global optimization algorithm and its convergence rate under the Wiener measure. *Informatica* **22**, 471–488 (2010)
20. Calvin, J.M., Chen, Y., Žilinskas, A.: An adaptive univariate global optimization algorithm and its convergence rate for twice continuously differentiable functions. *J. Optim. Theory Appl.* **155**, 628–636 (2011)
21. Calvin, J.M., Žilinskas, A.: On a global optimization of bivariate smooth functions. *J. Optim. Theory Appl.* **163**, 528–547 (2014)
22. Novak, E., Woźniakowski, H.: Tractability of Multivariate Problems, II. Tracts in Mathematics, vol. 12. European Mathematical Society, Zürich (2010)
23. Huyer, W., Neumaier, A.: Global optimization by multi-level coordinate search. *J. Glob. Optim.* **14**, 331–355 (1999)
24. Jones, D.R., Perttunen, C.D., Stuckman, C.D.: Lipschitz optimization without the lipschitz constant. *J. Optim. Theory Appl.* **79**, 157–181 (1993)
25. Sergejev, Y., Kvasov, D.: Global search based on efficient diagonal partitions and set of Lipschitz constants. *SIAM J. Optim.* **16**(3), 910–937 (2006)
26. Scholz, D.: Deterministic Global Optimization: Geometric Branch-and-Bound Methods and their Applications. Springer, Berlin (2012)
27. Paulavičius, R., Žilinskas, J.: Simplicial Global Optimization. Springer Briefs in Optimization. Springer, Berlin (2014)
28. Novak, E.: Deterministic and Stochastic Error Bounds in Numerical Analysis, Lecture Notes in Mathematics, vol. 1349. Springer, Berlin (1988)
29. Zhigljavsky, A., Žilinskas, A.: Stochastic Global Optimization. Springer, Berlin (2008)
30. Waldron, S.: Sharp error estimates for multivariate positive linear operators which reproduce the linear polynomials. In: Chui, C.K., Schumaker, L.L. (eds.) *Approximation Theory IX*, vol. 1, pp. 339–346. Vanderbilt University Press, Nashville (1998)
31. McGeoch, C.: Experimental analysis of algorithms. In: Pardalos, P., Romeijn, E. (eds.) *Handbook of Global Optimization*, vol. 2, pp. 489–514. Kluwer, Dodrecht (2002)
32. More, J., Wild, S.: Benchmarking derivative-free optimization algorithms. *SIAM J. Optim.* **20**, 172–191 (2009)
33. GSL, GNU Scientific Library. <https://www.gnu.org/software/gsl/>. Accessed on 1 Aug 2016
34. CGAL, Computational Geometry Algorithms Library. <http://www.cgal.org>. Accessed on 15 Oct 2014
35. Sobol, I.: On the systematic search in a hypercube. *SIAM J. Numer. Anal.* **16**, 790793 (1979)
36. Dixon, L.C.W., Szegő, G.P. (eds.): *Towards Global Optimization II*. North Holland, New York (1978)
37. Branin, F.: Widely convergent method for finding multiple solutions of simultaneous nonlinear equations. *IBM J. Res. Dev.* **16**(5), 504–522 (1972)
38. Hansen, P., Jaumard, B.: Lipschitz optimization. In: Horst, R., Pardalos, P. (eds.) *Handbook of Global Optimization*, pp. 407–493. Kluwer, Dodrecht (1995)
39. Liu, H., Xu, S., Ma, Y., Wang, X.: Global optimization of expensive black box functions using potential lipschitz constants and response surfaces. *J. Glob. Optim.* **63**, 229251 (2015)