On the convergence rate of the quasi – Monte Carlo method of search for extremum

A S Tikhomirov
Yaroslav-the-Wise Novgorod State University, ul. B. St. Petersburgskaya, 41
173003 Veliky Novgorod, Russia
E-mail: Alexey.Tikhomirov@novsu.ru

Abstract. The convergence rate of the quasi – Monte Carlo method of search for extremum is examined. It is shown that, if the objective function is nonsingular, then the number of its evaluations required to obtain the desired accuracy \( \varepsilon \) in the solution can be a slowly (namely, logarithmically) growing function as \( \varepsilon \) approaches zero.

1. Introduction

The work has been devoted to the theoretical examination of the convergence rate of a variant of the quasi – Monte Carlo method of search for function global extremum. Let the objective function \( f: (0, 1]^d \mapsto \mathbb{R} \) take the minimal value in the unit point \( x^* \). Quasi – Monte Carlo method is used to find the minimum point \( x^* \) with the given accuracy \( \varepsilon > 0 \) (“argument” approximation). As the characteristic of the algorithm convergence rate, we use the number of calculations of the target function required to achieve the given accuracy \( \varepsilon \) of the solution. The main consideration is given to the growth order of the characteristic of calculations of the target function which is required to achieve the given accuracy \( \varepsilon \) of the solution, when \( \varepsilon \) tends to zero.

No doubt, at practical calculations the major role may be played by constants in corresponding number exponents. However, it is well known that to find these constants in theory is much more difficult task than to define accurate number exponents, and that even to get good evaluations of exact number exponents in the tasks of global optimization is a complex problem. That is why we limit ourselves to studying mainly the convergence rate characteristics.

Before stating the results of our work, we present certain known evaluations of convergence rates of various methods of optimization of objective functions (of not very special kind). Thus, the certain scale of characteristics of convergence rates of optimization algorithms will be set, and it will be evident what place in this scale is taken by the obtained results.

Conditionally, the optimization theory may be divided into two parts: local optimization theory and theory of search for global extremum. For the wide class of local methods of optimization and regular target functions, the number of search steps to achieve accuracy \( \varepsilon \) looks like \( O(|\ln \varepsilon|) \). If the requirements to the functions optimized are more strict, for the wide range of local methods the number of steps has the characteristic \( O(\ln(\ln \varepsilon)) \) (q.v., e.g. [1, 2]).

For the methods of stochastic global optimization (q.v., e.g. [3–22]), the typical result is power (i.e. \( O(1/\varepsilon^a) \) in \( a > 0 \)) law of the required number of calculations of the objective function from \( \varepsilon \) that is much more worse than for the local methods. Thus, for the wide class of results there is a large gap between logarithmic exponents of “quick” local methods and power exponents of “slow” methods of stochastic global optimization.

Quick methods of stochastic global optimization are certain algorithms of Markovian monotonous random search. In the works [14–18], such homogeneous Markovian monotonous searches were constructed that have the required number of calculations of the objective function as \( O(\ln^2 \varepsilon) \). In the works [14, 15, 21], heterogeneous Markovian monotonous searches with the required number of calculations of the target function \( O(|\ln \varepsilon| \times \ln |\ln \varepsilon|) \) were considered.
In the work [20], it has been shown that, for any Markovian symmetric random searches, the inequality $N(\varepsilon,\psi) \geq \psi(\ln(p(x, x_0)/\varepsilon) + 2)$ is true, where $N(\varepsilon,\psi)$ is such number of search steps at which the achievement of the $\varepsilon$–neighborhood of minimum point is guaranteed with the probability $\psi$, and $x$ is the initial searching point. Thus, for the wide class of Markovian algorithms of random search, the number of calculations of objective function which is required to achieve the given accuracy $\varepsilon$ and reliability $\gamma$ can not be less than $\gamma(\ln(p(x, x_0)/\varepsilon) + 2)$.

The given article presents such variants of the quasi–Monte Carlo method in which (for non-degenerate objective functions) the required number of calculations of objective function looks like $O(|\ln \varepsilon|)$. The working results demonstrate that the quasi–Monte Carlo method can be quick (from the point of view of characteristic of dependence of labor intensity upon $\varepsilon$), provided it is well organized. The convergence rate of the examined method of search for global extremum of the function can be compared by the characteristic with the convergence rate of quick methods of local optimization.

The considered quasi Monte-Carlo method is presented in [23] (q.v.also [24]). At this, in [23] there are neither theorems about the convergence of the regarded method, nor evaluations of the convergence rate. The similar method is briefly described in [25]. The given work adds to the results [23–25].

2. Optimization Space

Let the set of optimization $X$ supplied with metric $\rho$ be called the optimization space. We limit ourselves to the case $X = \mathbb{I}^d = (0, 1]^d$ and the following variants of metrics $\rho(x, y)$:

$$
\rho_p(x, y) = \left( \sum_{i=1}^{d} \rho_i^p(x_i, y_i) \right)^{1/p}, \quad \rho_{\infty}(x, y) = \max_{1 \leq i \leq d} \rho_i(x_i, y_i),
$$

where $p \geq 1$ is any fixed number, $x = (x_1, ..., x_d)$ and $y = (y_1, ..., y_d)$, $\rho_i(x_i, y_i) = \min(|x_i - y_i|)$.

For any metric (1) the space $(\mathbb{I}^d, \rho)$ topologically is a $d$-dimensional torus. We’ll denote $\Gamma = \text{diam } X$.

Closed ball of radius $r$ with the centre in point $x$ is denoted as $B_r(x) = \{y \in X: \rho(x, y) \leq r\}$. Let $\varphi(r) = \mu(B_r(x))$, where $\mu$ means the $d$-dimensinal Lebesgue measure. Notice that $\varphi(r)$ does not depend upon $x$, at this for all sufficiently small $r$

$$
\varphi(r) = A r^d,
$$

where $A$ is a constant which depends upon the space dimensionality and metric $\rho$. Besides $\varphi(r) = 1$ at $r \geq \Gamma$.

The choice of the torus metric in the cube $\mathbb{I}^d$ instead the Euclidean metric allows to avoid studying “marginal effects” arising near the boundary of the cube. In the same time, for the exponents of the convergence rate of search for extremum the choice of torus as the optimization space is not principal. Indeed, at $x_0 = \text{arg min}\{f(x): x \in X\} \in (0, 1)^d$ small neighborhood of the point $x_0$ in the cube and the torus coincide, and the results about the convergence rate obtained for the torus are automatically transferred to the cube. The choice of the torus metric affects here only the construction of the search.

3. Objective Function

Further, we will assume that the objective function $f: X \mapsto \mathbb{R}$ is bounded below, measurable, and meets the following conditions.

**Condition 1.** The function $f$ takes its minimal value at a unique point $x_*= \text{arg min}\{f(x): x \in X\}$.

**Condition 2.** The function $f$ is continuous at the point $x_*$.

**Condition 3.** The inequality $\inf\{f(x): x \in B_r^e(x_*)\} > f(x_*)$ is satisfied for any $r > 0$.

Here, with $B_r^e(x_*)$ completing the ball $B_r(x_*)$ is denoted in the space $X$. Due to the condition 3, the convergence $f(x_n) \rightarrow f(x_*)$ implies the convergence $\rho(x_n, x_*) \rightarrow 0$. Notice that the functions of the mentioned class may be multi-extremal in any neighborhood of global minimum.

4. Search Objective

Voluntary sequence of points $\{\xi_t\}_{t \geq 1}$ in the optimization space $X$ will be called search. If the inequality $f(\xi_{t+1}) \leq f(\xi_t)$ is performed for any $t \geq 1$, then the search is called monotonous. Further we will study monotonous searchings of special kind.

The search is used to find the point of minimum $x_*$ with the given accuracy $\varepsilon$ (“argument” approximation). At argument approximation we have to be interested in the search getting into the ball $B_\varepsilon(x_*)$. However, it may occur that the search that will be in $B_\varepsilon(x_*)$ on the step $t$, leaves $B_\varepsilon(x_*)$ on one of the following steps. To avoid the analysis of such effects, we introduce sets
\[ M_r = \{ x \in B_r(x_0) : f(x) < f(y) \quad \forall y \in B_r^c(x_0) \}. \]

It is easy to see that sets \( M_r \) possess the following properties: a) if \( r_1 < r_2 \), then \( M_{r_1} \subset M_{r_2} \), b) if \( x \in M_r \) and \( y \notin M_r \), then \( f(x) < f(y) \). Once entering the set \( M_r \), the monotonous search leaves it no more. So, we will study the moment of the search entering the set \( M_r \), where \( \varepsilon \) maintains the value of the required search accuracy.

Below, all used information about the objective function \( f \) will be kept in the form of the function

\[ V^f(r) = \sup \{ v > 0 : B_r(x) \subset M_r \text{ for certain } x \in M_r \}. \quad (3) \]

It is easy to demonstrate that, for objective functions meeting the conditions 1–3, the inequality \( V^f(r) > 0 \) is performed at all \( r > 0 \). Functions which have \( \lim \inf V^f(r)/r > 0 \) at \( r \rightarrow 0 \) will be called \( V \)-nondegenerate (briefly, nondegenerate).

5. Quasi–Monte Carlo Method of Search for Extremum

To find the point of global maximum of the objective function \( f \), we will use the quasi–Monte Carlo method [23]. Any finite set of points \( Q = \{ x_1, ..., x_n \} \) from \( A \) is called a grid in the set \( A \subset X \). For the grid \( Q \) the number \( n \) of its points is denoted as \( |Q| \).

In correspondence with [5, 23, 24], as the characteristic of “density” of the grid \( Q = \{ x_1, ..., x_n \} \) in the set \( A \) we will use the dispersion of the grid \( Q \) from the set \( A \):

\[ \sigma(Q, A) = \sup_{y \in A} \min_{x \in Q} \rho(x, y). \]

It is clear that the less the value \( \sigma(Q, A) \) is, the better the grid \( Q \) “feels” the set \( A \).

Let us suppose that some method \( Q \) of constructing grids in balls \( B_r(x) \) at all possible values \( r > 0 \) and \( x \in X \) has been fixed, and write \( Q \subset Q' \), if the grid \( Q \) in \( B_r(x) \) is constructed with the help of this method. We will demand from the method \( Q \) for any ball \( B_r(x) \) and any number \( t \in (0, \Gamma] \) to have such grid \( Q \subset Q' \) in \( B_r(x) \), that \( \sigma(Q, B_r(x)) \leq t \).

Let us suppose that

\[ n(x, r, t) = \min \{ |Q| : Q \in Q, \sigma(Q, B_r(x)) \leq t \} \]

and denote as \( Q_{\min}(t, B_r(x)) \) the (best) grid, at which the minimum on the right part is achieved (4). Due to the conditions on \( Q \), values \( n(x, r, t) \) are finite. In addition, we will demand for these values not to depend upon \( x \), and denote them as \( n(r, t) \).

The grid in the ball \( B_r(x) \) can be made, using a simple transformation (q.v. [23, 24]) of the “initial” grid constructed in the cube \([0, 1]^d\). At this way of constructing grids in balls, values \( n(x, r, t) \) will not depend upon \( x \). For example, the simplest cubic grids or other nonrandom grids described in [5, 23, 24] meet all introduced requirements.

Imagine the algorithm of search \( x_*=\min\{f(x):x\in X\} \) for the examined method of grids. Parameters of the algorithm are the method \( Q \) of constructing grids, the number of grids \( m \), radii \( a_2, ..., a_m \) of balls in which the grids are constructed (the very first grid is constructed in the cube \([0, 1]^d\)), and upper boundaries \( \tau_1, ..., \tau_m \) of deviations of the grids.

Algorithm 1

Step 1. \( i \leftarrow 1 \). Construct the grid \( \{ x_1^{(1)}, ..., x_{n_1}^{(1)} \} = Q_{\min}(\tau_1, X) \subset Q \), where \( n_1 = n(\Gamma, \tau_1) \).

Step 2. \( \xi_1 \leftarrow \arg \min \{ f(x_1^{(1)}), ..., f(x_{n_1}^{(1)}) \} \).

Step 3. If \( i = m \), then complete the work of the algorithm, otherwise \( i \leftarrow i + 1 \).

Step 4. Construct the grid \( \{ x_1^{(i)}, ..., x_{n_i}^{(i)} \} = Q_{\min}(\tau_i, B_{\xi_i}(\xi_{i-1})) \subset Q \), where \( n_i = n(a_i, \tau_i) \).

Step 5. \( \xi_i \leftarrow \arg \min \{ f(x_1^{(i)}), ..., f(x_{n_i}^{(i)}, f(x_{n_i}^{(i)}) \} \) and pass to step 3.

Let us think that

\[ \arg \min[f(x_1), ..., f(x_n)] = x_0, \]

where

\[ i = \max\{ j \in \{ 1, ..., n \} : f(x_j) = \min\{ f(x_1), ..., f(x_n) \} \}. \]

The search algorithm regarded is similar to the search method regarded in [23]. At this in [23] there are neither regarded search convergence theorems nor convergence rate evaluations.
Let us set a priori information about the objective function $f$ in the form of such function $v(r)$ that $0 < v(r) < V'(r)$ at $0 < r < \Gamma$ (where the function $V'(r)$ is given by the formula (3)). Then, for the algorithm 1 search, the following statement is justified.

**Theorem 1.** Let $m \geq 1$ and $0 < n_m \leq \varepsilon < r_{m-1} < \cdots < r_1 < \Gamma$. Suppose $a_i = \min\{r_i + r_{i-1}, \Gamma\}$ at $2 \leq i \leq m$. Let $t_i = v(r_i)$ at $1 \leq i \leq m$. Then, to find algorithm 1 the ratios $\xi_i \in M_{r_i}$ are performed. In particular, $\xi_m \in M_{\varepsilon}$.

**Proof.** For unity of notations let us suppose $r_0 = a_1 = \Gamma$ and take the arbitrary point $\xi_0 \in X$. Since $M_{\varepsilon} = X$, then $\xi_0 \in M_{\varepsilon}$.

Now suppose that $\xi_{i-1} \in M_{r_{i-1}}$ and prove that $\xi_i \in M_{r_i}$.

After determining the function $v$, $V'$ and value $t_i$, we will find such $y$ that $B_{t_i}(y) \subseteq M_{r_i}$. Since $\xi_{i-1} \in B_{r_{i-1}}(x_i)$ and $a_i = \min\{r_i + r_{i-1}, \Gamma\}$, then $B_{a_i}(\xi_{i-1}) \supseteq B_{r_i}(x_i)$. Thus $B_{a_i}(\xi_{i-1}) \supseteq B_{r_i}(x_i) \supseteq M_{r_i} \supseteq B_{t_i}(y)$.

To be brief, let us denote $Q_i = Q_{\min}(t_i, B_{a_i}(\xi_{i-1}))$. Since $\sigma(Q_i, B_{a_i}(\xi_{i-1})) \leq t_i$, we will find such point $x_j^{(i)} \in Q_i$, that $\rho(x_j^{(i)}, y) \leq t_i$. Thus, $x_j^{(i)} \in B_{t_i}(y)$, and, so, $x_j^{(i)} \in M_{r_i}$. And since by constructing $f(\xi_i) \leq f(x_j^{(i)})$, then (due to the properties of the set $M_{r_i}$) $\xi_i \in M_{r_i}$. The theorem has been proved.

At performing algorithm 1, the values of the target function $f$ are calculated $N(\varepsilon) = n_1 + \cdots + n_m$ times. So, the value $N(\varepsilon)$ is the natural characteristic of labor intensity of algorithm 1. In the following statement the behavior $N(\varepsilon)$ is studied at small $\varepsilon$.

**Theorem 2.** Let the target function $f$ be $V$-nondegenerate, and for the function $v$ the inequality $\vartheta = \inf v(r)/r : 0 < r < \Gamma$ has been performed. After choosing $0 < q < 1$, let us suppose $R = \Gamma/(1 + q)$, $m = [\ln(\varepsilon/R)/\ln q]$ and $r_i = Rq^i$ at $1 \leq i \leq m$. Let us suppose that $n(r, t) \leq v(r/t)$, where nondecreasing function $v$ depends upon the selected method of constructing grids, the metric used, and the space dimensionality. Then, in the conditions of theorem 1

$$N(\varepsilon) \leq v((1 + 1/q)/\vartheta)[\ln(\varepsilon/R)/\ln q].$$

**Proof.** Let us denote $\omega = \ln(\varepsilon/R)/\ln q$ and note that $m = [\omega]$ and

$$Rq^\omega = \text{Re}x \left( \frac{\ln(\varepsilon/R)}{\ln q} \right) = \text{Re}x(\ln(\varepsilon/R)) = \frac{R\varepsilon}{R} = \varepsilon.$$

Further we will obtain

$$r_m = Rq^m = Rq^{[\omega]} = Rq^{\omega + [\omega]} = Rq^\omega q^{[\omega] - \omega} = \varepsilon q^{[\omega] - \omega} \quad \text{and} \quad \varepsilon q < q^{[\omega] - \omega} \leq \varepsilon.$$

Thus, $\varepsilon q < r_m \leq \varepsilon$.

Now let us evaluate $n_i = n(a_i, t_i)$. We have $v(r)/r \geq \vartheta$, and, so, $t_i = v(r_i) \geq \vartheta r_i$. For $i \geq 2$ $a_i = \min\{r_i + r_{i-1}, \Gamma\} \leq r_i + r_{i-1}$ has been performed, and

$$\frac{a_i}{t_i} \leq \frac{r_i + r_{i-1}}{\vartheta r_i} = \frac{1}{\vartheta} \left( 1 + \frac{r_{i-1}}{r_i} \right) = \frac{1}{\vartheta} \left( 1 + \frac{Rq^{i-1}}{Rq^i} \right) = \frac{1}{\vartheta} \left( 1 + \frac{1}{q} \right) = \frac{1 + 1/q}{\vartheta}. \tag{6}$$

For $i = 1$ $a_1 = \Gamma$ has been performed, and

$$\frac{a_1}{t_1} \leq \frac{\Gamma}{\vartheta r_1} = \frac{\Gamma}{\vartheta Rq} = \frac{\Gamma(1 + q)}{\vartheta q} = \frac{1 + q}{\vartheta} = \frac{1 + 1/q}{\vartheta}. \tag{7}$$

Using (6) and (7), we will obtain

$$N(\varepsilon) = \sum_{i=1}^{m} n_i = \sum_{i=1}^{m} n(a_i, t_i) \leq \sum_{i=1}^{m} v\left( \frac{a_i}{t_i} \right) \leq \sum_{i=1}^{m} v\left( \frac{1 + 1/q}{\vartheta} \right) = v\left( \frac{1 + 1/q}{\vartheta} \right) \frac{\ln(\varepsilon/R)}{\ln q}.$$

The theorem has been proved.
Comment 1.1. If in algorithm 1 the grids $Q_{\max}(t_i, B_{a_i})$ are replaced by any other grids $Q_i c|Q_i| \leq v(a_i/t_i)$ and $\sigma(Q_i, B_{a_i}) \leq t_i$, then the inequality (5) will still be left.

2. Since for the metric $\rho_{\infty}$ and cubic grids the inequality $n(r, t) \leq [r/t]^d$ is performed, then (5) is performed with $v(z) = |z|^d$.

3. For the metric $\rho_{\infty}(x, y) = \max_{1 \leq i \leq d}|x_i - y_i|$ in $[0, 1]^d$ and the wide range of grids the inequalities $\sigma(Q, [0, 1]^d) < \alpha|Q|^{-1/d}$ are justified, where $|Q|$ is the number of points of the grid $Q$, and the constant $\alpha$ depends upon the kind of the grid used (q.v. [24]). Using these grids, it is easy to construct such grids in the space $(\mathbb{N}, \rho_{\infty})$, that $n(r, t) \leq \left[(2\alpha r/t)^d\right]$. In this case the inequality (5) is performed with $v(z) = \left[(2\alpha z)^d\right]$.

4. Choosing $q$, it is possible to decrease the right part of the inequality (5). E.g., it is possible to take $q = \arg\min((1 + 1/p)^d/|\ln p|: 0 < p < 1)$.

5. Theorem 1 demonstrates that the value $N(\varepsilon)$ is associated with the search getting into $M_{\max} \subset M_{\varepsilon}$, and not directly into $M_{\varepsilon}$. At this (q.v. statement 1 below), the number $q$ can be corrected so that the equality $r_m = \varepsilon$ can be performed. Corrected value $\hat{q}$ is given with the equation $\hat{q}^m/(1 + \hat{q}) = \varepsilon/\Gamma$. At $\varepsilon < \Gamma/2$, this equation has the only solution on the space $[q, 1)$.

6. It is clear that the right part of the inequality (5) looks like $O(\ln \varepsilon)$ at $\varepsilon \rightarrow 0$. Thus, for $V$-nondegenerate target functions $f$ (if there is a priori information about the behavior of $f$ in the neighborhood of the extremum) there is quasi-Monte-Carlo method, with the number of target function calculations dependent logarithmically on the required accuracy of the task solution.

Statement 1. In the conditions of theorem 2 $q\varepsilon < r_m \leq \varepsilon$ is performed. The number $q$ may be corrected so that the equality $r_m = \varepsilon$ can be performed. The corrected value $\hat{q}$ is given with the equation $\hat{q}^m/(1 + \hat{q}) = \varepsilon/\Gamma$. At $\varepsilon < \Gamma/2$, this equation has the only solution on the space $[q, 1)$.

Proof. In the conditions of theorem 2

$$r_m = R q^m = \frac{\Gamma q^m}{1 + q} \leq \varepsilon.$$  

At fixed $m$, $\varepsilon$ and $\Gamma$, we will consider the equation

$$p^m/(1 + p) = \varepsilon/\Gamma.$$  

(8)

At $p \in [0, 1]$, we will set the function

$$h(p) = \frac{p^m}{1 + p}.$$  

Then

$$h'(p) = \frac{mp^{m-1}(1 + p) - p^m}{(1 + p)^2} = \frac{mp^{m-1} + mp^m - p^m}{(1 + p)^2} = \frac{mp^{m-1} + p^m(m - 1)}{(1 + p)^2},$$

and $h'(p) > 0$ at $p \in (0, 1)$. It means that the function $h$ increases by $[0, 1]$, and

$$\min_{p \in [0, 1]} h(p) = h(0) = 0, \quad \max_{p \in [0, 1]} h(p) = h(1) = 1/2.$$  

That is why at $\varepsilon/\Gamma < 1/2$ the equation (8) has the only solution $\hat{q} \in (0, 1)$.

Since, according to the condition, $r_m = \Gamma q^m/(1 + q) \leq \varepsilon$, then $h(q) = q^m/(1 + q) \leq \varepsilon/\Gamma$, and, so, $q \leq \hat{q}$. The statement has been proved.

To apply the search of theorem 2, you need a priori information about objective function $f$ in the form of such function $v(r)$ that $0 < v(r_i) < V_f(r_i)$ at $1 \leq i \leq m$. If there is no such information, it is possible to organize cyclic repetitions of algorithm 1 search, starting with the certain given value $v_1(r)$ at the first iteration, and then, for example, supposing $v_j(r) = a^j r_1(r)$ where $a < 1$ at $j$ iteration. It is clear that, after the finite number of iterations, the inequalities $v_j(r_i) < V_f(r_i)$ will become just. A variant of the cyclic search is presented in algorithm 2. The given variant of the search does not contain the conditions of stopping and generates the infinite sequence $\{\xi_1\}_{i=1}^\infty$. 

5
Algorithm 2
Step 1. $i \leftarrow 1, j \leftarrow 1$.
Step 2. Construct the grid $\{x^{(1)}_1, ..., x^{(1)}_{n_1}\} = Q_{\min}(t_1, X) \in Q$, where $t_1 = v_j(r_1), n_1 = n(\Gamma, t_1)$.
Step 3. If $j = 1$, then $\xi_i \leftarrow \arg\min \{f(x^{(1)}_1), ..., f(x^{(1)}_{n_1})\}$, otherwise
$$\xi_i \leftarrow \arg\min \{f(x^{(1)}_1), ..., f(x^{(1)}_{n_1}), f(\eta_{j-1})\}.$$  
Step 4. If $i = m$, then pass to step 7, otherwise $i \leftarrow i + 1$.
Step 5. Construct the grid $\{x^{(i)}_1, ..., x^{(i)}_{n_i}\} = Q_{\min}(t_i, B_{a_i}(\xi_{i-1}))$, where $t_i = v_j(r_i), n_i = n(a_i, t_i)$.
Step 6. $\xi_i \leftarrow \arg\min \{f(x^{(i)}_1), ..., f(x^{(i)}_{n_i}), f(\xi_{i-1})\}$ and pass to step 4.
Step 7. $\eta_j \leftarrow \xi_{i, j} = j + 1, v_j(r) \leftarrow \alpha v_j(r), i \leftarrow i + 1$ and pass to step 2.

With $N_j(\varepsilon)$ we will denote the number of calculations of the objective function $f$, performed by the algorithm at $j$-iteration. The following theorem is just.

Theorem 3. Let the objective function $f$ be $V$-nondegenerate and $\theta \defeq \min\{V(r)/r : 0 < r < \Gamma\}$. Having selected $q, \phi, \alpha \in (0, 1)$ we suppose $R = \Gamma/(1 + q), m = [\ln(\varepsilon/R)/\ln q], r_i = R^{q^i}$ at $1 \leq i \leq m, a_i = \min[r_i + r_{i-1}, \Gamma]$ at $2 \leq i \leq m, v_j(r) = \phi a^{i-1} r$ at $j \geq 1, f = [\ln(\theta/\phi)/\ln \alpha] + 2$. Then, to find algorithm 2, $\eta_j \in M_\varepsilon$ is performed.

Besides, let $n(r, t) \leq v(r/t)$, where the nondecreasing function $v$ depends upon the selected method of constructing grids, the metric used, and the space dimensionality. Then, the following inequality is just.

$$\sum_{j=1}^{j} N_j(\varepsilon) \leq \overline{N}_j(\varepsilon) = [\ln(\varepsilon/R)/\ln q] \sum_{j=1}^{j} v((1 + 1/q)/(\phi a^{i-1})).$$

\[ (9) \]

Proof. Let us denote $\omega = [\ln(\varepsilon/R)/\ln q]$ and note that $m = [\omega]$, and
$$Rq^\omega = R \exp(\ln q \ln(\varepsilon/R)/\ln q) = R \exp(\ln(\varepsilon/R)) = R \varepsilon/R = \varepsilon.$$  

Further, we will obtain
$$r_m = Rq^m = Rq^{\omega+\omega+\omega} = Rq^{\omega} q^{\omega+\omega} = q^{\omega+\omega} \varepsilon < \varepsilon \omega \leq \varepsilon.$$  

Thus, $\omega \leq r_m \leq \varepsilon$.

Let us evaluate the number of iteration $j$, for which the inequality $V^j(r) > v_j(r)$ becomes just. We have $V^j(r) / r \geq \theta^j$, and, so, $V^j(r) \geq \theta^j r$. We will find the value $j$, at which, for $v_j(r) = \phi a^{i-1} r$, the inequality $\theta^j > v_j$ is true. We have $\theta > \phi a^{i-1}, a^{i-1} < \theta/\phi, (j-1) \ln \alpha > \ln(\phi^j)/\ln (\varepsilon/R)$, and, thus, the inequality $V^j(r) > v_j(r)$ is true at $j = j = [\ln(\theta/\phi)/\ln \alpha] + 2$. So, $\eta_j \in M_\varepsilon$.

Provided $n(r, t) \leq v(r/t)$, let us evaluate the value $N_j(\varepsilon)$. For this we will evaluate the values $n_i = n(a_i, t_i)$ at $j$-iteration. We have $t_i = v_j(r_i) = \phi a^{i-1} r_i$. For $i \geq 2 a_i = \min[r_i + r_{i-1}, \Gamma] \leq r_i + r_{i-1}$ is performed, and
$$a_i \leq \frac{r_i + r_{i-1}}{\phi a^{i-1} r_i} \leq \frac{1 + r_{i-1}}{\phi a^{i-1} r_i} = \frac{1 + Rq^{i-1}}{\phi a^{i-1}} = \frac{1 + q}{\phi a^{i-1}}.$$  

\[ (10) \]

For $i = 1 a_1 = \Gamma$ is performed, and
$$a_1 \leq \frac{r_1}{\phi a^{i-1} r_1} = \frac{1}{\phi a^{i-1} Rq} = \frac{1 + q}{\phi a^{i-1} q} = \frac{1 + q}{\phi a^{i-1}}.$$  

\[ (11) \]

Using the formulas (10) and (11), we’ll obtain
$$N_j(\varepsilon) = \sum_{i=1}^{m} n(a_i, t_i) \leq \sum_{i=1}^{m} v(a_i/t_i) \leq \sum_{i=1}^{m} v \left( \frac{1 + q}{\phi a^{i-1}} \right) \leq \left( \frac{1 + q}{\phi a^{i-1}} \right) [\ln(\varepsilon/R)/\ln q].$$  

\[ (12) \]

From the formula (12), immediately the inequality (9) follows. The theorem has been proved.
Comment 2.1. The value $\bar{N}_f(\varepsilon)$ is the upper estimate of the number of calculations of the objective function which is necessary for the first penetration of the search into $\varepsilon$-neighborhood of the extremum. It is clear that $\bar{N}_f(\varepsilon) = O(\ln\varepsilon)$ at $\varepsilon \to 0$. Thus, for $V$-nondegenerate objective functions $f$ (at the absence of a priori information about the behavior $f$) there is such quasi Monte-Carlo method that the number of calculations of the target function, which is necessary for the first penetration of the search into $\varepsilon$-vicinity of the extremum, is dependent logarithmically on the required accuracy of the task solving.

2. In cyclic search of algorithm 2 every next iteration depends on the position of the “better” found earlier point $\eta_{j-1}$. Algorithm 2 may be changed so that the value $\eta_{j-1}$ will not be considered at obtaining new search points. The number of calculations of the nondegenerate objective function $f$, which is necessary for such cyclic search to enter the set $M_0$ for the first time, also look like $O(\ln\varepsilon)$ at $\varepsilon \to 0$. This possibility may be useful at organizing parallel calculations.

3. It is clear that you may use also the other rules of giving values $v_j(r)$ (different from $v_j(r) = \phi a_{j-1}r$), at which the inequalities $v_j(r_j) < V^f(r_j)$ become just after the finite number of iterations.

4. Let us consider the ratio $\bar{N}_j(\varepsilon)/N_j(\varepsilon)$, in which $N_j(\varepsilon)$ is the number of calculations of the objective function $f$, performed by the search of algorithm 2 at $f$-iteration when the values of the search parameters proved to be selected “correctly” (i.e., the inequality $V^f(r_j) > v_j(r_j)$ is true). Thus, the ratio $\bar{N}_j(\varepsilon)/N_j(\varepsilon)$ characterizes the loss of search effectiveness due to the absence of a priori information about the behavior $f$. Let us take the metric $\rho_a$ and cubic grids for which $v(z) = |z|^d$. Using approximate formula $v(z) = z^d$, we will obtain that $\bar{N}_j(\varepsilon)/N_j(\varepsilon) = (1 - a^d)/(1 - a^d)$. In particular, for example, at $\alpha = 1/2$ and $d = 4$ we’ll obtain $(1 - a^d)/(1 - a^d) < 1/(1 - a^d) = 1/15$. That is, the ratio $\bar{N}_j(\varepsilon)/N_j(\varepsilon)$ has proven to be quite close to unity. Thus, considering the ratio $\bar{N}_j(\varepsilon)/N_j(\varepsilon)$, the absence of a priori information about the behavior of $f$ does not tell much on the search effectiveness. It is much more essential that a priori information about the behavior $f$ allows to obtain the condition of stopping the search.

6. Conclusion
The results obtained demonstrate that in the presented variants of the quasi – Monte Carlo method (at optimizing non-degenerate objective functions) the required number of target function calculations look like $O(\ln\varepsilon)$. The results of the work demonstrate that quasi – Monte Carlo methods can be quick (considering the characteristic of dependence of labor intensity on $\varepsilon$), if they are well organized. The convergence rate of the considered methods of search for global function extremum can be compared by the characteristic with the convergence rate of quick methods of local optimization.

References
[1] Vasilev F P 1988 Numerical methods of solving extremal problems Nauka Moscow
[2] Nemirovsky A S and Yudin D B 1979 Complexity of problems and effectiveness of optimization methods Nauka Moscow
[3] Ermakov S M and Zhigljavsky A A 1983 On random search of global extremum Probability theory and its Applications Vol 1 129–136
[4] Ermakov S M, Zhigljavsky A A and Kondratovich M V 1989 On comparison of certain procedures of random search of global extremum Journal of Computational Mathematics and Mathematical Physics. Vol 29 163–170
[5] Zhigljavsky A A and Zilinskas A G 1991 Methods of global extremum search Nauka Moscow
[6] Zhigljavsky A A and Zilinskas A 2008 Stochastic global optimization Springer-Verlag Berlin
[7] Zilinskas A and Zhigljavsky A 2016 Stochastic global optimization: a review on the occasion of 25 years of Informatica Informatica Vol 27 229–256
[8] Spall J C 2003 Introduction to stochastic search and optimization: estimation, simulation, and control Wiley New Jersey
[9] Ingber L 1989 Very fast simulated re-annealing Mathl. Comput. Modelling. Vol 12 967–973
[10] Lopatin A S 2005 Annealing method Stochastic optimization in informatics I 133–149
[11] Granichin O N and Polyak B T 2003 Randomized estimation and optimization algorithms under almost arbitrary noise Nauka Moscow
[12] Sushkov Yu A 1972 On one way to organize random search Operations Research and Statistical Modeling Publishing House LSU Leningrad I 180–186
[13] Tarlowski D 2017 On the convergence rate issues of general Markov search for global minimum, Journal of Global Optimization Vol 69 869–888
[14] Nekrutkin V V and Tikhomirov A S 1993 Speed of convergence as a function of given accuracy for random search methods Acta Applicandae Mathematicae 33 89–108
[15] Tikhomirov A S and Nekrutkin V V 2004 Markovian monotonous search for extremum. Review of some theoretical results. Mathematical models. Theory and applications 4 3–47
[16] Tikhomirov A S 2006 On the Markov homogeneous optimization method Computational Mathematics and Mathematical Physics Vol 46 3 361–375
[17] Tikhomirov A S 2007 On the convergence rate of the Markov homogeneous monotone optimization method Computational Mathematics and Mathematical Physics Vol 47 5 780–790
[18] Tikhomirov A, Stojunina T and Nekrutkin V 2007 Monotonous random search on a torus: integral upper bounds for the complexity Journal of Statistical Planning and Inference Vol 137 12 4031–4047
[19] Tikhomirov A S 2010 On the convergence rate of the simulated annealing algorithm Computational Mathematics and Mathematical Physics Vol 50 1 19–31
[20] Tikhomirov A S 2011 Lower bounds on the convergence rate of the Markov symmetric random search Computational Mathematics and Mathematical Physics Vol 51 9 1524–1538
[21] Tikhomirov A S 2011 On the rate of convergence of one inhomogeneous Markov algorithm of search for extremum Vestnik St. Petersburg University. Mathematics Vol 44 4 309–316
[22] Tikhomirov A S 2018 On the program implementation of a Markov homogeneous monotonous random search algorithm of an extremum IOP Conference Series: Materials Science and Engineering Vol 441 012055 1–8
[23] Niederreiter H and Peart P 1986 Localization of search in quasi-Monte Carlo methods for global optimization SIAM J. Sci. Statist. Computing Vol 7 2. 660–664
[24] Niederreiter H 1992 Random number generation and quasi-Monte Carlo methods CBMS-NSF Regional Conference Series in Applied Math Vol. 63 Society for Industrial and Applied Mathematics
[25] Tikhomirov A S 1992 On labor intensity of search for function extremum Bulletin of St. Petersburg University. Series I 3 (15) 106–107