JOURNAL OF
COMPUTATIONAL AND
APPLIED MATHEMATICS

ELSEVIER Journal of Computational and Applied Mathematics 56 (1994) 283–293

On the crude multidimensional search

Ilya M. Sobol'*, Sergey G. Bakin

Institute for Mathematical Modelling of the Russian Academy of Sciences, 4 Miusskaya Square, Moscow 125047, Russia

Received 5 January 1993; revised 4 August 1993

Abstract

Multivariable trial functions that depend on random parameters are maximized by crude global search. Analytical and numerical investigations of error distributions confirm recent conclusions that in practice random searching points perform better than rectangular lattices, and that quasi-random searching points are even more efficient.

Keywords: Optimization; Optimal algorithms; Crude search; Random search; Quasi-random search

1. Introduction

For several decades a crude search in an n-dimensional cube of large n, say $n > 4$, was regarded as absolutely inefficient. Indeed, if the set of all functions with bounded first partial derivatives is considered and N optimal searching points are selected, the convergence rate will be only $N^{-1/n}$. However, it was stressed in [5] that the last estimate cannot be improved for "bad" functions only and these are functions equally depending on all n variables. On the contrary, if a function depends strongly on a few of these variables, say m and $m \ll n$, the convergence rate may be much better, even $N^{-1/m}$.

Such situations are often encountered in multicriteria optimum design of machines where the total number of decision variables is large and cannot be reduced; however, each individual objective depends strongly on a small number of its own "leading variables". In these problems, on the first stage of investigation a crude search is rather efficient: the accuracy requirements are moderate and all the objectives can be estimated at a relatively small number of common trial points [8].

Modern complex computational problems often include functions that are defined by programs rather than by explicit formulas; it is like a "black box": you put in a point $x = (x_1, \ldots, x_n)$ and

* Corresponding author. E-mail: sobol#3@imamod.msk.su.

SSDI 0377-0427(93)E0204-Y

obtain the value $f(x)$. Of course, it is much more difficult to estimate the influence of each variable than to find (approximately) the supremum of $f(x)$. Therefore, it was recommended in [6, 7] to use computational algorithms that are "uniformly good" (i.e. do not depend on the number of leading variables) rather than optimal algorithms [11] whose construction relies upon the unknown bounds of partial derivatives $\partial f/\partial x_j$.

In the present paper the dependence of error distributions on m (the effective number of leading variables) is investigated. A similar approach was used in [9] where the trial functions were different from ours and the error distributions were different also. Nevertheless, the results of both investigations support the above-mentioned recommendation.

Besides, in [9] there was no counterpart to Theorem 5 from Section 6.

2. Classes of functions

Denote by I the unit interval $0 \leqslant x \leqslant 1$ so that I^n is the n-dimensional unit cube. Consider the set $H = H(L_1, \ldots, L_n)$ of functions $f(x)$ that satisfy the following condition: for an arbitrary $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in I^n

$$|f(x) - f(y)| \leqslant \max_{1 \leqslant j \leqslant n} L_j |x_j - y_j|, \tag{1}$$

where all the constants $L_j \geqslant 0$.

Clearly, classes H are similar to Lipschitz classes: since

$$\max_{1 \leqslant j \leqslant n} L_j |x_j - y_j| \leqslant \sum_{j=1}^{n} L_j |x_j - y_j| \leqslant n \max_{1 \leqslant j \leqslant n} L_j |x_j - y_j|,$$

the set of all functions $f(x)$ that satisfy (1) with all possible nonnegative L_j is identical with the set of all $f(x)$ that satisfy a Lipschitz condition (again, with all possible nonnegative L_j).

3. The crude search

A set of arbitrary fixed searching points $x^{(1)}, \ldots, x^{(N)}$ is called a *net*. As an approximation to

$$f^* = \sup_{x \in I^n} f(x),$$

we may consider the value

$$f_N^* = \max_{1 \leqslant k \leqslant N} f(x^{(k)}).$$

This is the crude searching algorithm.

The usual definition of the approximation error for the class H is

$$d_N = \sup_{f \in H} (f^* - f_N^*).$$

An optimal net for the class H is defined [11] by the requirement

$$d_N \to \min. \tag{2}$$

Theorem 1. *For an arbitrary net $x^{(1)}, \ldots, x^{(N)}$ in I^n the approximation error*

$$d_N \geqslant c_N,$$

where the lower bound

$$c_N = \tfrac{1}{2}\max(L_{j_1} \ldots L_{j_s}/N)^{1/s} \tag{3}$$

and the maximum is extended over all $1 \leqslant j_1 < \cdots < j_s \leqslant n$ and $1 \leqslant s \leqslant n$.

Proof (schematic). The proof of Theorem 1 is similar to the proof of the corresponding theorem in [6]. Only the volume of an s-dimensional parallelepiped must be used rather than the volume of an s-dimensional pyramid in [6]. □

The next theorem shows that the order of the lower bound c_N is the best possible.

Theorem 2. *Consider an arbitrary P_τ-net in I^n containing $N = 2^\nu$ points. If these points are used as searching points in I^n then*

$$d_N \leqslant A c_N,$$

where $A = A(n, \tau)$ depends neither on N nor on $L_1, \ldots, L_n$.

Proof (*schematic*). First, the "worst" function in H (for a fixed net) should be introduced:

$$R(x) = \min_{1 \leqslant k \leqslant N} \max_{1 \leqslant j \leqslant n} L_j |x_j - x_j^{(k)}|; \tag{4}$$

one may easily verify that

$$\sup_{x \in I^n} R(x) = d_N.$$

Second, one may compare c_N and d_N with c_p and d_p in [6] and notice that the ratios d_N/d_p and c_N/c_p are bounded. Then Theorem 2 becomes a corollary of the corresponding theorem in [6]. □

4. Trial functions

Let us consider trial functions

$$f(x, \xi) = -\max_{1 \leqslant j \leqslant n} L_j |x_j - \xi_j|, \tag{5}$$

with a parameter $\xi = (\xi_1, \ldots, \xi_n) \in I^n$. The approximation error for the function (5) can be easily computed: on the one hand, $f^* = -f(\xi, \xi) = 0$ and on the other

$$f_N^* = -\min_k \max_j L_j |x_j^{(k)} - \xi_j| = -R(\xi),$$

so that $f^* - f_N^* = R(\xi)$.

Assume that ξ is a random point uniformly distributed in I^n. Then $\xi_1, \ldots, \xi_n$ are independent random variables uniformly distributed in I. We shall investigate the scaled random error

$$\eta = R(\xi)/c_N. \tag{6}$$

And we shall consider several sets of constants that define the class H:

$$L_1 = \cdots = L_m = 1, \qquad L_{m+1} = \cdots = L_n = 0 \tag{7}$$

for $1 \leqslant m \leqslant n$.

5. Rectangular lattice

Assume that each side of I^n is divided into M equal parts by parallel hyperplanes (Fig. 1). Then I^n is divided into $N = M^n$ equal cubes. Consider the centers of these cubes as searching points.

One may easily verify that the values of $R(x)$ are repeated in each of these cubes: for an arbitrary x the "nearest" point of the net is the center. Even more, the values of $R(x)$ are repeated in each of the 2^n hyperoctants of every cube. Therefore, denoting by $l = (2M)^{-1}$ the side of such a hyperoctant we may conclude that the probability

$$P\{R(\xi) < t\} = P\left\{\max_{1 \leqslant j \leqslant n} L_j \zeta_j < t\right\},$$

where $\zeta_1, \ldots, \zeta_n$ are independent random variables uniformly distributed in $0 \leqslant x \leqslant l$.

It follows from (4) that for the rectangular lattice in Fig. 1

$$d_N = l \max_{1 \leqslant j \leqslant n} L_j.$$

Consider now the set of constants (7). Then

$$P\{R(\xi) < t\} = P\{\zeta_1 < t, \ldots, \zeta_m < t\} = (t/l)^m.$$

At the same time according to (3), $c_N = 1/(2N^{1/m})$. So,

$$P\{\eta < z\} = P\{R(\xi) < c_N z\} = (c_N z/l)^m,$$

and this result can be formulated as follows.

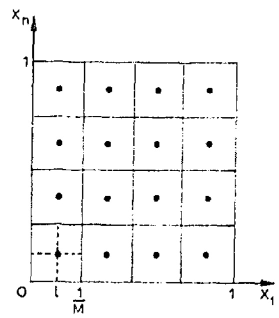

Fig. 1. Rectangular lattice, $N = M^n$, $M = 2$.

Theorem 3. *For the set of constants* (7) *and for the rectangular lattice the distribution function of the random variable* η *is*

$$P\{\eta < z\} = (z/T)^m, \quad 0 \leqslant z \leqslant T, \tag{8}$$

where $T = N^{1/m - 1/n}$.

Corollary 4. *If the assumptions of Theorem* 3 *are fulfilled, the expectation and the standard deviation of* η *are*

$$M\eta = \frac{m}{m+1} T, \qquad \sigma(\eta) = \frac{1}{m+1}\sqrt{\frac{m}{m+2}}\, T.$$

In all optimization theories before 1987 only the symmetric case $L_1 = \cdots = L_n$ (i.e. $m = n$) has been considered. In that case [4] the rectangular lattice is optimal (cf. (2)): $d_N = c_N = 1/(2N^{1/n})$, and the upper bound of η is $T = 1$. However for $m < n$ the upper bound $T \to \infty$ as $N \to \infty$. And for $m \ll n$ the rectangular lattice is catastrophically bad (cf. Section 8).

Remark. In our numerical experiment we have used more sophisticated sets of constants:

$$L_1 = \cdots = L_m = 1, \qquad L_{m+1} = \cdots = L_n = \varepsilon, \tag{9}$$

with $\varepsilon \ll 1$. If $N^{1/m}\varepsilon < 1$ then from (3) one may see that still $c_N = 1/(2N^{1/m})$. And the distribution function (8) will be changed to

$$P\{\eta < z\} = \begin{cases} (z/T)^n \varepsilon^{-(n-m)}, & 0 \leqslant z \leqslant \varepsilon T, \\ (z/T)^m, & \varepsilon T \leqslant z \leqslant T. \end{cases}$$

The formulas for $M\eta$ and $\sigma(\eta)$ given in the last corollary will acquire a factor $1 + O(\varepsilon^2)$.

6. Random nets

Consider searching points $x^{(1)}, \ldots, x^{(N)}$ that are independent random points uniformly distributed in I^n. According to (4), the probability

$$P\{R(\xi) \geqslant t\} = P\left\{\max_{1 \leqslant j \leqslant n} L_j |x_j^{(k)} - \xi_j| \geqslant t \text{ for } 1 \leqslant k \leqslant N\right\}.$$

If the set (7) of constants is considered then

$$P\{R(\xi) \geqslant t\} = P\left\{\max_{1 \leqslant j \leqslant m} |x_j^{(k)} - \xi_j| \geqslant t \text{ for } 1 \leqslant k \leqslant N\right\}.$$

Let us introduce an auxiliary function λ_t whose value is equal to the volume of the union of parallelepipeds belonging to I^m where the following condition is satisfied:

$$\lambda_t(y_1, \ldots, y_m) = \operatorname{mes}\left\{x \in I^m \,\middle|\, \max_{1 \leqslant j \leqslant m} |x_j - y_j| \geqslant t\right\}.$$

One may easily see that

$$P\{R(\xi) \geqslant t\} = \int_0^1 \cdots \int_0^1 [\lambda_t(y_1, \dots, y_m)]^N \, \mathrm{d}y_1 \cdots \mathrm{d}y_m.$$

It is clear from the definition of λ_t that

$$\lambda_t(y_1, \dots, y_m) = 1 - \operatorname{mes}\left\{x \in I^m \,\middle|\, \max_{1 \leqslant j \leqslant m} |x_j - y_j| < t\right\}.$$

$$= 1 - \prod_{j=1}^{m} \operatorname{mes}\{x_j \in I \,|\, |x_j - y_j| < t\} = 1 - \prod_{j=1}^{m} \mu_t(y_j),$$

where

$$\mu_t(y) = \operatorname{mes}\left\{x \in I \,\middle|\, |x - y| < t\right\}.$$

From geometric considerations, $\mu_t(1 - y) = \mu_t(y)$. Therefore, in the last integral each integration from 0 to 1 can be replaced by two integrals from 0 to $\frac{1}{2}$. Hence,

$$P\{R(\xi) \geqslant t\} = 2^m \int_0^{1/2} \cdots \int_0^{1/2} \left\{1 - \prod_{j=1}^{m} \mu_t(y_j)\right\}^N \mathrm{d}y_1 \cdots \mathrm{d}y_m, \tag{10}$$

and for $\mu_t(y)$ explicit expressions can be written: for $0 \leqslant t \leqslant \frac{1}{2}$

$$\mu_t(y) = \begin{cases} t + y, & 0 \leqslant y \leqslant t, \\ 2t, & t \leqslant y \leqslant \frac{1}{2} \end{cases}$$

and for $\frac{1}{2} \leqslant t \leqslant 1$

$$\mu_t(y) = \begin{cases} t + y, & 0 \leqslant y \leqslant 1 - t, \\ 1, & 1 - t \leqslant y \leqslant \frac{1}{2}. \end{cases}$$

In the proof of the next theorem only the first pair of these formulas will be used.

Theorem 5. *For the set of constants* (7) *and for random searching points the limit distribution function of the random variable* η *is*

$$\lim_{N \to \infty} P\{\eta < z\} = 1 - \exp(-z^m), \quad 0 \leqslant z < \infty. \tag{11}$$

Proof. Since $[0, \frac{1}{2}] = [0, t) + [t, \frac{1}{2}]$, the region of integration in (10) may be split into a sum of 2^m parallelepipeds. We shall prove that as $N \to \infty$ the main term of the distribution function is the integral over $[t, \frac{1}{2}]^m$ where all $\mu_t \equiv 2t$.

First, let us compute this term:

$$2^m \int_t^{\frac{1}{2}} \cdots \int_t^{\frac{1}{2}} [1 - (2t)^m]^N \mathrm{d}y_1 \cdots \mathrm{d}y_m = (1 - 2t)^m (1 - (2t)^m)^N.$$

Second, consider one of the other parallelepipeds, e.g. $[0,t)^s \times [t,\frac{1}{2}]^{m-s}$, with $0 < s \leqslant m$; the corresponding integral

$$2^s \int_0^t \cdots \int_0^t \left\{1 - \prod_{j=1}^{s} (t + y_j)(2t)^{m-s}\right\}^N \mathrm{d}y_1 \cdots \mathrm{d}y_s (1 - 2t)^{m-s}$$

does not exceed $(2t)^s$.

We are now able to estimate the probability

$$P\{\eta \geqslant z\} = P\{R(\xi) \geqslant z/(2N^{1/m})\}.$$

Using (10) with $t = z/(2N^{1/m})$ we obtain the main term

$$(1 - z/N^{1/m})^m (1 - z^m/N)^N$$

that with $N \to \infty$ tends to $\exp(-z^m)$. Each one of the other terms does not exceed $(2t)^s = z^s/N^{s/m}$ and tends to zero as $N \to \infty$. □

Corollary 6. *If the assumptions of Theorem 5 are fulfilled, the expectation and the standard deviation of η tend to*

$$\lim_{N\to\infty} M\eta = \Gamma\left(1 + \frac{1}{m}\right), \qquad \lim_{N\to\infty} \sigma(\eta) = \left(\Gamma\left(1 + \frac{2}{m}\right) - \Gamma^2\left(1 + \frac{1}{m}\right)\right)^{1/2}.$$

Remark. Probability distribution (11) is often called the reduced Weibull distribution.

7. Computational experiments

In the eight-dimensional cube I^s functions (5) with constants (9) at $\varepsilon = 0.001$ and three types of searching nets were considered.

RND: For each computation a net of $N = 256$ independent random points and a random point ξ were selected; from (4) the value $R(\xi)$ and from (6) the value η were computed. Using 1024 independent values of η a normed histogram approximating the density $p_\eta(z)$ was constructed (with $\Delta z = 0.1$), and $M\eta$, $D\eta$ and $\sup \eta$ were estimated.

LAT: A rectangular lattice (Fig. 1) with $N = 2^8 = 256$ points was fixed. For each computation a random point ξ was selected, from (4) and (6) the values $R(\xi)$ and η were computed, and so on...

LPT: As a searching net, $N = 256$ initial points of a quasi-random LP_τ-sequence [1, 2, 10] were fixed. Initial segments containing $N = 2^\nu$ points of such sequences at all $\nu \geqslant \nu_0$ are P_τ-nets and satisfy Theorem 2. For each computation a random point ξ was selected, and so on...

In all experiments the random points ξ were obtained from a pseudo-random number generator that had been tested in [3]. The random searching points in RND were computed from nonoverlapping parts of the same pseudo-random sequence.

Table 1
Computational results for different m at $\varepsilon = 0.001$

m		1	2	3	4	5	6	7	8
	LAT	63.7	5.23	2.33	1.59	1.25	1.07	0.96	0.89
$\bar{\eta}$	RND	1.04	0.90	0.93	1.00	1.03	1.07	1.11	1.15
	LPT	0.56	0.74	0.84	0.91	0.98	1.04	1.07	1.11
	LAT	128.0	8.0	3.2	2.0	1.5	1.3	1.1	1.0
η_N^*	RND	6.9	2.7	2.4	2.3	2.1	2.0	2.0	1.9
	LPT	1.8	1.9	1.9	1.9	1.9	2.0	1.8	1.9
	LAT	37.0	1.9	0.62	0.32	0.22	0.16	0.12	0.10
σ	RND	0.91	0.47	0.36	0.33	0.29	0.27	0.24	0.23
	LPT	0.22	0.31	0.29	0.27	0.25	0.24	0.23	0.22

8. Numerical results

Table 1 contains the computed empirical estimates $\bar{\eta} \approx M\eta$, $\sigma \approx (D\eta)^{1/2}$ and $\eta_N^* \approx \sup \eta = d_N/c_N$. Although the distributions of η are different from those obtained in [9], the conclusions from Table 1 confirm the main conclusions in [9].

First, at $m = 8$ when the trial functions depend on all the variables $x_1, \ldots, x_8$ equally the rectangular lattice is optimal, the least mean error $\bar{\eta}$ is for LAT. However, the difference between the performances of all three types of searching points is small. In other words, the class of functions H with $L_1 = \cdots = L_n$ is so "bad", that an average random net is almost as good as the optimal one.

As m is decreased, the mean error for LAT is increasing while the mean errors for RND and LPT are decreasing. At $m = 6$ the mean errors for all three types of nets are practically equal. But at $m \leqslant 3$ the mean errors for RND and LPT are much smaller than for LAT.

Second, at all m's the mean errors for LPT are better than the mean errors for RND, though at $m \geqslant 4$ the difference is slight.

Third, the increase of the mean errors for LAT at $m = 2$ and $m = 1$ may be regarded as catastrophic.

The maximum values η_N^* in Table 1 behave, in general, like $\bar{\eta}$. Here the line containing η_N^* for LPT looks very spectacular: all the values are almost the same. This can be interpreted as an illustration to Theorem 2: the ratios d_N/c_N are bounded by A that does not depend on $L_1, \ldots, L_n$.

Despite $\varepsilon \neq 0$, almost all values in Table 1 for LAT agree with values obtained from formulas in Section 5. Table 2 enables the comparison of computed $\bar{\eta}$ for RND with limit values obtained from Theorem 5 at $N = \infty$.

Clearly, the most reliable net for practical computations is LPT though here we have no analytical estimates of $\bar{\eta}$.

Table 2
Computed $\bar{\eta}$ and $\lim_{N\to\infty} M\eta$ for RND

m	1	2	3	4	5	6	7	8
$N = 256$	1.04	0.90	0.93	1.00	1.03	1.07	1.11	1.15
$N = \infty$	1.00	0.886	0.893	0.906	0.918	0.927	0.935	0.942

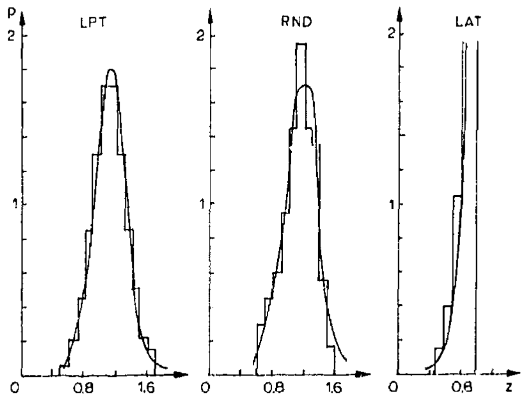

Fig. 2. Error distributions for $m = 8$ (symmetric case).

9. Error distributions

At $m = 8$ (Fig. 2) the error distributions for RND and LPT are approximately Gaussian $(a; \sigma)$ with $a = \bar{\eta}$ and σ from Table 1. The histogram for LAT is in full agreement with the power law (8).

As m is decreased, the distributions for RND and LPT deviate for Gaussian. In Fig. 3 all the distributions at $m = 3$ are plotted. Here again the histogram for LAT is in good agreement with (8), while the histogram for RND may be regarded as an approximation to the asymptotic law (11).

In general, the agreement of RND histograms with the law (11) worsens as m is increased. This may be explained by the fact that the main omitted term in the proof of Theorem 5 was $O(N^{-1/m})$.

Finally, consider the limit case $m = 1, \varepsilon = 0$. In this case the classes H are identical with Lipschitz classes that were investigated in [9] where analytical error distributions were found: for LPT the density is

$$p_\eta(z) = \begin{cases} 1 - (2N)^{-1}, & 0 \leqslant z < 1, \\ (2N)^{-1}, & 1 < z \leqslant 2; \end{cases}$$

for RND the density is $p_\eta = e^{-z}, 0 < z < \infty$; and for LAT the density is (8): uniform distribution in a huge interval $0 < z < T = 128$.

The histograms in Fig. 4 are for RND and LPT. In both cases there are remarkable distortions at small z due to $\varepsilon \neq 0$. The smooth dotted curves are $p = 8240z^7$.

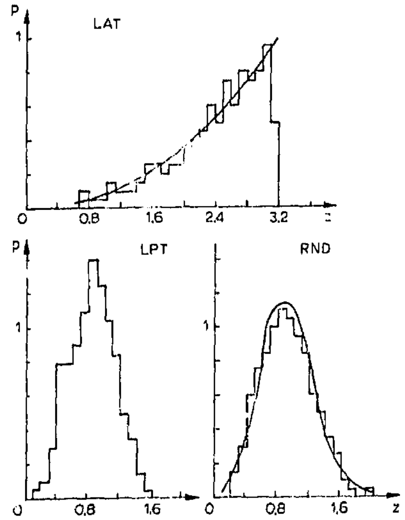

Fig. 3. Error distributions for $m = 3$ (nonsymmetric case).

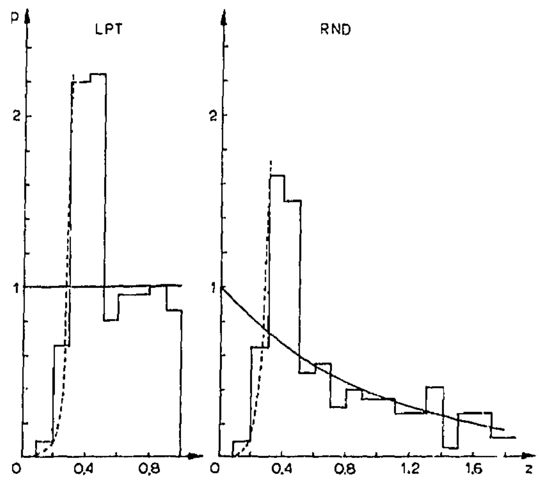

Fig. 4. Error distributions for $m = 1$ (strong nonsymmetry).

The power law z^{n-1} can be derived from quite general considerations. Assume that each node $x^{(k)}$ is surrounded by a box where

$$\max_{1 \leqslant j \leqslant n} L_j |x_j - x_j^{(k)}| < t.$$

Its volume is $(2t/L_1) \cdots (2t/L_n)$. For sufficiently small t intersections of such boxes with the boundaries of I^n can be neglected and the probability

$$P\{R(\xi) < t\} = N(2t)^n/(L_1 \dots L_n).$$

In our case $L_1 = 1$, $L_2 = \cdots = L_n = \varepsilon$, $c_N = (2N)^{-1}$ and we conclude that for sufficiently small z

$$P\{\eta < z\} = z^n/(\varepsilon N)^{n-1}.$$

Hence, the density of η is proportional to z^{n-1}.

References

[1] F. Bratley and B.L. Fox, Implementing Sobol's quasi-random sequence generator, *ACM Trans. Math. Software* **14** (1) (1988) 88–100.

[2] Yu.L. Levitan, N.I. Markovich, S.G. Rozin and I.M. Sobol', On quasirandom sequences for numerical computations, *USSR Comput. Math. Math. Phys.* **28** (3) (1988) 88–92.

[3] Yu.L. Levitan and I.M. Sobol', On a pseudo-random number generator for personal computers, *Math. Modelling* **2** (8) (1990) 119–126 (in Russian).

[4] H. Niederreiter, Quasi-Monte Carlo methods for global optimization, in: *Proc. 4th Pannonian Symp. on Math. Statistics* (Austria, 1983) (Akad. Kiado, Budapest, 1986) 251–267.

[5] I.M. Sobol', On functions satisfying a Lipschitz condition in multidimensional problems of numerical mathematics, *Dokl. Akad. Nauk SSSR* **293** (1987) 1314–1319.

[6] I.M. Sobol', On the search for extremal values of functions of several variables satisfying a general Lipschitz condition, *USSR Comput. Math. Math. Phys.* **28** (2) (1988) 112–118.

[7] I.M. Sobol', Quadrature formulas for functions of several variables satisfying a general Lipschitz condition, *USSR Comput. Math. Math. Phys.* **29** (3) (1989) 201–206.

[8] I.M. Sobol', An efficient approach to multicriteria optimum design problems, *Surveys Math. Indust.* **1** (4) (1992) 259–281.

[9] I.M. Sobol' and Yu.L. Levitan, On stochastic modelling of error distribution in a crude global search, *Stochastic Optim. Design* **1** (1) (1992) 73–84.

[10] I.M. Sobol', V.I. Turchaninov, Yu.L. Levitan and B.V. Shukhman, *Quasirandom Sequence Generators* (Inst. Appl. Math. Moscow, 1992).

[11] J.F. Traub and H. Woźniakowski, *A General Theory of Optimal Algorithms* (Academic Press, New York, 1980).