

## AVERAGE PERFORMANCE OF QUASI MONTE CARLO METHODS FOR GLOBAL OPTIMIZATION

Hisham A. Al-Mharmah

Industrial Engineering Department  
University of Jordan  
Amman 11942, JORDAN

### ABSTRACT

In this paper we compare the average performance of one class of low-discrepancy quasi-Monte Carlo sequences for global optimization. Weiner measure is assumed as the probability prior on all optimized functions. We show how to construct van der Corput sequences and we prove their consistency. Numerical experimentation shows that the van der Corput sequence in base 2 has a better average performance.

### 1 INTRODUCTION

The Monte Carlo method may be described, in simple terms, as a numerical method based on random sampling. It is therefore a method with a strong statistical and probabilistic flavor. Quasi-Monte Carlo methods are the deterministic version of the classical Monte Carlo methods, in the sense that the random samples in the Monte Carlo method are replaced by well-chosen deterministic points.

For numerical integration, the Monte Carlo method promises integration errors for which the order of magnitude, in terms of the number of nodes, is independent of the dimension. However, it yields only a probabilistic (unguaranteed) bound on the integration error. Moreover, the analysis shows that a deterministic error bound can be established if deterministic nodes are used. This leads to the idea of selecting deterministic nodes in such a way that the error bound is as small as possible. This idea expresses the fundamental principle of quasi-Monte Carlo Method. The main aim is to select deterministic points for which the deterministic bound is smaller than the probabilistic Monte Carlo error bound. In fact, for many Monte Carlo methods, it is possible to develop corresponding quasi-Monte Carlo methods as their deterministic versions.

There are quasi-Monte Carlo methods not only for numerical integration, but also for various other numerical

problems such as optimization. Quasi-Monte Carlo based (deterministic) algorithms have shown a superior average performance (asymptotically) compared to other Monte Carlo based (random) algorithms for approximating the global maximum of a Brownian motion. Calvin (1995) has shown that, for the Brownian motion prior, if observations form a deterministic equi-spaced grid, then the average error is about 82% as large as if the points are chosen at random uniformly over the unit interval. Ritter (1990) investigated the optimality of the equi-spaced grid algorithm for finite number of observations. He showed that it is optimal for two observations, but not in general. Al-Mharmah and Calvin (1997) has shown that a deterministic algorithm that uses the images of a certain deterministic sequence under the inverse cumulative probability function of the beta distribution has a better asymptotic average performance than the optimal Monte-Carlo based algorithm. The question of the optimal deterministic sequence that could minimize the average value of the approximation error is still an interesting question that needs to be answered.

The purpose of this paper is to present a comparison of the average performance on several quasi-Monte Carlo sequences for approximating the global maximum of one-dimensional real-valued functions defined on the unit interval. The average performance is defined as the expected difference between the observed maximum value and the actual global maximum. For that, Brownian motion will be used as a probabilistic model for the one-dimensional continuous functions, and the objective function is to be taken as one realization of a Brownian motion process. We use an average-case framework, which can be thought of as averaging the error over many independent realizations of the algorithm on different objective functions. The concept of average optimality is more useful in this setting than the worst case analysis where the error can be arbitrarily large unless stringent conditions are placed on the set of objective functions, such

as convexity and differentiability. Optimization algorithms based on average optimality are surveyed in Mockus (1989), Törn and Žilinskas (1989), and Betró (1991). Applications of quasi-Monte Carlo methods are surveyed in Niederreiter (1992).

In the next section we introduce the problem and the terminology. In Section 3 we show how to construct van der Corput sequences as a class of low discrepancy quasi-Monte Carlo sequences and we discuss their dispersion and consistency characteristics. In Section 4 we compare the average performance of different van der Corput sequences.

## 2 NOTATION AND TERMINOLOGY

Given the set of continuous real-valued functions defined on the unit interval  $C([0,1])$ , let  $f \in C([0,1])$  be the objective function to be maximized. Also, let  $f^* = \max\{f(t); t \in [0,1]\}$  denote its global maximum. To approximate  $f^*$  we assume that we are allowed to observe the function  $f$  at  $n$  locations. Let  $t_1, t_2, \dots, t_n$  be the sequence of the observation sites in  $[0,1]$ , and let  $M_n$  be the maximum of the  $n$  observed values;

$$M_n = \max_{1 \leq i \leq n} f(t_i).$$

Our goal is to compare the average performance of several deterministic sequences based on their average approximation error  $E(\Delta_n)$ , where

$$\Delta_n = \max_{1 \leq i \leq n} f(t_i) - f^*.$$

Suppose that we have some prior knowledge about the relative likelihood of various functions, and that we can formalize this knowledge in a form of a probability measure  $\mu$  on  $C([0,1])$ . Consequently, we can view any function  $f \in C([0,1])$  as a sample path of a stochastic process. Hence,

$$E(\Delta_n) = \int_{f \in C([0,1])} \left( \max_{1 \leq i \leq n} f(t_i) - f^* \right) d\mu(f). \quad (1)$$

The Wiener measure on  $C([0,1])$  will be taken as the probability distribution; i.e.,  $f$  is taken to be a sample path of a Brownian motion process. It is natural to use a Gaussian measure, such as the Wiener measure, as a model for a random objective function that has multiple local minima with positive probability. Brownian motion is one of only a few non-trivial stochastic processes for which the distribution of the minimum is even known.

Consistency of any optimization method means that the observed maximum converges to the global maximum as the number of observations increases to  $\infty$ , i.e.

$$\lim_{n \rightarrow \infty} \max\{f(t_1), f(t_2), \dots, f(t_n)\} \rightarrow f^*,$$

for all continuous functions on  $C[0,1]$ . Discrepancy of a set of observation sites can be viewed as a quantitative measure for the irregularity of distribution. For the one-dimensional case, discrepancy can be expressed in different forms such as star discrepancy  $D_n^*(S)$  and extreme discrepancy  $D_n(S)$ , where (for  $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ )

$$D_n^*(t_1, t_2, \dots, t_n) = \frac{1}{2n} + \max_{1 \leq i \leq n} \left| t_i - \frac{2i-1}{2n} \right|,$$

and

$$D_n(t_1, t_2, \dots, t_n) = \frac{1}{n} + \max_{1 \leq i \leq n} \left( \frac{i}{n} - t_i \right) - \min_{1 \leq i \leq n} \left( \frac{i}{n} - t_i \right).$$

A low-discrepancy sequence is informally defined as a sequence  $S$  for which  $D_n^*(S)$  or  $D_n(S)$  is small for  $n \geq 1$ .

Dispersion may be viewed as a measure for the deviation from denseness. Formally speaking, let  $(T, d)$  be a bounded metric space and let  $S$  be a sequence that consists of  $t_1, t_2, \dots, t_n \in T$ . If we order the points of  $S$  such that  $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ , then the dispersion  $d_n(S)$  of  $S$  in  $T$  is defined as

$$d_n(S) = \max (t_1, 0.5(t_2 - t_1), 0.5(t_3 - t_2), \dots, 0.5(t_n - t_{n-1}), 1 - t_n).$$

Discrepancy and dispersion of a certain set (or sequence) of observations are related to each other. Niederreiter (1992) shows that every low-discrepancy set (or sequence) is a low-dispersion point set (or sequence) but not conversely. Therefore, we will limit our average performance comparison to the average performance of one class of low-discrepancy sequences, the van der Corput sequences.

A composite non-adaptive algorithm is one that maintains its form as the number of observations increases; see Zhigljavsky (1991). If we denote the set of observations made by an algorithm up to time  $n$  by  $S_n = \{t_1, t_2, \dots, t_n\}$ , then we will call an algorithm composite if  $S_{n+1} \supset S_n$ . One clear advantage of van der Corput sequences is compositeness. A consequence of compositeness is that there is no need to determine in advance how many observations are to be taken in order to construct the observation set. In contrast, non-composite algorithms do not adapt gracefully as the number of observations changes. An example is the "uniform grid" algorithm (non-composite) that takes equally spaced observations; if a total of  $n$  observations are to be made, they are placed at  $1/n, 2/n, \dots, 1$ . However, if the number of observations is increased to  $n+1$ , there is no way to add an observation point so as to maintain a uniform grid.

### 3 VAN DER CORPUT SEQUENCES

#### 3.1 Construction

Van der Corput sequence in base  $b$  is a low-discrepancy sequence and can be constructed as follows (see Niederreiter 1992): for an integer  $b \geq 2$ , we put  $Z_b = \{0, 1, \dots, b-1\}$ ; i.e.,  $Z_b$  is the least residue system mode  $b$ . Every integer  $n \geq 0$  has a unique digit expansion

$$n = \sum_{j=0}^{\infty} a_j(n)b^j \tag{2}$$

in base  $b$ , where  $a_j(n) \in Z_b$  for all  $j \geq 0$  and  $a_j(n) = 0$  for all sufficiently large  $j$ . Also, let  $\phi_b$  be the radical-inverse function in base  $b$  for an integer  $b \geq 2$ , where

$$\phi_b(n) = \sum_{j=0}^{\infty} \frac{a_j(n)}{b^{j+1}} \tag{3}$$

for all integers  $n \geq 0$ . Thus  $\phi_b(n)$  is obtained from  $n$  by a symmetric reflection of the expansion (2) in the decimal point. Then, for an integer  $b \geq 2$ , the van der Corput sequence in base  $b$  is the sequence  $S_b = \{t_1, t_2, \dots\}$  with

$$t_n = \phi_b(n), \tag{4}$$

for all  $n \geq 0$ . Following the above approach, Table 1 shows the first ten observation sites for five van der Corput sequences in base 2,3,4, and 5 respectively.

For the purpose of comparing the average performance of the above sequences, we continue generating sites from all sequences as will be shown in Section 4.

#### 3.2 Consistency

Törn and Žilinskas (1989) discussed the necessary and sufficient conditions for the convergence of a global optimization method for continuous functions. They proved that a global optimization algorithm converges to the global extremum of any continuous function if and only if the sequence of the sampling observations is everywhere dense in the function domain. Hence to prove consistency of a certain algorithm, it suffices to show that the observations become dense everywhere. The above discussion suggests that the desirable observation sites should be evenly distributed over the unit interval. In Theorem 1 we show that any van der Corput sequence in base  $b$  is consistent.

**Theorem 1** *Let  $f$  be a continuous real-valued function on  $C[0, 1]$ . The set of observation sites  $\{t_1, t_2, \dots, t_n\}$  generated as a van der Corput sequence  $S_b$  and applied to  $f$  is dense in  $[0, 1]$ .*

*Proof.*

Let  $0 \leq k_1 \leq k_2 \leq 1$ . Consider an algorithm that places the observation sites according to a van der Corput sequence  $S_b$  in base  $b$ , and assume that no observation will ever be made in the interval  $(k_1, k_2)$ . Therefore,

$$t_n = \sum_{j=0}^{\infty} \frac{a_j(n)}{b^{j+1}}$$

should be either less than  $k_1$  or greater than  $k_2$  for all  $n = 1, 2, \dots$ . However, as  $n$  increases to  $\infty$ , there exists  $m$  such that

$$n = \prod_{j=0}^m (b-1)b^j,$$

where  $m$  is positive integer. Hence, the constants in Equation (2) will take the values

$$a_0 = a_1 = \dots = a_m = b-1,$$

and

$$a_{m+1} = a_{m+2} = \dots = a_{\infty} = 0.$$

Therefore, from Equation (4),

$$\begin{aligned} t_n &= \sum_{j=0}^m \frac{b-1}{b^{j+1}}, \\ &= \frac{b-1}{b} + \frac{b-1}{b^2} + \dots + \frac{b-1}{b^{m+1}}, \\ &= 1 - \frac{1}{b} + \frac{1}{b} - \frac{1}{b^2} + \frac{1}{b^2} - \dots - \frac{1}{b^m} \\ &= 1 - \frac{1}{b^m}, \end{aligned}$$

which converges to 1 as  $n \rightarrow \infty$ . Therefore,  $t_n$  will be, eventually, greater than  $k_2$ . Similarly, we can show that  $t_n$  will eventually converge to zero, and, hence, less than  $k_1$  if we pick

$$n = \prod_{j=0}^m b^j,$$

and let  $m \rightarrow \infty$  as  $n \rightarrow \infty$ . Therefore, eventually, there exists  $t_n \in (k_1, k_2)$  which contradicts our assumption.

### 4 NUMERICAL EXPERIMENTATION

Simulation is used to compare the average performance of van der Corput sequences in base 2,3,4, and 5 by estimating their average approximation error. All averages are based on 50000 independent runs. In each run the function values are generated at the sampling locations of a van der Corput sequence from a Brownian motion prior. Tables 2, 3, 4, and 5 show the average value of the observed maximum  $E(M_n)$  as well as the average

Table 1: Sample Observation Sites for the Van Der Corput Sequences in Base 2,3,4,and 5

$n$	$S_2$	$S_3$	$S_4$	$S_5$
1	0.5	0.333	0.25	0.2
2	0.25	0.667	0.5	0.4
3	0.75	0.111	0.75	0.6
4	0.125	0.444	0.0625	0.8
5	0.625	0.778	0.3125	0.04
6	0.375	0.222	0.5625	0.24
7	0.875	0.556	0.8125	0.44
8	0.0625	0.889	0.125	0.64
9	0.5625	0.03704	0.375	0.84
10	0.3125	0.3704	0.625	0.08

Table 2: Simulation Output for Van Der Corput Sequence  $S_2$

$n$	$E(M_n)$	$E(\Delta_n)$
2	0.481798	0.316087
3	0.520939	0.276946
4	0.555338	0.242547
5	0.574304	0.223581
6	0.588510	0.209374
7	0.600230	0.197655
8	0.617022	0.180863
9	0.626513	0.171372
10	0.632028	0.165857
20	0.676564	0.121320
30	0.696680	0.101205
40	0.710129	0.087755
50	0.718799	0.079086

Table 3: Simulation Output for Van Der Corput Sequence  $S_3$

$n$	$E(M_n)$	$E(\Delta_n)$
2	0.477751	0.320133
3	0.526043	0.271842
4	0.551657	0.246227
5	0.569908	0.227976
6	0.589711	0.208174
7	0.601422	0.196463
8	0.611748	0.186136
9	0.626158	0.171726
10	0.634196	0.163688
20	0.675856	0.122029
30	0.698688	0.099197
40	0.708912	0.088972
50	0.717341	0.080544

Table 4: Simulation Output for Van Der Corput Sequence  $S_4$

$n$	$E(M_n)$	$E(\Delta_n)$
2	0.472260	0.325625
3	0.520682	0.277203
4	0.555369	0.242515
5	0.573351	0.224533
6	0.585096	0.212788
7	0.596053	0.201831
8	0.609566	0.188319
9	0.619486	0.178399
10	0.626613	0.171271
20	0.675261	0.122623
30	0.692840	0.105045
40	0.707126	0.090759
50	0.718415	0.079470

Table 5: Simulation Output for Van Der Corput Sequence  $S_5$

$n$	$E(M_n)$	$E(\Delta_n)$
2	0.467452	0.330433
3	0.513808	0.284077
4	0.549663	0.248221
5	0.576710	0.221175
6	0.590387	0.207498
7	0.598750	0.199134
8	0.606350	0.191535
9	0.614038	0.183847
10	0.623717	0.174168
20	0.675547	0.122337
30	0.696444	0.101441
40	0.705855	0.092029
50	0.714849	0.083036

approximation error  $E(\Delta_n)$  after placing  $n$  observations as four van der Corput sequences in base 2,3,4, and 5 respectively.

Linear regression is used to fit linear models to the estimates of the expected error values as function of the observation number for van der Corput sequences in base 2,3,4, and 5 respectively. Table 6 summarizes the outcomes, and shows that placing observations as an  $S_5$  sequence will be as efficient as 98% compared to  $S_2$ .

Table 6: Linear Regression Models

Sequence	The Best Fit Model
$S_2$	$E(\Delta_n) = \frac{0.442755}{n^{0.35}}$
$S_3$	$E(\Delta_n) = \frac{0.444450}{n^{0.35}}$
$S_4$	$E(\Delta_n) = \frac{0.446752}{n^{0.35}}$
$S_5$	$E(\Delta_n) = \frac{0.448246}{n^{0.35}}$

## 5 CONCLUSIONS

In this paper we compare the average performance of four low discrepancy quasi-Monte Carlo sequences, the van der Corput sequences in base 2,3,4,and 5. Numerical experimentation has shown that the sequence in base 2 has slightly better average performance as compared to other types of sequences. The asymptotic average performance of the investigated sequences will be addressed elsewhere.

## REFERENCES

Al-Mharmah, H. and J. Calvin 1997. Comparison of Monte Carlo and deterministic methods for non-adaptive optimization. *Proceedings of the 1997 Winter Simulation Conference*: 348-351.

Betró, B. 1991. Bayesian methods in global optimization. *Journal of Global Optimization* 1:1-14.

Calvin, J. 1995. Average performance of non-adaptive algorithms for global optimization. *Journal of Mathematical Analysis and Applications* 191:608–617.

Mockus, J. 1989. *Bayesian Approach to Global Optimization: Theory and Applications*. Dordrecht: Kluwer.

Niederreiter, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, Philadelphia.

Ritter, K. 1990. Approximation and optimization on the Wiener space, *Journal of Complexity* 6:337-364.

Törn, A. and A. Žilinskas 1989. *Global Optimization*. Berlin: Springer-Verlag.

Zhigljavsky, A. 1991. *Theory of Global Random Search*, Dordrecht: Kluwer.

## AUTHOR BIOGRAPHIES

**HISHAM A. AL-MHARMAH** is an Assistant Professor in the Department of Industrial Engineering at the University of Jordan. He received a B.S. in Civil Engineering and an M.S. in Industrial Engineering from the University of Jordan in 1986 and 1990 respectively. He joined the school of Industrial and Systems Engineering at Georgia Institute of Technology in 1991, and received a Ph.D. in Industrial Engineering in 1993. His research interests include Applied Probability and Stochastic Optimization.