

# A New Optimized Stochastic Approach for Multidimensional Integrals in Machine Learning

Venelin Todorov

Bulgarian Academy of Sciences  
Institute of Mathematics and Informatics  
ul. G. Bonchev 8, 1113 Sofia, Bulgaria  
Bulgarian Academy of Sciences

Institute of Information and Communication Technologies  
ul. G. Bonchev 25A, 1113 Sofia, Bulgaria  
Email: vtodorov@math.bas.bg, venelin@parallel.bas.bg

Stoyan Apostolov

Sofia University  
Faculty of Mathematics and Informatics  
1764 Sofia, Bulgaria  
Email: stoyanapostolov@gmail.com

Ivan Dimov

Bulgarian Academy of Sciences  
Institute of Information and Communication Technologies  
ul. G. Bonchev 25A, 1113 Sofia, Bulgaria  
Email: ivdimov@bas.bg

Stefka Fidanova

Bulgarian Academy of Sciences  
Institute of Information and Communication Technologies  
ul. G. Bonchev 25A, 1113 Sofia, Bulgaria  
Email: stefka@parallel.bas.bg

**Abstract**—Stochastic techniques have been developed over many years in a range of different fields, but have only recently been applied to the problems in machine learning. A fundamental problem in this area is the accurate evaluation of multidimensional integrals. An introduction to the theory of the stochastic optimal generating vectors has been given. A new optimized lattice sequence with a special choice of the optimal generating vector has been applied to compute multidimensional integrals up to 30-dimensions. Clearly, the progress in the area of machine learning is closely related to the progress in reliable algorithms for multidimensional integration.

## I. INTRODUCTION

**M**ONTE Carlo methods are suitable for mathematical modelling of multi-dimensional problems [10], since their computational complexity increases polynomially, but not exponentially with the dimensionality [2]. A general problem in neural networks and machine learning is the accurate evaluation of multidimensional integrals. In 2011 Shaowei Lin in his works [4], [5] consider the problem of evaluating multidimensional integrals in Bayesian statistics which are used

Stoyan Apostolov is supported by the Bulgarian National Science Fund under Young Scientists Project KP-06-M32/2 - 17.12.2019 "Advanced Stochastic and Deterministic Approaches for Large-Scale Problems of Computational Mathematics". Venelin Todorov is supported by the National Scientific Program "Information and Communication Technologies for a Single Digital Market in Science, Education and Security (ICT in SES)", contract No DOI-205/23.11.2018, financed by the Ministry of Education and Science in Bulgaria and by the National Scientific Program for young scientists and post doctoral scientists of Ministry of Education and Science 2020-2021. The work is also supported by the Bulgarian National Science Fund under Project DN 12/5-2017 "Efficient Stochastic Methods and Algorithms for Large-Scale Problems".

in neural network and machine learning. We will primarily be interested in two kinds of integrals. The first has the form

$$\int_{\Omega} p_1^{u_1}(x) \dots p_s^{u_s}(x) dx, \quad (1)$$

where  $\Omega \in \mathcal{R}^s$ ,  $x = (x_1, \dots, x_s)$ ,  $p_i(x)$  are polynomials and  $u_i$  are integers. The second kind of integrals has the form

$$\int_{\Omega} e^{-Nf(x)} \phi(x) dx, \quad (2)$$

where  $f(x)$  and  $\phi(x)$  are  $s$ -dimensional polynomials and  $N$  is a natural number. The asymptotics of such integrals is well understood for models in machine learning, but little was known for singular models until a breakthrough in 2001 [9].

## II. QMC METHODS BASED ON LATTICE RULES

**Lattice point sets** are a special type of low-discrepancy stochastic sequences based on the use of deterministic sequences instead of random sequences [8]. The monographs of Sloan and Kachoyan [7] and Hua and Wang [3] provide comprehensive expositions on the theory of integration lattices.

In our study we will use the following a particular rank-1 lattice sequence [8]:

$$\mathbf{x}_k = \left\{ \frac{k}{N} \mathbf{z} \right\}, \quad k = 1, \dots, N, \quad (3)$$

where  $N$  is an integer,  $N \geq 2$ ,  $\mathbf{z} = (z_1, z_2, \dots, z_s)$  is an integer vector modulo  $N$  of dimensionality  $s$  called a generator of the set and  $\{z\}$  denotes the fractional part of  $z$ . We denote by  $P_N = \{x_1, x_2, \dots, x_N\}$ ,  $x_i \in [0, 1)^s$  the integration nodes of the formula.

**Definition 1:** [7] We say that  $f(x)$  belongs to the class of functions  $E_s^\alpha(c)$  for  $\alpha > 1$  and  $c > 0$ , if  $f$  is a periodic

function with period 1 for every of its components  $x_i, i = 1, 2, \dots, s$ , defined over the unit cube  $[0, 1]^s$  and its Fourier coefficients satisfy the following inequalities:

$$|a(m)| < \frac{c}{(\bar{m}_1 \dots \bar{m}_s)^\alpha}, \quad (4)$$

where

$$\bar{m} = \begin{cases} |m|, & |m| \neq 0, \\ 0, & m = 0, \end{cases}$$

and the constant  $c$  does not depend on  $m_1, \dots, m_s$ .

The discrepancy and the "worst case" error are two important characteristics for the quality of the lattice sequences.

*Definition 2:* Consider the point set  $X = \{x_i \mid i = 1, 2, \dots, N\}$  in  $[0, 1]^s$  and  $N > 1$ . Denote by  $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(s)})$  and  $J(v) = [0, v_1] \times [0, v_2] \times \dots \times [0, v_s]$ . Then the discrepancy of the set is defined as

$$D(N) := \sup_{0 \leq v_j \leq 1} \left| \frac{\#\{x_i \in J(v)\}}{N} - \prod_{j=1}^s v_j \right|.$$

*Definition 3:* For  $f \in E_\alpha^s(c)$  the worst case error is defined as [8]

$$P_\alpha(z, N) = \sum_{z, a \equiv 0 \pmod{N}, a \neq 0} \frac{c}{(\bar{m}_1 \dots \bar{m}_s)^\alpha}.$$

The quantity  $P_\alpha(N, z)$  and the discrepancy are similar measures of the quality of the lattice point set.

In 1959 Bahvalov proved that [1] there exists an optimal choice of the generating vector  $\mathbf{z}$ , for which the error of integration satisfies

$$\left| \frac{1}{N} \sum_{k=0}^{N-1} f\left(\left\{\frac{k}{N}\mathbf{z}\right\}\right) - \int_{[0,1]^s} f(u) du \right| \leq cd \frac{(\log N)^{\beta(s, \alpha)}}{N^\alpha}, \quad (5)$$

for the function  $f \in E_s^\alpha(c)$ , where  $\alpha > 1$  and  $d(s, \alpha), \beta(s, \alpha)$  do not depend on  $N$ .

Moreover, if  $N$  is a prime number, then  $\beta(s, \alpha) = \alpha(s-1)$ . The generating vector  $\mathbf{z}$ , for which inequality (5) is satisfied, is an optimal generating vector and the point set  $P_N$  is a set of good integration points and the numerical integration method is called Good Lattice Point method (GLP). While the theoretical result establish the existence of optimal generating vectors, the difficulty of the development of GLPs is in the construction of the optimal vectors and this is especially difficult with increasing the dimensionality of the integral and dramatically increases the computational complexity.

The first generating vector that we are going to use is based on the generalized Fibonacci numbers of the corresponding dimension. Let  $F_n^{(s)}$  is the  $n$ -th term of the corresponding generalized Fibonacci sequence [8] of dimensionality  $s$ . It's a sum of previous  $s$  terms from this sequence:

$$F_n^{(s)} = \sum_{i=n-s}^{n-1} F_i^{(s)}, \quad \text{where } n \text{ is an integer and } n \geq s \quad (6)$$

and the following initial conditions hold:

$$F_0^{(s)} = F_1^{(s)} = \dots = F_{s-2}^{(s)} = 0, \quad F_{s-1}^{(s)} = 1. \quad (7)$$

Consider the following generating vector [8]:

$$\mathbf{z} = (1, F_n^{(s)}(2), \dots, F_n^{(s)}(s)), \quad (8)$$

where we use that

$$F_n^{(s)}(j) := F_{n+j-1}^{(s)} - \sum_{i=0}^{j-2} F_{n+i}^{(s)} \quad (9)$$

and  $F_{n+l}^{(s)}$  ( $l = 0, \dots, j-1, j$  is an integer,  $2 \leq j \leq s$ ) is the corresponding term of the generalized Fibonacci sequence of dimensionality  $s$ .

The generating vector (8) is transformed into [3], [8]:

$$\mathbf{z} = (1, F_{n-1}^{(s)} + F_{n-2}^{(s)} + \dots + F_{n-s+1}^{(s)}, \dots, F_{n-1}^{(s)} + F_{n-2}^{(s)}, F_{n-1}^{(s)}). \quad (10)$$

Hua and Wang in 1981 [3] proved that the lattice point set with  $N = F_n^{(s)}$  points obtained by using generating vector based on generalized Fibonacci numbers of corresponding dimensionality

$$\left( \left\{ \frac{1}{F_n^{(s)}} k \right\}, \left\{ \frac{F_n^{(s)}(2)}{F_n^{(s)}} k \right\}, \dots, \left\{ \frac{F_n^{(s)}(s)}{F_n^{(s)}} k \right\} \right), \quad (11)$$

$1 \leq k \leq F_n^{(s)}$ , has discrepancy

$$D_{F_n^{(s)}}^* = \mathcal{O} \left( F_n^{(s)^{-\frac{1}{2} - \frac{1}{2s+1} \cdot \frac{1}{\ln 2} - \frac{1}{2 \cdot 2^{s+3}}} \right)$$

and the worst case error is

$$P_\alpha(\mathbf{z}, F_n^{(s)}) = \mathcal{O} \left( (F_n^{(s)})^{-\frac{\alpha}{2} - \frac{\alpha}{2s+1} \cdot \frac{1}{\log 2} - \frac{\alpha}{2 \cdot 2^{s+4}}} \right).$$

If we change the generating vector to be optimal in the way described in [6] we have improved the lattice sequence. The optimal generating vector that we are going to use is constructed recently by Dirk Nuyens [6]. This is a 600-dimensional base-2 generating vector of prime numbers for up to  $2^{20} = 1048576$  points. The method is improved by generating the points from a lattice sequence in base 2 in gray coded radical inverse ordering. This generating vector is generated by the fast component-by-component algorithms, developed in his PhD thesis. The special choice of this optimal generating vector is better than the generating vector from generalized Fibonacci numbers for higher dimensions, which is only optimal for the two dimensional case [8].

### III. NUMERICAL EXAMPLES

We considered three different examples of 4,7,10 and 30 dimensional integrals, respectively, for which we have computed their referent values.

Example 1.  $s = 4$ .

$$\int_{[0,1]^4} x_1 x_2^2 e^{x_1 x_2} \sin(x_3) \cos(x_4) \approx 0.108975. \quad (12)$$

Example 2.  $s = 7$ .

$$\int_{[0,1]^7} e^{1-\sum_{i=1}^3 \sin(\frac{\pi}{2} \cdot x_i)} \cdot \arcsin \left( \sin(1) + \frac{\sum_{j=1}^7 x_j}{200} \right) \approx 0.7515. \quad (13)$$

Example 3.  $s = 10$ .

$$\int_{[0,1]^{10}} \frac{4x_1x_3^2e^{2x_1x_3}}{(1+x_2+x_4)^2} e^{x_5+\dots+x_{10}} \approx 14.808435. \quad (14)$$

Example 4.  $s = 30$ .

$$\int_{[0,1]^{30}} \frac{4x_1x_3^2e^{2x_1x_3}}{(1+x_2+x_4)^2} e^{x_5+\dots+x_{20}} x_{21} \dots x_{30} \approx 3.244. \quad (15)$$

The results are given in the tables below. We make a comparison between plain Monte Carlo (CRUDE) optimized lattice sequence with an optimal generating vector (OPT), Fibonacci lattice sets (FIBO), Sobol sequence (SOBOL) and Matousek scrambling for Sobol sequence (SCR). Each Table contains information about the stochastic approach which is applied, the obtained relative error (RE), the needed CPU-time in seconds and the number of points. Note that when the FIBO method is tested, the number of sampled points are always Generalized Fibonacci numbers of the corresponding dimensionality.

Table I  
ALGORITHM COMPARISON OF THE RE FOR THE 4-DIMENSIONAL INTEGRAL.

# of points	OPT	t,s	FIBO	t,s	SOBOL	t,s	SCR	t,s
1490	6.11e-4	0.002	1.01e-3	0.004	9.46e-4	0.43	3.78e-3	0.47
10671	2.13e-5	0.01	8.59e-5	0.02	5.28e-4	1.4	6.10e-4	1.59
20569	6.56e-6	0.02	3.89e-5	0.03	3.52e-5	4.32	1.97e-5	4.54
39648	9.14e-7	0.06	3.01e-5	0.07	2.68e-5	7.77	9.67e-6	8.26
147312	4.78e-7	0.15	3.71e-6	0.24	2.29e-6	23.7	1.40e-6	27.91

Table II  
ALGORITHM COMPARISON OF THE RE FOR THE 4-DIMENSIONAL INTEGRAL

t, s	OPT	FIBO	SOBOL	SCR
1	5.66e-7	5.62e-6	7.54e-4	6.32e-4
5	3.12e-7	5.38e-7	3.26e-5	1.23e-5
10	5.14e-8	3.77e-7	1.50e-5	8.48e-6
20	3.18e-8	2.67e-8	3.55e-6	1.16e-6

Numerical results show essential advantage for the optimized lattice sets algorithm based on an optimal generating vector in comparison with Fibonacci generalized numbers and Sobol scramble sequence (1-2 orders). For lower dimensions FIBO and Sobol gives results of the same order-see Table I,II. For higher dimensions Scramble sequence SCR is better than FIBO and Sobol by at least 1 order. The results for relative

Table III  
ALGORITHM COMPARISON OF THE RE FOR THE 7-DIMENSIONAL INTEGRAL.

# of points	OPT	t,s	FIBO	t,s	SOBOL	t,s	SCR	t,s
2000	6.39e-4	0.14	2.81e-3	0.23	5.45e-3	1.04	2.51e-3	1.42
7936	3.23e-4	0.64	1.38e-3	0.87	1.28e-3	2.08	1.16e-3	3.08
15808	1.23e-5	0.95	9.19e-4	1.73	9.65e-4	3.26	7.58e-4	5.89
62725	3.15e-6	2.54	2.78e-5	3.41	5.18e-4	12.3	3.11e-4	15.64
124946	1.12e-6	6.48	6.87e-5	6.90	1.09e-4	25.4	8.22e-5	31.41

Table IV  
ALGORITHM COMPARISON OF THE RE FOR THE 7-DIMENSIONAL INTEGRAL

t, s	OPT	FIBO	SOBOL	SCR
0.1	7.38e-4	2.38e-3	8.85e-3	8.37e-3
1	1.17e-5	6.19e-4	5.85e-3	1.37e-3
5	2.32e-6	8.81e-5	1.79e-3	8.38e-4
10	9.11e-7	1.88e-5	7.36e-4	4.78e-4
20	7.43e-7	3.87e-6	1.96e-4	9.87e-5

Table V  
ALGORITHM COMPARISON OF THE RE FOR THE 10-DIMENSIONAL INTEGRAL.

# of points	OPT	t,s	FIBO	t,s	SOBOL	t,s	SCR	t,s
1597	3.14e-4	0.002	4.39e-3	0.003	6.31e-3	0.02	1.46e-3	0.05
17711	6.21e-5	0.02	1.81e-3	0.04	5.31e-4	0.11	1.83e-4	0.21
121393	4.34e-6	0.15	1.20e-3	0.16	1.78e-4	1.21	3.12e-5	1.47
832040	4.11e-7	0.75	1.19e-5	0.70	3.24e-5	12.1	8.25e-6	14.41
3524578	5.32e-8	6.35	2.63e-6	6.45	4.57e-6	121.5	7.71e-7	139.1

Table VI  
ALGORITHM COMPARISON OF THE RE FOR THE 10-DIMENSIONAL INTEGRAL

t, s	OPT	FIBO	SOBOL	SCR
0.1	4.95e-6	9.19e-6	5.31e-4	4.19e-4
1	8.10e-7	5.63e-6	1.81e-4	1.21e-4
5	3.56e-8	2.15e-6	8.07e-5	7.21e-5
10	4.31e-8	1.79e-6	4.77e-5	3.51e-5
20	9.13e-9	8.61e-7	8.42e-6	7.09e-6

Table VII  
ALGORITHM COMPARISON OF THE RE FOR THE 30-DIMENSIONAL INTEGRAL.

# of points	OPT	t,s	SCR	t,s	SOBOL	t,s	FIBO	t,s
1024	1.21e-2	0.02	5.78e-2	0.53	1.18e-1	0.42	8.81e-1	0.02
16384	4.11e-3	0.16	1.53e-2	5.69	8.40e-2	4.5	6.19e-1	0.14
131072	5.24e-4	1.34	1.35e-3	42.1	1.18e-2	30.2	2.78e-1	1.16
1048576	8.81e-5	9.02	6.78e-4	243.9	9.20e-3	168	9.86e-2	8.61

errors corresponding to FIBO and Sobol are similar especially for higher sample number, see Tables III. If the computational time is fixed the advantage of Fibonacci lattice sets in terms of relative error in comparison with Sobol approach is clearly

Table VIII  
ALGORITHM COMPARISON OF THE RE FOR THE 30-DIMENSIONAL  
INTEGRAL

time,sec	OPT	SCR	SOBOL	FIBO
1	3.48e-3	2.38e-2	1.01e-1	2.38e-1
5	4.23e-4	5.46e-3	7.76e-2	1.81e-1
10	8.91e-5	1.25e-3	5.71e-2	9.48e-2
20	2.33e-5	6.11e-4	1.28e-2	7.87e-2

seen, see Tables IV. In general scrambling procedure improves the relative error of the unscrambled nets as it is the case for Sobol sequence and its scrambled version by Matousek linear scrambling as can be seen from the results in Tables V,VI. For very high dimensions the optimized lattice rule outperforms not only the scramble sequence, but also the FIBO method and Sobol sequence by at least 2 orders - see Table VII,VIII. The experiments show that the optimized lattice sequence with a special choice of the optimal generating vector is the best method in terms of lower relative errors with increasing the dimensionality of the integral. The optimized lattice sequence gives the best results compared to the other stochastic approaches also for a fixed computational times which show that the presented algorithm is the most computationally efficient.

#### IV. CONCLUSION

In this paper an optimized lattice rule has been presented and tested on multidimensional integrals used in machine learning. A comprehensive experimental study of optimized lattice rule, Fibonacci lattice sets, Sobol sequence and Matousek scrambling for Sobol sequence has been done on some case test functions. This approaches are the only possible algorithms for high dimensional integrals because the deterministic

algorithms need an huge amount of time for the evaluation of the integral. The numerical tests show that the optimized lattice rule is the most efficient for multidimensional integration and especially for computing high dimensional integrals. It is an important element since this may be crucial in order to achieve a more reliable interpretation of the results in Bayesian statistics which is foundational in artificial intelligence and machine learning.

#### REFERENCES

- [1] N. Bahvalov (1959) On the approximate calculation of multiple integrals, *Journal of Complexity*, Volume 31, Issue 4, 2015, Pages 502-516, ISSN 0885-064X, <https://doi.org/10.1016/j.jco.2014.12.003>.
- [2] Dimov I., *Monte Carlo Methods for Applied Scientists*, New Jersey, London, Singapore, World Scientific, 2008, 291p.
- [3] Hua, L.K. and Wang, Y., *Applications of Number Theory to Numerical analysis*, 1981.
- [4] Lin S., "Algebraic Methods for Evaluating Integrals in Bayesian Statistics," Ph.D. dissertation, UC Berkeley, May 2011.
- [5] Lin, S., Sturmfels B., Xu Z.: Marginal Likelihood Integrals for Mixtures of Independence Models, *Journal of Machine Learning Research*, Vol. 10, pp. 1611-1631, 2009, <https://doi.org/10.5555/1577069.1755838>.
- [6] Kuo, F.Y., Nuyens, D. Application of Quasi-Monte Carlo Methods to Elliptic PDEs with Random Diffusion Coefficients: A Survey of Analysis and Implementation. *Found Comput Math* 16, 1631–1696 (2016). <https://doi.org/10.1007/s10208-016-9329-5>.
- [7] Sloan I.H. and Kachoyan P.J., Lattice methods for multiple integration: Theory, error analysis and examples, *SIAM J. Numer. Anal.* 24, pp. 116–128, 1987, <https://doi.org/10.1137/0724010>.
- [8] Wang Y., Hickernell F.J. (2002) An Historical Overview of Lattice Point Sets. In: Fang KT., Niederreiter H., Hickernell F.J. (eds) *Monte Carlo and Quasi-Monte Carlo Methods 2000*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-56046-0\\_10](https://doi.org/10.1007/978-3-642-56046-0_10).
- [9] Watanabe S., Algebraic analysis for nonidentifiable learning machines. *NeuralComput.*(13), pp. 899–933, April 2001, <https://doi.org/10.1162/089976601300014402>.
- [10] Zheleva, I., Georgiev, I., Filipova, M., & Menseidov, D. (2017, October). Mathematical modeling of the heat transfer during pyrolysis process used for end-of-life tires treatment. In *AIP Conference Proceedings* (Vol. 1895, No. 1, p. 030008). AIP Publishing LLC, <https://doi.org/10.1063/1.5007367>.