JOURNAL OF COMPUTATIONAL PHYSICS 33, 249–258 (1979)

An Improved Low-Discrepancy Sequence for Multidimensional Quasi–Monte Carlo Integration*

ERIC BRAATEN AND GEORGE WELLER

Department of Physics, University of Wisconsin, Madison, Wisconsin 53706

Received November 7, 1978; revised February 13, 1979

We present an improved method of generating vectors for Monte Carlo integration, which produces a significant improvement in rate of convergence over previous methods for problems in more than eight dimensions.

1. INTRODUCTION

Scientific calculations frequently require numerical integration of functions of many variables. Iterated Gaussian quadrature is very effective in spaces of less than about five dimensions. But for more than five, this method requires too many function evaluations to be practical.

Some applications require the evaluation of integrals in 10 or even 20 dimensions. In such cases, Monte Carlo (MC) and quasi–Monte Carlo (QMC) methods are more effective. In the MC method, an approximation to the integral is calculated from the values of the function at a sequence of "random" points. In the QMC method, a uniformly distributed sequence of points is used.

2. DRAWBACKS OF THE MONTE CARLO METHOD

Two problems are associated with the MC method. First, its asymptotic convergence rate is considerably slower than that of the QMC method. Second, the pseudo-random-number generators commonly used in the MC method have poor distribution properties in spaces of many dimensions.

Many integration problems can be reduced to integrating a function f over the k-dimensional unit hypercube:

$$I = \int_0^1 dx_1 \cdots \int_0^1 dx_k f(x_1, \ldots, x_k).$$

In the MC and QMC methods, a sequence $\mathbf{a}_1, \ldots, \mathbf{a}_N$ of vectors in the unit hypercube

0021-9991/79/110249–10$02.00/0

is generated and the integral is approximated by the average value of the function at those points:

$$I_N = \sum_{n=1}^{N} f(\mathbf{a}_n)/N.$$

If f is square-integrable, the Law of the Iterated Logarithm [1] gives the best possible asymptotic upper bound on the error of integration for a random sequence:

$$| I_N - I | < K(\log(\log N)/N)^{1/2}, \tag{1}$$

where K is a constant which depends on the function f. So for the values of N encountered in applications, the MC method gives an upper bound on the error which decreases essentially like $1/N^{1/2}$. On the other hand, the QMC method uses a "quasi-random" sequence which is designed to be uniformly distributed in the hypercube. There are quasi-random sequences which give upper bounds on $| I_N - I |$ that decrease like $(\log N)^k/N$, and hence have much better asymptotic behavior than the MC method.

In the MC method, the pseudo-random vectors are often generated by taking k successive integers b_j from a linear congruential sequence of the form

$$b_{j+1} = (qb_j + r) \qquad \text{mod } M.$$

These k integers are then normalized by dividing by M to form the components of $\mathbf{a}_n$. Thus $\mathbf{a}_n$ must fall on the lattice of points whose coordinates are multiples of $1/M$. Marsaglia [2, 3] has obtained two results which raise questions about the suitability of these pseudo-random sequences for numerical integration in more than one dimension:

(1) All vectors generated by this method will fall in fewer than $(k!M)^{1/k}$ parallel hyperplanes, whereas the lattice of points whose coordinates are multiples of $1/M$ cannot be spanned by fewer than M parallel hyperplanes.

2. Any $k + 1$ points enclose a volume which is a multiple of $1/(k!M)$ while $k + 1$ random points on the lattice should enclose volumes as small as $1/(k!M^k)$.

As an example, consider the linear congruential generator RANUN used by the Madison Academic Computing Center, for which $q = 5^{15}$, $r = 1$, and $M = 2^{35}$. Vectors in 8-space generated by RANUN will fall in fewer than 79 parallel hyperplanes, as compared with the more than 3.4×10^{10} expected of a random sequence on the lattice. Thus the vectors generated by this method are far from randomly distributed. This difficulty is avoided in the QMC method, which aims for a uniformly distributed sequence rather than a random one.

For a critique of Marsaglia's results and a recent survey of quasi–Monte Carlo methods and pseudo-random numbers, see Ref. [9].

3. Discrepancy as a Measure of Nonuniformity

The root-mean-square discrepancy [4] T_N, which we shall simply call the "discrepancy," has been found to be a useful measure of the nonuniformity of the distribution of a sequence $\mathbf{a}_1, \ldots, \mathbf{a}_N$ in the unit hypercube. The discrepancy is the root-mean-square value of the local discrepancy function $g(\mathbf{x})$ defined by

$$g(\mathbf{x}) = \sum_{n=1}^{N} \left(\prod_{i=1}^{k} \theta(x_i - a_{ni}) \right) \Big/ N - \prod_{i=1}^{k} x_i , \tag{2}$$

where a_{ni} is the ith component of $\mathbf{a}_n$ and θ is the step function: $\theta(t) = 0$ for $t < 0$, $\theta(t) = 1$ for $t \geqslant 0$. g is essentially the deviation of the cumulative distribution of the sequence $\mathbf{a}_1, \ldots, \mathbf{a}_N$ from the cumulative uniform distribution.

The discrepancy is useful because it gives a strict upper bound on the error integration [5]:

$$| I_N - I | < \sum_{(\cdots)} K(\cdots)\, T_N(\cdots),$$

where the sum is over the $2^k - 1$ nonempty subsequences $(i_1, \ldots, i_j)$ of $(1, 2, \ldots, k)$, $K(i_1, \ldots, i_j)$ is a constant which depends on the function f, and $T_N(i_1, \ldots, i_j)$ is the discrepancy of the j-dimensional sequence consisting of the i_1th,..., and i_jth components of the sequence $\mathbf{a}_n$. This upper bound holds only for functions which are "of bounded variation in the sense of Hardy and Krause" [6] (BVHK). It is a rather restricted class of functions, but it does include all functions f which satisfy both:

(1) $| \partial/\partial x_1 \cdots (\partial/\partial x_k) f |$ is integrable.

(2) The restriction of f to each boundary hypersurface of the hypercube is also BVHK.

4. The Halton Sequence

Warnock [7] has made a comparative study of various low-discrepancy sequences, calculating their discrepancies out to $N = 1000$. (One thousand points is too small for many numerical integration problems, but calculating discrepancies beyond $N = 1000$ is rather expensive.) Warnock found that the sequences with the lowest discrepancies in eight or fewer dimensions are the radical inverse sequences, the prototype of which is the Halton sequence.

The uniformity of the distribution of the Halton sequence is based on the properties of prime numbers. If p is a prime, any integer n can be written in p-ary notation as follows:

$$n = e_j p^j + \cdots + e_1 p + e_0 \qquad \text{where} \quad 0 \leqslant e_i \leqslant p - 1.$$

So n can be represented by the integer string $e_j \cdots e_1 e_0$ (base p). The radical inverse of n to the base p, $R_p(n)$, is then obtained by reflecting through the radical point:

$$R_p(n) = 0.e_0 e_1 \cdots e_j \text{ (base } p) = e_0/p + e_1/p^2 + \cdots + e_j/p^{j+1}. \tag{3}$$

This gives a very uniformly distributed sequence in the interval (0, 1) for each prime p. The Halton sequence in k dimensions consists of a distinct radical inverse sequence for each coordinate:

$$\mathbf{a}_n = (R_2(n), R_3(n),..., R_{p_k}(n)),$$

where p_k is the kth prime number.

Halton [8] has found an upper bound on the discrepancy of his sequence:

$$T_N < C_k(\log N)^k/N, \tag{4}$$

where C_k is a constant that depends on the dimension. This bound, together with Eq. (2), gives a strict upper bound on the error of integration using the Halton sequence:

$$| I_N - I | < C(\log N)^k/N.$$

This upper bound guarantees a much faster asymptotic convergence rate than that of the MC method, which is given by Eq. (1).

5. Scrambling the Halton Sequence

Halton's upper bound on T guarantees that, asymptotically, his sequence will give a much lower integration error than a random sequence. However, numerical calculations show that in eight dimensions, the discrepancy of the Halton sequence does not fall permanently below the root-mean-square discrepancy of a random sequence until $N \simeq 700$. The situation becomes progressively worse as k increases.

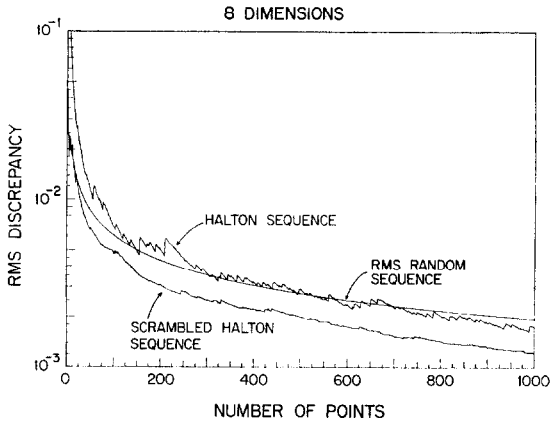

Fig. 1. Discrepancies T_N in $k = 8$ dimensions for Halton sequence and Scrambled Halton sequence out to $N = 1000$ points. The smooth curve is the root-mean-square discrepancy for a random sequence: $\langle T_N^2 \rangle = ((1/2^k - 1/3^k)/N)^{1/2}$.

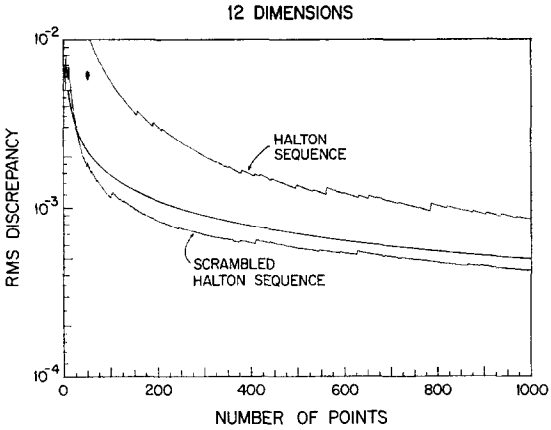

FIG. 2. Discrepancies in 12 dimensions.

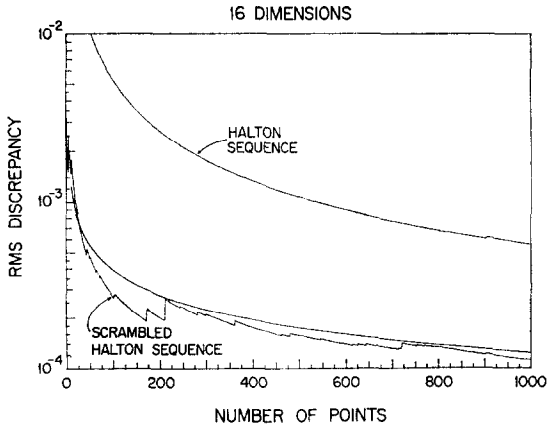

FIG. 3. Discrepancies in 16 dimensions.

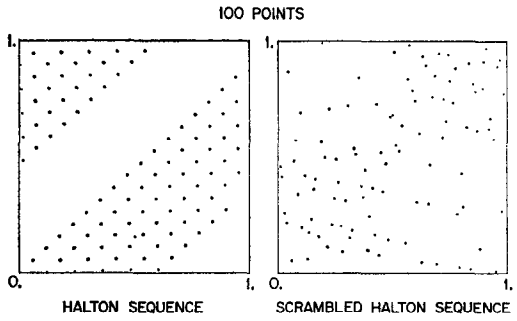

FIG. 4. Projection of the first 100 points of the Halton and Scrambled Halton sequences onto the seventh and eighth coordinate plane.

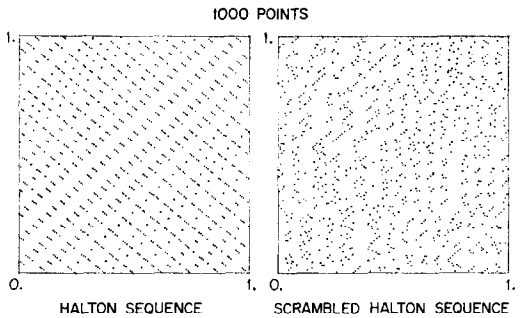

FIG. 5. Projection of the first 1000 points.

TABLE I

Permutations for Scrambled Halton Sequence

Prime	Permutation[a] of (0 1 2 ⋯ $p-1$)
2	(0 1)
3	(0 2 1)
5	(0 3 1 4 2)
7	(0 4 2 6 1 5 3)
11	(0 5 8 2 10 3 6 1 9 7 4)
13	(0 6 10 2 8 4 12 1 9 5 11 3 7)
17	(0 8 13 3 11 5 16 1 10 7 14 4 12 2 15 6 9)
19	(0 9 14 3 17 6 11 1 15 7 12 4 18 8 2 16 10 5 13)
23	(0 11 17 4 20 7 13 2 22 9 15 5 18 1 14 10 21 6 16 3 19 8 12)
29	(0 15 7 24 11 20 2 27 9 18 4 22 13 26 5 16 10 23 1 19 28 6 14 17 3 25 12 8 21)
31	(0 15 23 5 27 9 18 2 29 12 20 7 25 11 17 3 30 14 22 1 21 8 26 10 16 28 4 19 6 24 13)
37	(0 18 28 6 23 11 34 3 25 14 31 8 20 36 1 16 27 10 22 13 32 4 29 17 7 35 19 2 26 12 30 9 24 15 33 5 21)
41	(0 20 31 7 26 12 38 3 23 34 14 17 29 5 40 10 24 1 35 18 28 9 33 15 21 4 37 13 30 8 39 22 2 27 16 32 11 25 6 36 19)
43	(0 21 32 7 38 13 25 3 35 17 28 10 41 5 23 30 15 37 1 19 33 11 26 42 8 18 29 4 39 14 22 34 6 24 12 40 2 31 20 27 9 36 16)
47	(0 24 12 39 6 33 20 44 3 29 16 36 10 42 22 8 31 26 14 46 1 35 18 28 5 40 19 37 11 25 43 4 30 15 34 9 45 21 2 32 17 41 13 27 7 38 23)
53	(0 26 40 9 33 16 49 4 36 21 45 12 29 6 51 23 38 14 43 1 30 19 47 10 34 24 42 3 27 52 15 18 39 7 46 31 11 35 20 48 2 28 41 8 22 50 13 32 17 44 5 37 25)

[a] For each permutation, we can generate another permutation with the same minimal discrepancy by replacing each nonzero integer n by $p - n$. For example, for $p = 11$, the other permutation is (0 6 3 9 1 8 5 10 24 7).

See Figs. 1, 2, and 3. The reason for this behavior is that the sequence $R_p(n)$ defined by Eq. (3) consists of cycles of length p of monotonically increasing numbers. For example, in eight dimensions, the last two coordinates are given by $R_{117}(n)$ and $R_{19}(n)$, and consist of 17 and 19 increasing numbers, respectively. This produces a strong correlation between the seventh and eight coordinates of the sequence. Figures 4 and 5 show the projection of the Halton sequence onto the seventh and eight dimensions for $N = 100$ and $N = 1000$.

In an attempt to improve the behavior of the Halton sequence in many dimensions we looked for a way of scrambling the cycles of length p without sacrificing the low discrepancy. We have proven that Halton's upper bound Eq. (4) continues to hold for the family of sequences described below. For each prime p, let π_p be a permutation on the digits $(0, 1,..., p-1)$ which holds 0 fixed. We define the scrambled radical inverse sequence $S_p(n)$ in analogy with Eq. (3):

$$S_p(n) = \pi_p(e_0)/p + \pi_p(e_1)/p^2 + \cdots + \pi_p(e_j)/p^{j+1}$$

if n has the expansion $n = e_j \cdots e_1 e_0$ (base p). Our Scrambled Halton sequence is then given by

$$\mathbf{a}_n = (S_2(n), S_3(n),..., S_{p_k}(n)).$$

Note that the Halton sequence is just the special case in which each π_p is chosen to be the identity permutation.

We have not been able to find optimal choices for the permutations $\pi_3, \pi_5, \ldots$. A choice that we have found to be very effective is the following: for each prime p, we picked a permutation π_p which minimized the 1-dimensional discrepancy after each of the first $p-1$ steps of the sequence $S_p(n)$. That is, having picked $\pi_p(1),...,$ and $\pi_p(j)$, we chose $\pi_p(j+1)$ so as to minimize the discrepancy of the set $(\pi_p(1)/p,..., \pi_p(j)/p, \pi_p(j+1)/p$. This procedure does not specify a unique permutation. For example, $\pi_p(1)$, for p odd, can be chosen to be either $\frac{1}{2}(p-1)$ or $\frac{1}{2}(p+1)$. We made our calculations with the permutations listed in Table I, but the calculations are insensitive to the choice of permutation. Figures 4 and 5 show how our permutations break up the correlations between the seventh and eight coordinates of the Halton sequence.

6. Tests and Comparisons

In Figs. 1–3, we compare the discrepancy of our Scrambled Halton sequence with that of the original Halton sequence in 8, 12, and 16 dimensions. In 8 dimensions, the discrepancy of the scrambled sequence is 29% lower than that of the Halton sequence at $N = 1000$ points. It is significantly lower than any of the sequences that were examined by Warnock. In 16 dimensions, our sequence has a discrepancy that

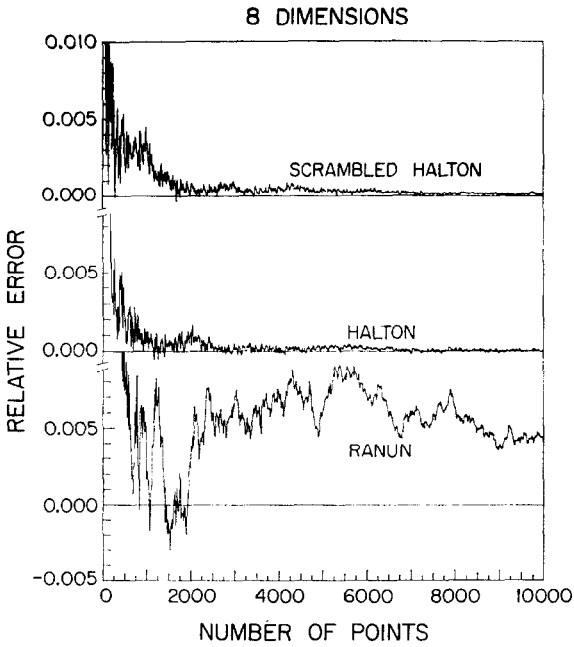

FIG. 6. Relative error $(I_N - I)/I$ in eight dimensions for Scrambled Halton sequence, Halton sequence, and RANUN out to $N = 10\,000$ points. The integrand is a Gaussian peaked at the center of the hypercube: $f(\mathbf{x}) = \prod_{i=1}^{8} \exp(-2(x_i - 0.5)^2)$.

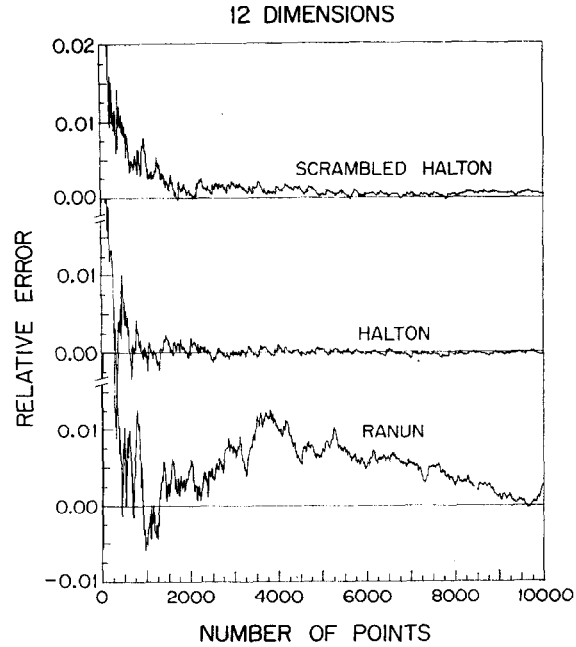

FIG. 7. Relative error in 12 dimensions.

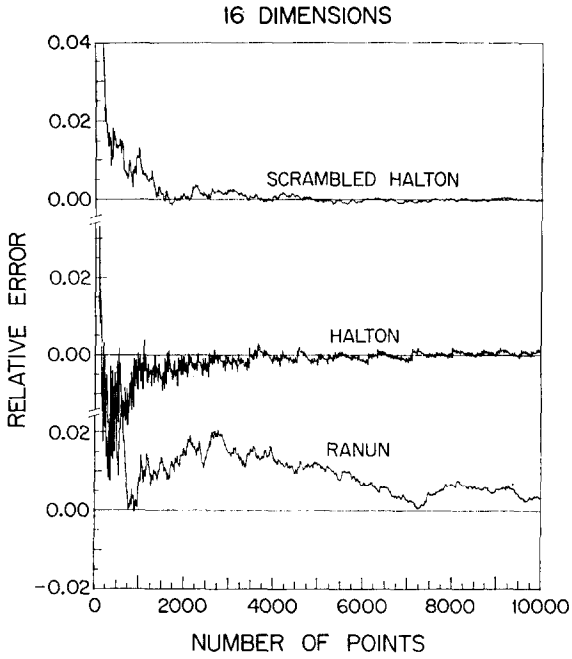

FIG. 8. Relative error in 16 dimensions.

is 80% lower than Halton's at $N = 1000$, and it always stays below the root-mean-square discrepancy of a random sequence.

Test integrals are a very subjective way of comparing integration methods, but they can be useful for showing convergence trends. In Figs. 6–8, we compare the error of integration for the Scrambled Halton sequence, the Halton sequence, and the linear congruential generator RANUN described in Section 2, on a simple test integral in 8, 12, and 16 dimensions. The slow convergence of the MC method is obvious in every dimension. The improvement that results from scrambling the Halton sequence is evident in 16 dimensions, where the Halton sequence produces much larger fluctuations in the error of integration.

7. CONCLUSION

We have shown that considerable improvement in the convergence rate of the quasi–Monte Carlo method can be obtained by reordering the Halton sequence. Discrepancy calculations indicate that our sequence shows drastic improvement over the Halton sequence in spaces of high dimension, remaining below the expected discrepancy of a random sequence. This was accomplished without sacrificing Halton's upper bound on the discrepancy, which guarantees a much faster asymptotic convergence rate than one can get with a Monte Carlo calculation.

ACKNOWLEDGMENTS

One of us (GW) gratefully acknowledges helpful discussions with John H. Halton at an early stage of this work. We also thank Vernon Barger for attempting the calculations that made this work necessary.

This research was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, and in part by the Department of Energy under contract EY-76-C-02-0881, COO-881-55.

REFERENCES

1. L. BREIMAN, "Probability," p. 64, Addison-Wesley, Reading, Mass. 1968.
2. G. MARSAGLIA, *Proc. Nat. Acad. Sci. U.S.A.* **61** (1968), 25.
3. G. MARSAGLIA, *Numer. Math.* **16** (1968), 8.
4. S. K. ZAREMBA, *SIAM Rev.* **10** (1968), 303.
5. J. HALTON, Estimating the Accuracy of Quasi–Monte Carlo Integration, *in* "Application of Number Theory to Numerical Analysis" (S. K. Zaremba, Ed.), Academic Press, New York, 1972.
6. S. HABER, *SIAM Rev.* **12** (1970), 481.
7. T. WARNOCK, Computational Investigations of Low-discrepancy Point Sets, *in* "Applications of Number Theory to Numerical Analysis" (S. K. Zaremba, Ed.), Academic Press, New York, 1972
8. J. HALTON, *Numer. Math.* **2** (1960), 84. Note erratum in *Numer. Math.* **2** (1960), 196.
9. H. NIEDERREITER, *Bull. Amer. Math. Soc.* **84** (1978), 957.