JOURNAL OF COMPUTATIONAL PHYSICS 122, 218–230 (1995)

Quasi-Monte Carlo Integration

WILLIAM J. MOROKOFF AND RUSSEL E. CAFLISCH*

Mathematics Department, University of California, Los Angeles, California 90095-1555

Received August 6, 1993; revised May 2, 1994

The standard Monte Carlo approach to evaluating multidimensional integrals using (pseudo)-random integration nodes is frequently used when quadrature methods are too difficult or expensive to implement. As an alternative to the random methods, it has been suggested that lower error and improved convergence may be obtained by replacing the pseudo-random sequences with more uniformly distributed sequences known as quasi-random. In this paper quasi-random (Halton, Sobol', and Faure) and pseudo-random sequences are compared in computational experiments designed to determine the effects on convergence of certain properties of the integrand, including variance, variation, smoothness, and dimension. The results show that variation, which plays an important role in the theoretical upper bound given by the Koksma–Hlawka inequality, does not affect convergence, while variance, the determining factor in random Monte Carlo, is shown to provide a rough upper bound, but does not accurately predict performance. In general, quasi-Monte Carlo methods are superior to random Monte Carlo, but the advantage may be slight, particularly in high dimensions or for integrands that are not smooth. For discontinuous integrands, we derive a bound which shows that the exponent for algebraic decay of the integration error from quasi-Monte Carlo is only slightly larger than $\frac{1}{2}$ in high dimensions.

1. INTRODUCTION

1.1. *Applications of Quasi-Random Sequences*

Quasi-random sequences have been greatly touted for their theoretical properties, for example in recent articles in *Scientific American*, *Science*, and *SIAM News*. They have seen little real application, however, because the theoretical advantages are often difficult to attain. In fact, previous computational studies [1, 6, 8, 17, 19] showed that quasi-random methods can provide significant improvement, which, however, falls well short of their theoretical potential.

Our paper is an attempt to fill in that gap by careful numerical experiment to clarify the real advantages and weaknesses of quasi-Monte Carlo in computations. The computations presented here clearly show the influence of variation, variance, dimension, and smoothness on the convergence rates of quasi-Monte Carlo integration. We expect that the conclusions of this study will be useful for the application of quasi-Monte Carlo methods to particular problems, as well as for the development of variance reduction and other Monte Carlo techniques using quasi-random sequences.

In Sections 2 and 3 the integrals of several test functions are approximated by using the Halton, Sobol', and Faure sequences. For each example the integration error is approximated by the form $cN^{-\alpha}$, in which N is the number of quasi-random integration points. In general, the results show that quasi-Monte Carlo integration is superior to standard Monte Carlo in exponent α (which is $\frac{1}{2}$ for standard Monte Carlo) or in the constant c, but that the improvement degrades as the smoothness of the integrand decreases or the dimension increases. These conclusions and their implications for further use of quasi-random sequences will be discussed in Section 4.

1.2. *Low Discrepancy Sequences*

Monte Carlo methods use independent, uniformly distributed random numbers on the s-dimensional unit cube I^s as the source of integration nodes. In the simplest case, if $\{\mathbf{z}_i\}$ is such a sequence of random points in I^s, then the integral of the function $f(x_1, \ldots, x_s)$ over I^s is approximated by the average of f evaluated at the points $\mathbf{z}_i$. The error,

$$\varepsilon = \int_{I^s} f(\mathbf{x})\, d\mathbf{x} - \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{z}_i),$$

satisfies the relationship involving the expectation $E(\cdot)$ of a random variable

$$E(\varepsilon^2) = \frac{\sigma^2(f)}{N},$$

where $\sigma^2(f)$ is the variance of f defined by

$$\sigma^2(f) = \int_{I^s} f^2(\mathbf{x})\, d\mathbf{x} - \left(\int_{I^s} f(\mathbf{x})\, d\mathbf{x}\right)^2.$$

This shows the familiar convergence rate of $N^{-1/2}$ associated with random methods. The key property of the random sequence

* Research supported in part by the Air Force Office of Scientific Research under Grant AFOSR 90-0003.

0021-9991/95 $12.00

is its uniformity, so that any contiguous subsequence is well spread throughout the cube. This idea has lead to the suggestion that using other sequences which are more uniformly distributed than a random sequence may produce better results. Such sequences are called quasi-random or low discrepancy sequences.

Initially it may appear that a grid would provide optimal uniformity. However, grids suffer from several difficulties. First, in high dimension, the number of points required to create even a course mesh is exponentially large in dimension. Also, grids have rather high discrepancy, a quantity which measures the uniformity of a set of points. This is defined and discussed below. Finally, the only obvious method for increasing accuracy of a uniform grid is to halve the mesh size, which requires adding 2^s times the current number of points, so that the accuracy of a uniform grid cannot be increased incrementally.

A solution to this problem is to use infinite sequences of points such that for every N, the first N terms of the sequence are uniformly distributed throughout the cube. In order to quantify this, the discrepancy of a set of N points is defined as follows: Let Q be a rectangle contained in I^s with sides parallel to the coordinate axes, and let $m(Q)$ be its volume. The discrepancy D_N of the sequence $\{x_i\}$ of N points is

$$D_N = \sup_{Q \in I^s} \left| \frac{\#\text{ of points in } Q}{N} - m(Q) \right|.$$

By the law of iterated logarithms, the expectation of the discrepancy of a random sequence is bounded by $(\log \log N)N^{-1/2}$. There are many quasi-random sequences known for which the discrepancy is bounded by a constant times $(\log N)^s/N$, which suggests greater uniformity than a random sequence. Three such sequences, the Halton, Sobol', and Faure sequences, were chosen for comparison in this work and are briefly described now. A more thorough discussion of these sequences can be found in [9, 15, 16].

The Halton sequence [4] in one dimension is generated by choosing a prime p and expanding the sequence of integers 0, 1, 2, ..., N into base p notation. The nth term of the sequence is given by

$$z_n = \frac{a_0}{p} + \frac{a_1}{p^2} + \frac{a_2}{p^3} + \cdots + \frac{a_m}{p^{m+1}},$$

where the a_i's are integers taken from the base p expansion of $n - 1$

$$[n-1]_p = a_m a_{m-1} \ldots a_2 a_1 a_0,$$

with $0 \le a_i < p$. For example, if $p = 3$, the first 12 terms of the sequence ($n = 1, \ldots, 12$) are

$$\{0, \tfrac{1}{3}, \tfrac{2}{3}, \tfrac{1}{9}, \tfrac{4}{9}, \tfrac{7}{9}, \tfrac{2}{9}, \tfrac{5}{9}, \tfrac{8}{9}, \tfrac{1}{27}, \tfrac{10}{27}, \tfrac{19}{27}, \ldots\}.$$

Note that the numbers lie in cycles of p increasing terms and that, within the cycle, the terms are separated by $1/p$. The effect is that once a grid of refinement $1/p^m$ is filled in by repeated sweeps of these cycles, the next cycle starts filling in the grid at level $1/p^{m+1}$. The s-dimensional Halton sequence is generated by pairing s one-dimensional sequences based on s different primes; usually the first s primes are chosen. One difficulty with this sequence is that in high dimensions, the base p must be large, so that the cycle of increasing terms is rather long. When paired against another large prime-based sequence, the result is that the points lie on parallel lines which slowly sweep through the unit square. Thus the distribution of points is not very uniform.

In the definition of D_N, the restriction to rectangles with sides parallel to the coordinate axes is due to this construction of an s-dimensional sequences as a product of one-dimensional sequences.

As an alternative, the theory of (t, s)-nets has been developed by Sobol' [20], Faure [3], and Niederreiter [15]. The Sobol' sequence solves the problem of large primes by only using $p = 2$. The sequence is generated such that the first 2^m terms of each dimension for $m = 0, 1, 2, \ldots$ are a permutation of the corresponding terms of the Halton sequence with prime base 2 (also known as the van der Corput sequence). If the proper choice of permutations is used, the resulting s-dimensional sequence can be shown to have good uniformity properties. However, as dimension increases, more permutations must be used, and the possibility increases that a bad pairing may exist between two dimensions leading to a highly nonuniform distribution in that plane.

The Faure sequence is similar to the Sobol' sequence in that each dimension is a permutation of a Halton sequence; however, the prime used for the base is chosen as the smallest prime greater than or equal to the dimension. If this prime is labeled $p(s)$, then each dimension of the Faure sequence is generated such that the first $p(s)^m$ terms for $m = 0, 1, 2, \ldots$ are a permutation of the corresponding terms of the Halton sequence base $p(s)$. The advantage of this is that as long as $s \le p(s)$, an optimal set of permutations can be prescribed. However, as dimension increases, the problems of large primes arise, though not as quickly as with the Halton sequence.

In his review papers [13] and [14], Niederreiter summarizes the properties and theory of these sequences. The key point is that their discrepancy satisfies the relationship

$$D_N \le c_s \frac{(\log N)^s}{N} + \mathcal{O}\left(\frac{(\log N)^{s-1}}{N}\right),$$

where c_s, different for each sequence, is constant in N, but depends on s. This is an optimal upper bound in the sense that for any infinite sequence, there exist an infinite number of N such that

$$D_N \geq c\frac{(\log N)^s}{N}$$

for some constant c. These bounds suggest that, at least for large N, the quasi-random sequences described above will be considerably more uniform than a random sequence. The relationship between discrepancy and integration error is described next.

1.3. *Error Bounds*

As stated above the expectation of error for random Monte Carlo integration is $\sigma(f)N^{-1/2}$. Because probability theory can no longer be used, a different approach is necessary to bound the integration error when quasi-random sequences are used. It turns out that it is the variation of the integrand which appears, and not the variance. If $f(x_1, ..., x_s)$ is sufficiently differentiable, then for all positive $k \leq s$ and all sets of k integers $1 \leq i_1 < i_2 < \cdots < i_k \leq s$, define the quantity

$$V^{(k)}(f; i_1, ..., i_k) = \int_{I^k} \left| \frac{\partial^k f}{\partial t_{i_1} \cdots \partial t_{i_k}} \right|_{t_j = 1, j \neq i_1, ..., i_k} dt_{i_1} \cdots dt_{i_k}, \quad (1)$$

which represents the variation of the projection of k variables. The variation in the sense of Hardy and Krause is then defined as

$$V(f) = \sum_{k=1}^{s} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq s} V^{(k)}(f; i_1, ..., i_k). \quad (2)$$

The restriction of differentiability is stronger than necessary and may be relaxed to the standard notion of bounded variation. The main result on integration error is known as the Koksma–Hlawka inequality, which states that if f is a function of bounded variation in the sense of Hardy and Krause, and $\{\mathbf{x}_i\}$ is a sequence of N points in I^s with discrepancy D_N, then

$$\left| \int_{I^s} f(\mathbf{x})\, d\mathbf{x} - \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_i) \right| \leq V(f)\, D_N. \quad (3)$$

As with the expectation of error in the random case, the effects of the integrand are separated from those of the sequence. The bound on the discrepancy of a random sequence indicates $N^{-1/2}$ type convergence and suggests that a sequence with smaller discrepancy than a random sequence will give smaller errors. The requirement that the integrand be of bounded variation is considerably more restrictive than the condition for the random case of finite variance, and it excludes many integrands of interest. For example the characteristic function of any nonrectangular set has infinite variation. The relationship of such functions to particle simulation is described in Section 3 as motivation for the convergence results presented there.

Another approach to describing integration error comes from considering the average case error for the class of continuous functions equipped with a Wiener sheet measure. In [21] Woźniakowski gives a result which relates this average integration error for such functions to an L_2 version of discrepancy. Further discussion and an alternative derivation of this can be found in [9]. This result suggests that low discrepancy sequences should out perform random sequences, at least for large enough N, even for functions that are only Hölder continuous with exponent $\frac{1}{2}$.

2. VARIATION, VARIANCE, SEQUENCES, AND DIMENSION

From the Koksma–Hlawka inequality (3), one might hope that variation would play the same role for quasi-random integration as variance does for standard Monte Carlo. However, certain observations suggest that variation is a suspect quantity for measuring integration error. For example, from the definition it can be seen that total variation involves a sum of variations over only those lower dimensional boundaries where all the restricted variables are equal to 1 (not 0). This can lead to considerable differences between the variation of two basically identical functions. It might be expected that the functions f and g,

$$f(x_1, ..., x_s) = \prod_{i=1}^{s} (1 - x_i)$$

$$g(x_1, ..., x_s) = \prod_{i=1}^{s} x_i,$$

would be integrated virtually identically by a quasi-random sequence. However, $V(f) = 1$ and $V(g) = 2^s - 1$. While this may be an extreme case, it does introduce some doubt as to the usefulness of variation in predicting integration errors.

2.1. *Setup of Numerical Experiments*

In order to investigate the actual role that variation plays in determining the integration error, a variety of test functions with a wide range of variations were integrated over the unit cube. Effects of variance, dimension, and choice of sequence were also observed. The functions were chosen so that they could be integrated analytically and so that the variance and variation could be at least estimated, if not explicitly calculated. The test functions used were all normalized so that they integrated to 1.

For each experiment, the results of error versus N are plotted on a log log scale (base 2), and an empirical convergence rate $cN^{-\alpha}$ is determined, in which N is the number of quasi-random points used and α is somewhere between 0 and 1 ($\alpha = 0.5$) is the expected behavior of a random sequence). The values of α and c are found by a least-squares linear fit of $cN^{-\alpha}$ to the

data on a log log scale in a certain range of N. Although this formula cannot be expected to be an asymptotic formula for the integration error, there is too much scatter in the numerical results to determine a more precise convergence rate. Moreover, for the range of computationally practical N, this is an adequate indicator of performance.

Within a given range of N, we choose m values of N, denoted N_i, $1 \le i \le m$, that are equally spaced on the logarithmic scale. For example for the range $N = 1{,}000$ to $N = 100{,}000$ and $m = 100$, we have that $N_0 = 1{,}000$, $N_1 = 1{,}047$, ... , $N_{99} = 95{,}499$, $N_{100} = 100{,}000$. While this tends to make the low end of the range of N have a few more points, it is much better than evenly spacing the N_i. On the log scale that causes almost all of the points to be clustered together at the high end.

The experiments were run by generating the first N_1 terms of the sequence in question. Because the Halton sequence starts with a point very close to zero in high dimensions (the first points is $1/2, 1/3, \ldots, 1/q(s)$), where $q(s)$ is the sth prime) and some of the integrands used are highly sensitive to this point, the first 200 terms of the sequence are discarded. This number is arbitrary, and discarding the first 10 terms would probably suffice. The Faure sequence has a similar problem, so following Bratley and Fox [1], it is also started in the middle at the $(p(s)^4 - 1)$th term, where $p(s)$ is the first prime greater than or equal to s.

Once these N_1 terms have been generated, the integral is approximated with them and the square error $(\int_{I^s} f\,dx - (1/N_1)\sum_{i=1}^{N_1} f(x_i))^2$ is recorded. Then the next N_2 terms of the sequence are generated, and the process is repeated. This leads to values of the square error at $N = N_1, \ldots, N_m$ and requires $\sum N_i$ points in the quasi-random sequence.

Unfortunately, just as in the random case, there is a great deal of scatter in the value of the error. A consistent and reliable value of the error is obtained by performing the numerical computation a number of times (or runs) and using the root mean square error for each value of N. The second run begins by calculating the next N_1 terms of the sequence, which are different from any of the terms used in the first run.

It is important to note that the error calculated for each value of N is "independent" of the error calculated for other values of N, since points in the quasi-random sequence are not reused. The advantage of doing this is that the errors associated with N_1 and N_2 will not be correlated, as they would be if the other method were used. This helps make convergence more easily indentifiable. On the other hand, the method of adding $N_2 - N_1$ points to the original N_1 to get the error using N_2 points is exactly the approach used to actually evaluate an integral, when the answer is unknown and determination of convergence rate is not the goal.

2.2. *Simple Multidimensional Integrals*

The first set of trial functions consisted of relatively simple multidimensional functions formed from products of one-dimensional functions, such as those used in [1] and [8]. The advantage of this was that the values of the integral, the variance, and the variation could be easily obtained. These functions could also be chosen to have specific properties such as high variation and low variance. Thus they were important for determining the effect of the various factors in question, even if the integrals themselves are of little practical interest. The functions, their variances, and their variations are given in Table I. For the first function both variance and variation grow exponentially with dimension. For the second function the variance decays to zero while the variation grows exponentially. The variance decays in the third example, while the variation remains nearly constant.

TABLE I

Test Functions for Determining Effect of Variation

Function	Variance	Variation
$\prod_{i=1}^{s} \lvert 4x_i - 2\rvert$	$\left(\frac{4}{3}\right)^s - 1$	$6^s - 2^s$
$\left(1+\frac{1}{s}\right)^s \prod_{i=1}^{s} x_i^{1/s}$	$\left(1+\frac{1}{s^2+2s}\right)^s - 1$	$\left(1+\frac{1}{s}\right)^s (2s-1)$
$\frac{1}{(s-0.5)^s}\prod_{i=1}^{s}(s-x_i)$	$\left(1+\frac{1}{12(s-0.5)^2}\right)^s - 1$	$\frac{s^s-(s-1)^s}{(s-0.5)^s}$

Figures 1 and 2 show plots of log(error) versus log N for the range $N_0 = 1{,}000$ to $N_{25} = 30{,}000$ for the last two functions from Table I. The error here is the root mean square (rms) error described above, obtained from averaging over 50 runs. Results are given for dimensions $s = 5, 10, 20, 30$ for the Sobol' sequence. For reference, the expected value of the rms error for a random sequence is also plotted for each dimension. The calculated errors are represented by points which are connected by lines to help clarify the trend. Because these are

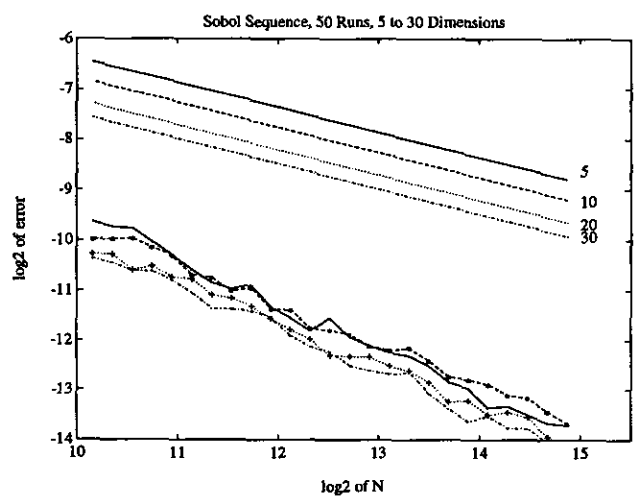

FIG. 1. Integration error for $F = \prod_{i=1}^{s} (1 + 1/s)x^{1/s}$.

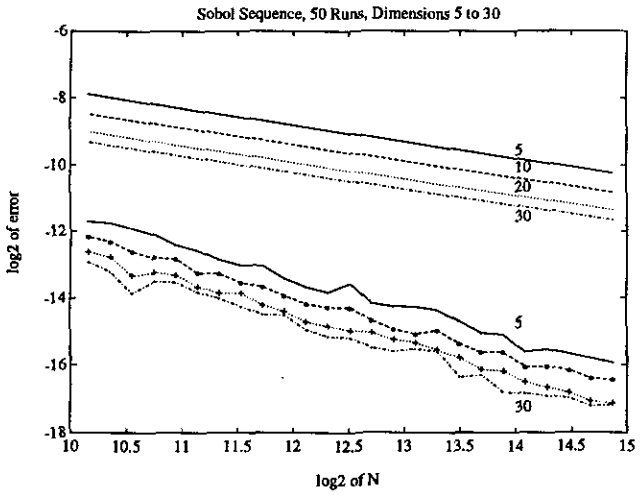

FIG. 2. Integration error for $F = \prod_{i=1}^{s} (s - x_i)/(s - 0.5)$.

log log plots, the linear behavior corresponds to an error of size $cN^{-\alpha}$.

These plots illustrate certain points which hold true in general for quasi-Monte Carlo integration. Of particular importance here is that for the first function, the variation grows exponentially with dimension, while in the second plot, the variation is virtually constant with dimension. Such trends in the error cannot be detected. However, both functions have similar variances, and also similar behavior with regard to error. From these experiments, as well as those described below, it can be seen that the use of quasi-random sequences generally results in lower error than with random sequences. As the random error is determined by the variance of the integrand, the variance may then be interpreted as part of an approximate upper bound on the integration error.

The results of this study verify what was suggested above, that variation is not an important quantity in determining integration error. These examples were of limited generality, however, since the integrands were all product functions. In the next three subsections, more realistic examples are presented, which contain many of the features of true scientific or engineering problems, but in a simplified or restricted setting where the exact solution is known. The roles of dimension and choice of sequence will be considered more carefully in the next section.

2.3. *Absorption*

Monte Carlo methods are frequently used to solve integral equations associated with transport problems. The behavior of quasi-random sequences in this setting was studied by Sarkar and Prasad in [19], where a fairly simple one-dimensional absorption problem was investigated. We consider an even simpler problem, the integral equation

$$y(x) = \int_x^1 \gamma y(x')\, dx' + x,$$

which describes particles traveling through a one-dimensional slab of length one. In each step the particle travels a distance which is uniformly distributed on [0, 1]. This may cause it to exit the slab; otherwise, it may be absorbed with probability $1 - \gamma$ before the next step. In the equation, x describes the current position of the particle, and $y(x)$ gives the probability that the particle will eventually leave the slab given that it has already made it to x. The quantity of interest to calculate is then $y(0)$, the probability that a particle entering the slab will leave the slab.

The solution of this test problem is

$$y(x) = \frac{1}{\gamma}(1 - (1 - \gamma)\exp(\gamma(1 - x))).$$

The solution may also be represented by an infinite-dimensional integral over the unit cube

$$y(x) = \int_{I^\infty} \sum_{n=0}^{\infty} F^n(x, \bar{R})\, d\bar{R},$$

where

$$F^n = \gamma^n \theta\left(1 - x - \sum_{j=1}^{n} R_j\right) \times \theta\left(\sum_{j=1}^{n+1} R_j - (1 - x)\right).$$

Here $\theta(z)$ is the Heaviside function

$$\theta(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}.$$

This corresponds to a Monte Carlo particle simulation where the particle does a simple forward random walk with jump size uniformly distributed on [0, 1]. If it leaves on its $(n + 1)$th jump, it contributes γ^n to the sum which approximates $y(x)$. Because the high dimensions represent the contributions of particles which undergo many collisions before leaving the slab, and the likelihood that a particle can go more than a few collisions before either leaving or being absorbed is quite small, $y(x)$ can be quite accurately represented by truncating the integral at a finite n. For the purposes of computational experiments, the cutoff $n = 20$ was chosen, although the same results would have been obtained with $n = 6$.

When $x = 0$ and the integrand is normalized so that the above integral is 1, the variance of the function can be calculated as

$$\sigma^2(F_{\text{normalized}}) = \frac{\gamma}{1 - (1 - \gamma)e^{\gamma}} - 1.$$

The variation of F is infinite, because it is composed of characteristic functions of nonrectangular sets (cf. Section 3). Figure

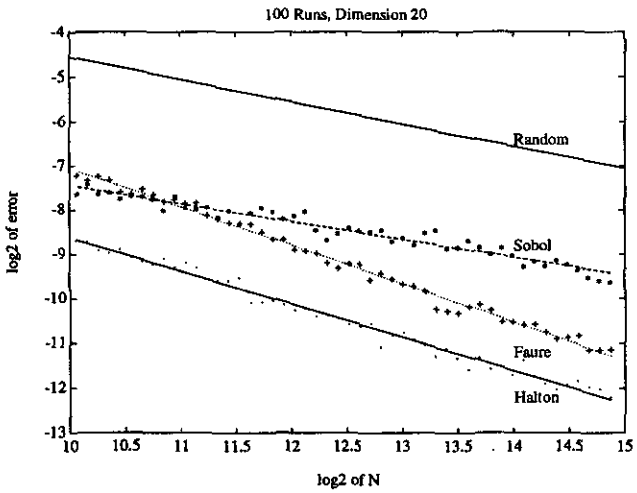

FIG. 3. Error for analog absorption particle simulation.

3 shows the results of running this simulation for a survival probability of $\gamma = 0.5$. The data were taken at 50 points evenly distributed on a log scale between $N = 1{,}000$ and $N = 30{,}000$. The data were averaged over 100 runs. The results are discussed below, where they are compared with results from a modified version of this simulation which has finite variation. This second simulation is described next.

Following the work of Sarkar and Prasad, the "analog" simulation described above can be adapted to produce a smooth integrand. The idea is that each jump that the particle makes should contribute something to the evaluation of the integral, so that there is not an all-or-nothing (discontinuous) contribution if the particle leaves the slab or is absorbed. In fact in the new simulation, the particle never leaves the slab at all and is never absorbed. In this framework, the solution $y(x)$ can be written as

$$y(x) = x + \int_{I^\infty} \sum_{n=0}^{\infty} F^n(x, \bar{R})\, d\bar{R},$$

where

$$F^n = \gamma^n (1 - x)^n \left(\sum_{i=1}^{n-1} R_i^{n-i} \right) \left(1 - (1 - x) \prod_{j=1}^{n} R_j \right).$$

For reasons given above, it is only necessary to go out to a finite n (number of jumps that the particle makes), here chosen to be 20. The corresponding simulation follows a particle through 20 jumps. If after n jumps the particle is at position x', the length of the next jump is sampled uniformly from the interval $[0, 1 - x']$ and the quantity $\gamma^n(1 - x')^n$ is added to the sum which approximates $y(x)$.

The most important difference between the second simulation approach and the first is that the integrand of the second is now smooth and has much lower variance. The variance of F starting from initial particle position $x = 0$ can be given explicitly by a rather lengthy formula. Once the integrand has been normalized, the value of this variance is 0.0742, whereas the variance for the first method (nonsmooth) is 1.8467, assuming a survival probability of $\gamma = 0.5$ is used. As stated above, the variation of the first method is infinite; because the integrand for the second method is smooth, it has finite variation.

Figure 4 shows the computational results of using quasi-random sequences to perform the second simulation and is to be compared with Fig. 3 for the non-smooth simulation. Again a γ of 0.5 was used. Since the effective dimension of this problem was only 6, it is not surprising that the smooth case results look quite similar to the results of the other low-dimensional test problems, with the Halton sequence giving the best performance. The Faure sequence has somewhat higher error than the others, but this can be attributed to the fact that the prime used to generate the sequence was 23, the first prime larger than the dimension 20. The rate of convergence for all the sequences is around $N^{-0.95}$ for the smooth case.

On the other hand, for the analog particle simulation of Fig. 3, Halton has the lowest error and a convergence rate of $N^{-0.7}$. Faure has larger error, but a slightly better convergence rate of $N^{-0.85}$. The slow convergence of the Sobol's sequence is unusual and puzzling; it illustrates the difficulty of predicting convergence for quasi-random sequences. In the next section more examples of characteristic functions are studied.

2.4. *Boltzmann Collision Integral*

The second example comes from the Boltzmann equation, which describes the evolution of a distribution function $f(\xi)$ for the density of a rarefied gas in velocity space. Suppose $f(\xi)$ can be written as

$$f(\xi) = f_1(\xi) + f_2(\xi),$$

where the f_i are Maxwellians, i.e., they have the form

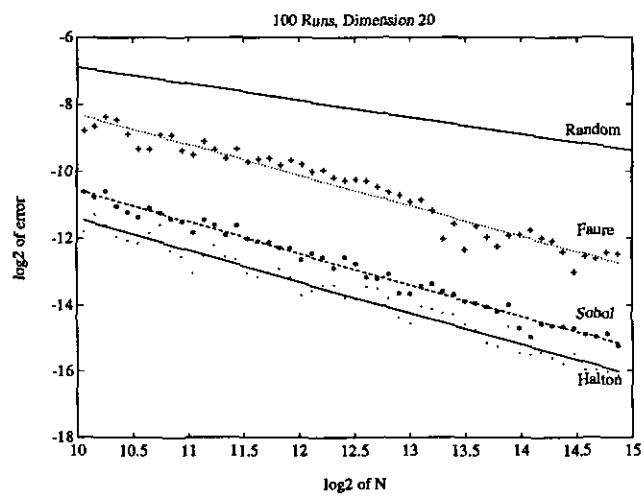

FIG. 4. Error for smooth absorption particle simulation.

$$f_i(\xi) = \rho_i(\beta_i/\pi)^{3/2} \exp\{-\beta_i(\xi - \mathbf{u}_i)^2\}.$$

Here ρ, $\mathbf{u}$, and β are the fluid dynamic density, velocity, and inverse temperature. In [2] Deshpande and Narasimha give an exact formula for this case for the gain in $f(\xi)$ due to binary collisions. This gain can be described by an integral over all possible collisions which result in a particle with velocity ξ.

If the hard sphere collision model is used with molecular diameter $\sigma = 2$, and the Maxwellians are chosen so that $\rho_1 = \rho_2 = 1$, $\beta_1 = \beta_2 = 1$, $\mathbf{u}_1 = (u, 0, 0)$, and $\mathbf{u}_2 = (0, u, 0)$, then the exact value of the gain integral for $f(0)$ is given by the series

$$\frac{4}{\pi} \exp\{-2u^2\} \sum_{k=0}^{\infty} c_k u^{2k},$$

where

$$c_k = \frac{3^{k+1} - 1}{1 \cdot 3 \cdot 5 \cdots (2k+1)}.$$

The gain term to which this corresponds is an integral over the velocity variable $\mathbf{w} = (w_1, w_2, w_3)$ and two angle parameters χ and ε which determine the type of collision. If $\mathbf{w}^2 = w_1^2 + w_2^2 + w_3^2$, $w = \sqrt{\mathbf{w}^2}$, and $w_{2,3} = \sqrt{w_2^2 + w_3^2}$, then the integral can be written as

$$\frac{1}{\pi^3} \int_{\mathbb{R}^3} \int_0^{2\pi} \int_0^{\pi} w \sin\chi \, e^{h(u,\mathbf{w},\chi,\varepsilon)} e^{-\mathbf{w}^2} \, d\chi \, d\varepsilon \, d\mathbf{w},$$

where

$$\begin{aligned} h(u, \mathbf{w}, \chi, \varepsilon) &= -2u^2 + u(w_1 + w_2 + (w_2 - w_1)\cos\chi) \\ &\quad + u \sin\chi \left[w_{2,3} \sin\varepsilon + \frac{w w_3 \cos\varepsilon + w_1 w_2 \sin\varepsilon}{w_{2,3}} \right]. \end{aligned}$$

This is the first example given here which is not explicitly an integral over the unit cube. But the standard Monte Carlo methods for sampling variables which are not uniformly distributed on the interval [0, 1] may also be used here to create quasi-random sequences with the required distribution. In this case these sampling methods are equivalent to the change of variables

$$\begin{aligned} w_i &= \operatorname{erf}^{-1}(2x_i - 1), \quad i = 1, 2, 3 \\ \chi &= \pi x_4 \\ \varepsilon &= 2\pi x_5. \end{aligned}$$

Here erf^{-1} is the inverse of the error function

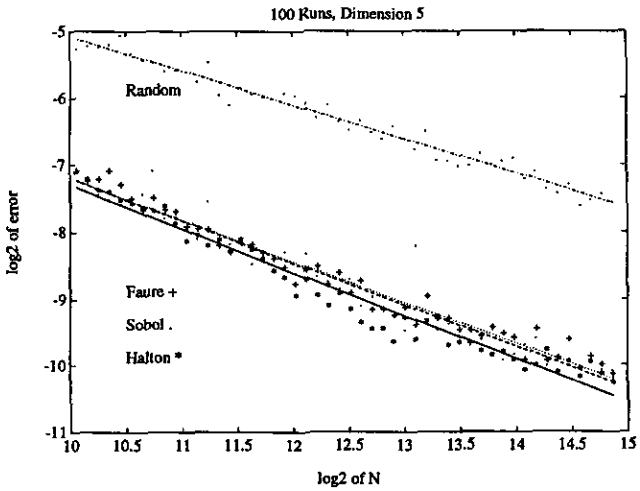

FIG. 5. Boltzmann collision integral for two maxwellians.

$$\operatorname{erf}(w) = \frac{2}{\sqrt{\pi}} \int_0^w e^{-t^2} \, dt.$$

Under this change of variables the integral becomes

$$2\sqrt{\pi} \int_{I^5} w \sin\chi \, e^{h(u,\mathbf{w},\chi,\varepsilon)} \, d\mathbf{x}.$$

It should be noted here that many sampling techniques, when viewed as a change of variables, produce nonsmooth integrands or may map the characteristic function of rectangular set to an irregular set. As discussed in the next section, this may cause a decay in the performance of the quasi-random sequences which is not observed in the random case. Thus care must be taken when using low discrepancy sequences with the standard methods of Monte Carlo, including sampling and variance reduction.

The results of the convergence experiment for this five-dimensional integral are given in Fig. 5. Here u was taken to be 0.25, and the results shown are relative error (the integral is normalized to one). All three quasi-random sequences show similar behavior which is significantly better than the pseudo-random behavior. The least-squares fit convergence rates range from $N^{-0.65}$ to $N^{-0.61}$ for these sequences, while pseudo-random shows a convergence rate of $N^{0.51}$.

2.5. *Feynman–Kac Path Integrals*

A third situation where Monte Carlo methods are frequently used is the evaluation of Feynman–Kac path integrals. Although diffusion Monte Carlo, Green's function Monte Carlo, or the Metropolis algorithm are more commonly used to evaluate such path integrals, this direct evaluation provides a challenging, but easily stated test problem for quasi-random integration.

If $V(x)$ is a potential function, then the quantity $u(x, t)$ given by

$$u(x, t) = E_x\left[\exp\left\{-\int_0^t V(b(\tau))\, d\tau\right\}\right] \tag{4}$$

may be considered. Here E_x is the expectation over all Brownian motion paths $b(\tau)$ that start at x. In [5] Kac shows that $u(x, t)$ satisfies

$$u_t = u_{xx} - V(x)u. \tag{5}$$

We obtain a high-dimensional test problem from Eq. (4) by discretizing in time and considering a random walk with Gaussian steps of variance Δt as an approximation to Brownian motion. Equation (4) is then expressed as

$$u^n(x) = \int_{\mathbb{R}^n} \prod_{k=1}^{n} \left(1 - \Delta t\, V\left(x + \sqrt{\Delta t} \sum_{i=1}^{k} g_i\right)\right) u^0(x)\, dG_1 \cdots dG_n.$$

Here

$$dG_i = \frac{1}{2\pi} e^{-g_i^2/2}\, dg_i.$$

Now $u^n(x)$ satisfies the time-discretized version of Eq. (5)

$$u^{n+1}(x) = (1 - \Delta t V(x))u^n(x) + \Delta t u^n_{xx}(x).$$

If $V(x)$ is taken to be e^{ix} and the initial data is 2π periodic, then the Fourier transform of the discrete equation shows that

$$\hat{u}^{n+1}(k) = (1 - \Delta t k^2)\hat{u}^n(k) - \Delta t \hat{u}^n(k - 1). \tag{6}$$

If the initial data are chosen to be constant, $u^0(x) = 1/2\pi$, then all the nonzero modes are initially zero, while $\hat{u}^0(0) = 1$. It follows from Eq. (6) that all negative modes of u^n remain zero, while the positive modes change and the zeroth mode remains constant. This means that

$$\int_0^{2\pi} u^n(x)\, dx = \int_0^{2\pi} u^0(x)\, dx$$

for all n. By setting $s = n$ and $\Delta t = 1/s$, the $(s + 1)$-dimensional integral representing the total mass of u, the solution of Eq. (5), at time $T = 1$, is

$$\int_0^{2\pi} \int_{\mathbb{R}^s} \prod_{k=1}^{s} \left(1 - \frac{1}{s} \exp\left\{i\left(x + \sqrt{\frac{1}{s}} \sum_{i=1}^{k} g_i\right)\right\}\right)$$
$$u^0(x)\, dG_1 \cdots dG_s\, dx.$$

This is just the integral of $u^s(x)$ which is equal to the integral of the initial data and thus equal to one.

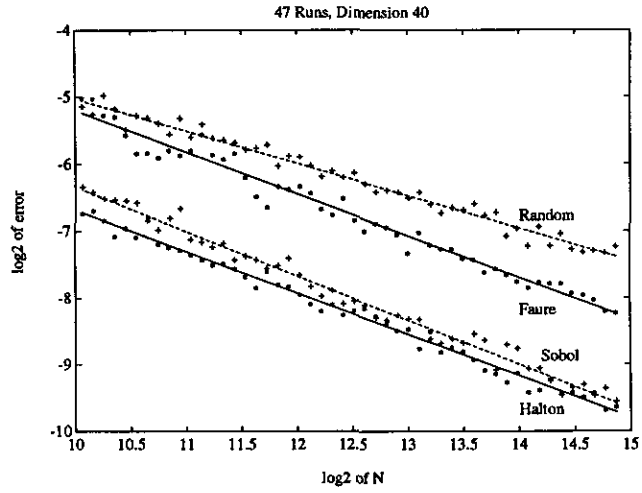

FIG. 6. Feynman–Kac path integral.

Again this is not an integral over the unit cube, but the same methods used for the Boltzmann collision integral may be used here. The inverse error function is used to generate Gaussian distributed points, and x is sampled uniformly on $[0, 2\pi]$ by multiplying uniform numbers on [0, 1] by 2π. The results of the experiment using pseudo- and quasi-random sequences to evaluate this integral in 40 dimensions ($s = 39$ time steps plus one spatial integration) are shown in Fig. 6. Even though the dimension is rather high, the Halton sequence still gives the lowest errors, although the Sobol' sequence is very close. On the other hand, the Faure sequence is noticeably worse. For small N it leads to errors roughly the size of those given by the pseudo-random sequence. All three quasi-random sequences show convergence between $N^{-0.65}$ and $N^{-0.61}$, while the least-squares fit for pseudo-random indicates convergence as $N^{-0.49}$. Although in this computation each dimension is of equal significance, it may be possible to formulate an improved method in which the higher dimensions are of decreasing significance.

3. CHARACTERISTIC FUNCTIONS, CONTINUITY, AND DIMENSION

The performance of quasi-random sequences in evaluating integrals of characteristic functions is of considerable interest because this indicates their potential for success in Monte Carlo particle simulations. As illustrated in the last section, a particle simulation may frequently be described by a multidimensional integral. Characteristic functions arise whenever a yes/no decision must be made, such as if the particle is absorbed. The characteristic function to be integrated corresponds to the subset of the parameter space where the parameter values indicate a positive decision. When the domain of the parameter space is mapped onto the unit cube, the volume of the "yes" set (i.e., the value of the integral of its characteristic function) is just the probability that the decision is accepted.

TABLE II

Radii of Sets Used for Characteristic Function Integration

Shape	$s = 2$	$s = 3$	$s = 4$	$s = 5$	$s = 6$
Cube	$\left(\frac{7}{8}\right)^3$	$\left(\frac{7}{8}\right)^2$	$\left(\frac{7}{8}\right)^{3/2}$	$\left(\frac{7}{8}\right)^{6/5}$	$\frac{7}{8}$
	0.670	0.766	0.818	0.852	0.875
Cone	$\frac{\pi^2}{360}$	$\frac{1}{2}\left(\frac{\pi}{30}\right)^{1/2}$	$\frac{1}{2}\left(\frac{\pi}{15}\right)^{1/3}$	$\frac{1}{2}\left(\frac{4}{9}\right)^{1/4}$	$\frac{1}{2}$
	0.027	0.162	0.297	0.408	0.5

In this section another side of the variation issue is also raised. In general characteristic functions have infinite variation, the exception being rectangles with sides parallel to the coordinate axes. The question of whether the infinite variation causes performance degradation is addressed by also considering integration of a continuous and a smooth function with support on a sphere inside the unit cube.

3.1. *Characteristic Functions of Cubes and Cones*

In order to determine the effects of jumps in the integrand, integration experiments were run using characteristic functions of a cube and a cone. The cube was used because it is the basis of the definition of discrepancy and has finite variation equal to 2^s. This is related to the fact that there is a jump of size one in the direction of the coordinate axes at each corner. The cone was chosen because it has infinite variation and is pointed. The experiments were run to compare sequences and to determine convergence rates as a function of dimension.

In order to compare results across dimensions, the integrands were normalized so that the integral of each function for all shapes and dimensions was one. In a given shape class, the largest radius possible was used in dimension six. The radii used for dimensions two through five were then computed such that these functions would have the same variance. This has the advantange that if random numbers were used for the calculation, the size of the error would be the same in all dimensions.

For the cube, the center was chosen to be $(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$, and side length was the "radius" parameter. In six dimensions, the side length was chosen to be $\frac{7}{8}$, which gives a normalized variance of 1.228. The cone was described by the set

$$\left\{(x_1, \ldots, x_s) : 4R^2\left(x_s - \frac{1}{2}\right)^2 - \sum_{i=1}^{s-1}\left(x_i - \frac{1}{2}\right)^2 \geq 0\right\}.$$

This actually describes two cones with a common tip at $(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$ and circular bases of radius R lying in the $x_s = 0$ and $x_s = 1$ planes. In six dimensions R was set to $\frac{1}{2}$, and the corresponding normalized variance was $360/\pi^2 - 1$, which is about 35.48. Table II shows the radii (or side lengths) used for each shape and each dimension.

Table III shows the results of using a pseudo-random sequence (R) and the three quasi-random sequences (H, S, F) to integrate the characteristic functions of the sets described above. The experiment consisted of 50 runs using 50 logarithmically evenly spaced values of N in dimensions two through six. For both the shapes, the size of the error increases with dimension. Convergence behavior is generally best in two dimensions and tends to decay as the dimension increases. For the most part, the Halton sequence gives slightly lower errors, especially in two dimensions. For the cube, convergence in two dimensions is around $N^{-0.95}$ for all sequences, while for dimensions three to six, the least-squares fit convergence rates ranged from about $N^{-0.71}$ to $N^{-0.89}$. The Faure sequence appears to start out with slightly higher error, but it converges a little faster, so that all the sequences end up with similar error size at $N = 30{,}000$.

For the cone, the error is considerably closer to the expected random error in size and in convergence rate than for the cube. For dimensions four to six, all the sequences and dimensions look about the same, except for the four-dimensional Sobol' sequence, which performs rather erratically. The convergence rates range from $N^{-0.54}$ to $N^{-0.62}$. Again, the Halton sequence is marginally better than the others, with the greatest difference being in dimension two. Otherwise, Halton and Faure look very much alike, while Sobol' gives errors in about the same range, but is less predictable.

A similar set of experiments were done for the characteristic function of a sphere in dimensions two through six. These showed error behavior similar to that of the cone. For the

TABLE III

Integration Error for Characteristic Functions of Cube and Cone

Shape	Dimension	Sequence	Convergence rate	Error at $N = 32{,}768$
Cube	2–6	R	−0.50	0.00609
	2	H	−0.96	0.00019
		S	−0.97	0.00031
		F	−0.92	0.00035
	4	H	−0.79	0.00065
		S	−0.81	0.00061
		F	−0.89	0.00086
	6	H	−0.76	0.00092
		S	−0.71	0.00094
		F	−0.82	0.00115
Cone	2–6	R	−0.50	0.033
	2	H	−0.79	0.0045
		S	−0.79	0.0068
		F	−0.66	0.0081
	4	H	−0.62	0.012
		S	−0.59	0.023
		F	−0.61	0.013
	6	H	−0.59	0.015
		S	−0.54	0.015
		F	−0.54	0.017

sphere, the two-dimensional convergence rate for all sequences was about $N^{-0.78}$, while the higher dimensions had a rate around $N^{-0.6}$. Again, Halton performed slightly better in two dimensions; however, the Sobol' sequence did not look particularly odd or unpredictable compared with the others.

The fact that quasi-random sequences performed better on the cube was to be expected. The sequences are constructed to minimize discrepancy, which is based on rectangles. Also, each multidimensional sequence is made from a combination of one-dimensional sequences. This corresponds to a cube being the product of one-dimesional characteristic functions. However, it should be noted that the $1/N$ type convergence suggested by the discrepancy bound (ignoring the log factors) is only approached in two dimensions, at least for the range of N considered here. For nonrectangular characteristic functions, it appears that a convergence rate only moderately better than that of a random sequence can be expected from quasi-random sequences. This is of particular importance for particle simulations where a decision process may be equivalent to such a function.

On the other hand, note that in all of these experiments the error for quasi-Monte Carlo is significantly lower than the error for standard Monte Carlo at the largest value of N. This indicates that quasi-Monte Carlo can decrease the constant c in the integration error size $cN^{-\alpha}$, even if it does not significantly improve the decay rate α. While this is reminiscent of variance reduction techniques, it is important to note that quasi-Monte Carlo concerns manipulating the source "random" points, whereas variance reduction involves changing the integrand and integration domain. Thus these are two different techniques which may be used simultaneously.

3.2. *Continuous and Smooth Functions on a Sphere*

The question arises now as to whether the poor performance of quasi-random sequences on characteristic functions relative to the smooth functions examined earlier is connected to the fact that these functions have infinite variation. To investigate this, two more sets of experiments were conducted. Both involved functions which have support on spheres contained within the unit cube. The first was the continuous but not differentiable function

$$F(x_1, \ldots, x_s) = \begin{cases} R(s) - r, & r \le R(s) \\ 0, & r > R(s) \end{cases}.$$

Here

$$r = \left(\left(x_1 - \frac{1}{2} \right)^2 + \cdots + \left(x_s - \frac{1}{2} \right)^2 \right)^{1/2},$$

TABLE IV

Radii for Functions with Spherical Support

Function	$s = 2$	$s = 3$	$s = 4$	$s = 5$	$s = 6$
Continuous	$\frac{\pi}{\sqrt{448}}$	$\left(\frac{\pi^2}{560}\right)^{1/3}$	$\left(\frac{5\pi}{1008}\right)^{1/4}$	$\left(\frac{15\pi}{3136}\right)^{1/5}$	$\frac{1}{2}$
	0.148	0.260	0.353	0.432	0.5
Smooth	$\left(\frac{\pi^2}{128}\frac{H(2)}{H(6)}\right)^{1/2}$	$\left(\frac{\pi^2}{256}\frac{H(3)}{H(6)}\right)^{1/3}$	$\left(\frac{\pi}{128}\frac{H(4)}{H(6)}\right)^{1/4}$	$\left(\frac{3\pi}{512}\frac{H(5)}{H(6)}\right)^{1/5}$	$\frac{1}{2}$
	0.146	0.256	0.350	0.430	0.5

and $R(s)$ is the radius given in Table IV. Because this function is not differentiable on I^s, it has infinite variation. The second example was the C^∞ function

$$F(x_1, \ldots, x_s) = \begin{cases} \exp[-(20z(1-z))^{-2}], & z \le 1 \\ 0, & z > 1 \end{cases}.$$

Here $z = r/R(s)$. The variation of the function can be estimated if it is assumed that in each quadrant the sth derivative will be of one sign. Because the function is zero at the boundary of the cube, the variation will be the sum of the function integrated over the boundary of each quadrant of the sphere. Because there are 2^s quadrants, it follows that the variation will grow exponentially with dimension.

A new set of radii had to be calculated for each function because the values of the integral and the variance are different than for the characteristic function of the sphere. Again the radii were chosen to maintain constant variance across dimension. For the continuous function, the normalized variance was $672/\pi^3 - 1$, or about 20.7. For the C^∞ function some additional notation is necessary to precisely give the radii and variance. Let

$$f(z) = \exp[-(20z(1-z))^{-2}]$$

for $0 \le z \le 1$. For dimension $s = 2$ to 6, define

$$H(s) = \frac{\int_0^1 z^{s-1} f^2(z)\, dz}{\left(\int_0^1 z^{s-1} f(z)\, dz \right)^2}.$$

Then the normalized variance of the smooth function is $64\, H(6)/\pi^3 - 1$, which is around 15.63. The appropriate radius necessary to maintain variance for each dimension and each of the functions is listed in Table IV.

TABLE V

Integration Error for Functions with Support on Sphere

Function	Dimension	Sequence	Convergence rate	Error at $N = 32768$
Continuous	2–6	R	−0.50	0.025
	2	H	−0.99	0.0006
		S	−1.00	0.0014
		F	−0.97	0.0015
	4	H	−0.95	0.0020
		S	−0.70	0.0037
		F	−0.75	0.0033
	6	H	−0.65	0.0048
		S	−0.73	0.0039
		F	−0.72	0.0047
Smooth	2–6	R	−0.50	0.022
	2	H	−0.90	0.0011
		S	−0.94	0.0023
		F	−0.94	0.0023
	4	H	−0.74	0.0039
		S	−0.62	0.0096
		F	−0.68	0.0049
	6	H	−0.62	0.0070
		S	−0.69	0.0050
		F	−0.65	0.0071

Table V shows the results of integrating the continuous function and the C^∞ function, respectively. For the continuous function, all the sequences in dimensions two and three are converging at a rate close to $1/N$, with the Halton sequence having the lowest error again. The higher dimensions look about the same across the sequences and have a convergence rate of around $N^{-0.7}$. For the smooth function, Halton is again the best in two dimensions and shows a steady decay in convergence rate with dimension, going from $N^{-0.9}$ for $s = 2$ to $N^{-0.62}$ for $s = 6$. The Sobol' sequence shows the same kind of erratic behavior it did for the cone, with the fourth dimension again being particularly bad. The errors are about the same size as with Halton (except for two dimensions). The convergence rates for Sobol' range from $N^{-0.94}$ in two dimensions to $N^{-0.62}$ in four dimensions, although these rates do not appear too reliable. The Faure sequence lies somewhere between Halton and Sobol', with error size and convergence rates of the same magnitude.

What is interesting about these results is that despite having infinite variation and higher variance, the continuous function has smaller error and better convergence than the smooth function. The convergence rates for the continuous function are about the same as for the cube and better than for the other characteristic functions. This suggests that continuity of the integrand is important for improved quasi-random convergence, but that differentiability does not necessarily help. These examples show that neither variation, variance, nor smoothness consistently predicts quasi-random integration error behavior.

3.3. *Theoretical Bounds for Integration of Characteristic Functions*

To try and approach the convergence of quasi-random sequences on characteristic functions of nonrectangular sets analytically, Kuipers and Niederreiter [7] define the isotropic discrepancy J_N of a sequence $(x_1, \ldots, x_N)$ in I^s as

$$J_N = \sup_{C \in \mathcal{C}} \left| \frac{A(C; N)}{N} - m(C) \right|,$$

where $\mathcal{C}$ is the class of all convex subsets of I^s, $A(C; N)$ is the number of x_i which are inside C, and $m(C)$ is the measure of C. In [Theorems 1.5 and 1.6, pp. 94–97], they prove the bound

$$D_N \leq J_N \leq C_s D_N^{1/s},$$

where C_s is a constant depending only on dimension. Because J_N is a bound on the integration error associated with characteristic functions of convex sets, this bound suggests that the convergence rate for such functions may only be $N^{-1/s}$. The computed examples of such functions show decreased convergence rates with increasing s, but nothing this extreme.

Press and Teukolsky [18] suggest another argument to explain the observed decline in convergence rates of quasi-random sequences when applied to characteristic functions. They reason that near the boundary of the set described by the characteristic function, whether a point of the sequence lands inside the set (and thus contributes a value of one to the average) or outside the set (where the function is zero) is essentially random for nonrectangular sets. Combining the random error near the boundary, which has $N^{-0.5}$ type behavior, with the superior quasi-random behavior over the rest of the set, described by the discrepancy D_N, leads to a new convergence estimate. As dimension increases, the boundary of any set plays an increasingly dominant role. This helps explain the decline in convergence rate with increasing dimension.

This argument can be made more rigorous by reworking the theorems mentioned above under the assumption of randomness near the boundary. The following lemma summarizes this result.

LEMMA 1. *Let $(x_1, \ldots, x_N)$ be a sequence of N points in I^s with discrepancy D_N and isotropic discrepancy J_N, such that $D_N \leq N^{-0.5}$. For any convex set C, assume that for a set of sufficiently small measure which contains the boundary of C, the error in approximating the measure of the small set with the sequence is bounded by a constant times the expectation for a random sequence. Then*

$$J_N \leq C_s [N^{-(s-1)} D_N]^{1/(2s-1)}.$$

Proof. Theorems 1.5 and 1.6 in [7] show that it can be assumed that C is a closed convex polytope, which means that it is possible to construct sets P_r which are unions of a finite

number of disjoint rectangles such that $P_r \subset C$ and $m(C) - m(P_r)$ is arbitrarily small. Consider such a set P_r. Then

$$\begin{aligned}\left|\frac{A(C;N)}{N} - m(C)\right| &= \left|\frac{A(C;N)}{N} - m(C) + \frac{A(P_r;N)}{N} - m(P_r) \right.\\ &\qquad \left. - \frac{A(P_r';N)}{N} + m(P_r)\right| \\ &\le \left|\frac{A(P_r;N)}{N} - m(P_r)\right| \\ &\quad + \left|\frac{A(C-P_r;N)}{N} - m(C-P_r)\right| \\ &\le \left|\frac{A(P_r;N)}{N} - m(P_r)\right| + k\left(\frac{m(C-P_r)}{N}\right)^{1/2}.\end{aligned}$$

This last inequality follows from the assumption of random-like behavior near the boundary of C. Here k is a constant. For any positive integer r, Kuipers and Niederreiter show how to construct a set P_r such that

$$\left|\frac{A(P_r;N)}{N} - m(P_r)\right| \le r^{s-1} D_N$$

and

$$m(C - P_r) \le \frac{2s\sqrt{s}}{r}.$$

Thus it follows that

$$\left|\frac{A(C;N)}{N} - m(C)\right| \le r^{s-1} D_N + k\left(\frac{2s\sqrt{s}}{rN}\right)^{1/2}.$$

Because r can be any positive integer, it may be chosen to optimize the bound. Let r be the first integer smaller than $(ND_N^2)^{-1/(2s-1)}$. Then it follows that $1/r \le 2(ND_N^2)^{1/(2s-1)}$, and the bound can be written

$$\begin{aligned}\left|\frac{A(C;N)}{N} - m(C)\right| &\le (ND_N^2)^{-(s-1)/(2s-1)} D_N + k_s (ND_N^2)^{1/2(2s-1)} N^{-1/2} \\ &= N^{-(s-1)/(2s-1)} D_N^{1-2(s-1)/(2s-1)} \\ &\quad + k_s N^{(1/2)(1/(2s-1)-1} D_N^{1/(2s-1)} \\ &= N^{-(s-1)/(2s-1)} D_N^{1/(2s-1)} + k_s N^{-(s-1)/(2s-1)} D_N^{1/(2s-1)} \\ &= C_s [N^{-(s-1)} D_N]^{1/(2s-1)}.\end{aligned}$$

Because the right-hand side of the inequality is independent of the convex set chosen, the left-hand side may be replaced by the sup over all convex sets. This gives the result stated in the lemma.

Even if the optimistic approximation for D_N of N^{-1} is used, the resulting bound on the integration error for characteristic functions of convex sets is

$$|error| \le C_s N^{-s/(2s-1)}.$$

This result suggests that for an integration problems with discontinuous integrand in a high dimension, convergence much better than that of a random sequence cannot be expected. This is a result of the fact that sets in high dimensions are almost entirely boundary, where the integration error of a quasi-random sequence behaves like the error for a random sequence.

A comparison of this theoretical bound with the above experimental results for characteristic functions shows again that the bound is somewhat pessimistic. For example, for the characteristic function of the cone, the Halton sequence in two dimensions has a calculated convergence rate of about $N^{-0.8}$, whereas the bound from the lemma has a convergence rate of $N^{-0.67}$. In six dimensions the Halton convergence rate is $N^{-0.6}$, while the bound predicts a convergence rate of $N^{-0.55}$. The actual integration error shows some similarities to the bound behavior, but the rates of convergence are better than predicted. Moreover the constant c in the integration error rate ($cN^{-\alpha}$) is often seen to be improved through quasi-Monte Carlo.

4. CONCLUSIONS

The computational experiments described above show that Quasi-Monte Carlo methods, using quasi-random (i.e., low discrepancy) sequences, provide an effective integration technique for many multidimensional integrals. Moreover, the error in the quasi-Monte Carlo integration method is found to be significantly less than the corresponding error for a standard Monte Carlo (i.e., random or pseudo-random) method. Of the quasi-random sequences tested, for low-dimensional problems up to around $s = 6$, the Halton sequence generally gives the best results. In higher dimensions, for most problems the Sobol' sequence was superior. The Faure sequence, which has the best theoretical bound [9], was generally better than random, but was for the most part outperformed by Halton or Sobol'.

It is important to note, however, that the error reduction for quasi-Monte Carlo methods is limited by several factors. For integration of smooth functions in one dimension, the error is of size $c_1 N^{-1}$, compared to error size $c_2 N^{-1/2}$ for random simulation (which has this error size for all dimensions). If the dimension is increased or the integrand function is less smooth, the observed error may be of size $c_1 N^{-\lambda}$ in which $\frac{1}{2} \le \lambda \le 1$. Still the error for quasi-Monte Carlo integration is almost always significantly better than that for standard Monte Carlo, using a random or pseudo-random sequence, due to either a larger algebraic decay rate λ or a smaller constant c. Thus for a

fixed error tolerance level, the quasi-random simulation requires significantly smaller number N of simulation points. In many problems with complicated integrands, use of the simulation points entails a lot of computation, so that the resulting reduction in computational effort will more than compensate for the increased work required to generate the quasi-random sequence.

In another paper [10] we have used a quasi-Monte Carlo method for simulating solutions of the heat equation. The computational results in [10] indicate a dependence on dimension that is quite similar to that for quasi-Monte Carlo integration seen above. They also show that use of quasi-random sequences is delicate, since the elements of the sequence are correlated. For example, efficient simulation of the heat equation is possible only if the particle labels are reordered. Also use of a quasi-random sequence in the Box–Muller method for sampling a Gaussian distribution results in significant loss of efficiency.

These results lead to several conclusions concerning application of quasi-random sequences to Monte Carlo methods. First, there are many problems for which direct application of quasi-Monte Carlo may be superior to standard quadrature or Monte Carlo. Examples include smooth integration problems in intermediate dimension, such as absorption or scattering from rough surfaces, and discontinuous integration problems in low dimension. For such problems, use of quasi-Monte Carlo is superior to standard Monte Carlo (even with variance reduction), since it provides a better convergence rate (not just a better constant).

For more difficult problems, however, the advantages of quasi-Monte Carlo over standard variance reduction are not as clear, and effective use of quasi-random sequences requires more effort. We expect that our results will guide the development of modified variance reduction and other Monte Carlo techniques employing quasi-random sequences. This may require reformulation of the technique to insure that the resulting integrands are smooth and low dimensional. One such example is the modified absorption method in Section 2. Our computational results show that this made the quasi-Monte Carlo method much more accurate.

In a related work, several modified Monte Carlo methods have been developed for effective application of quasirandom sequences. These include a smoothed acceptance-rejection method and a lower dimensional Feynman–Kac integration method in [12] and the Diffusion Monte Carlo method [11]. We have not yet succeeded in application of quasirandom sequences to other Monte Carlo methods, such as the Metropolis algorithm and stratification.

We expect that future progress with quasirandom methods will depend on further modifications of standard Monte Carlo methods for simulation and variance reduction.

REFERENCES

1. P. Bratley and B. L. Fox, *ACM Trans. Math. Software* **14,** 88 (1988).
2. S. M. Deshpande and R. Narasimha, *J. Fluid Mech.* **36**(3), 545 (1969).
3. H. Faure, *Acta Arithmetica* **41,** 337 (1982).
4. J. H. Halton, *Numer. Math.* **2,** 84 (1960).
5. M. Kac, On some connections between probability theory and differential and integral equations, in *Proceedings, 2nd Berkeley Symp. Math. Stat. Prob.*, edited by J. Neyman. (Univ. of California Press, Berkeley, 1951), p. 189.
6. A. Kersch, W. Morokoff, and A. Schuster, *Transp. Theory Stat. Phys.*, to appear.
7. L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley, New York, 1976.
8. Yu. L. Levitan, *et al.* Short communications on quasi-random sequences for numerical computations, *USSR Comput. Math. Math. Phys. (Engl. Transl.)* **28**(3), 88 (1988).
9. W. J. Morokoff and R. E. Caflisch, *SIAM J. Sci. Stat. Comp.* **15,** 1251 (1994).
10. W. J. Morokoff and R. E. Caflisch, *SIAM J. Num. Anal.* **30,** 1558 (1993).
11. B. Moskowitz, *Application of Quasi-Random Sequences to Monte Carlo Methods.* Ph.D. thesis, UCLA, 1993.
12. B. Moskowitz and R. E. Caflisch, *J. Math. Comp. Modeling*, to appear.
13. H. Niederreiter, *Bull. Amer. Math. Soc.* **84,** 957 (1978).
14. H. Niederreiter, *Quasi-Monte Carlo Methods for Multidimensional Numerical Integration, Numerical Integration III,* International Series of Numerical Math., Vol. 85, edited by H. Brass and G. Hämmerlin (Birkhäuser Verlag, Basel, 1988).
15. H. Niederreiter, *Diophantine Approximation and Its Applications*, edited by C. F. Osgood (Academic Press, New York, 1973), p. 129.
16. H. Niederreiter, *Random Number Generators and Quasi-Monte Carlo Methods, SIAM*, Philadelphia, 1992.
17. D. M. O'Brien, *J. Quant. Spectrosc. Radiat. Transfer* **48**(1), 41 (1992).
18. W. H. Press and S. A. Teukolsky, *Comput. Phys.* **Nov/Dec** 76 (1989).
19. P. K. Sarkar and M. A. Prasad, *J. Comput. Phys.* **68,** 66 (1987).
20. I. M. Sobol', *USSR Comput. Math. Math Phys. (Engl. Transl.)* **7**(4), 86 (1967).
21. H. Woźniakowski, *Bull. Amer. Math. Soc.* **24,** 185 (1991).