UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Smoothness and Dimension Reduction in Quasi-Monte Carlo Methods

Bradley Moskowitz
Russel E. Caflisch

March 1994

CAM Report 94-17

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90024-1555

Smoothness and Dimension Reduction in Quasi-Monte Carlo Methods

Bradley Moskowitz *
Russel E. Caflisch †
Mathematics Department, UCLA

March 5, 1994

Abstract

Monte Carlo integration using quasirandom sequences has theoretical error bounds of size $O(N^{-1}\log^d N)$ in dimension d, as opposed to the error of size $O(N^{-1/2})$ for random or pseudorandom sequences. In practice, however, this improved performance for quasirandom sequences is often not observed. The degradation of performance is due to discontinuity or lack of smoothness in the integrand and to large dimension of the domain of integration, both of which often occur in Monte Carlo methods. In this paper, modified Monte Carlo methods are developed, using smoothing and dimension reduction, so that the convergence rate of nearly $O(N^{-1})$ is regained. The standard rejection method, as used in importance sampling, involves discontinuities, corresponding to the decision to accept or reject. A smoothed rejection method, as well as a method of weighted uniform sampling, is formulated below and found to have error size of almost $O(N^{-1})$ in quasi-Monte Carlo. Quasi-Monte Carlo evaluation of Feynman-Kac path integrals involves high dimension, one dimension for each discrete time interval. Through an alternative discretization, the effective dimension of the integration domain is drastically reduced, so that the error size close to $O(N^{-1})$ is again regained.

*Los Angeles, California 90024. e-mail: mosko@math.ucla.edu. Research supported in part by the Air Force Office of Scientific Research under grant number F49620-94-1-0091

†Los Angeles, California 90024. e-mail: caflisch@math.ucla.edu. Research supported in part by the Army Research Office under grant number DAAL03-91-G-0162 and by the Air Force Office of Scientific Research under grant number F49620-94-1-0091

1 Introduction

Monte Carlo integration using quasirandom sequences has theoretical error bounds of size $O(N^{-1}\log^d N)$ [1] in dimension d, as opposed to the error of size $O(N^{-1/2})$ for random or pseudorandom sequences. In practice, however, this improved performance for quasirandom sequences is often not observed. The loss of accuracy was found by Morokoff and Caflisch [2, 3, 4] to be due to two causes:

The first is discontinuity or lack of smoothness in the integrand. The $O(N^{-1}\log^d N)$ error bound quoted above is established using the Koksma-Hlawka inequality, which states that for Monte Carlo integration of a function f on the unit cube in R^d using N points, the integration error $E_N(f)$ is bounded by

$$E_N(f) \leq D_N V(f) \tag{1}$$

in which D_N is the *discrepancy* of the sequence and $V(f)$ is the *variation* of f in the Hardy-Krause sense [1]. Moreover, for quasirandom sequences the discrepancy D_N is of size $O(N^{-1}\log^d N)$. For $V(f)$ to be finite in dimension $d > 1$, in general, f must be smooth (see [1] for a precise statement). On the other hand, discontinuous integrands arise often in practical Monte Carlo methods. For example, many methods involve a decision process, for which some component of the integrand is 1 or 0 if the decision is "yes" or "no." For such a discontinuous integrand, the inequality (1) is inapplicable, and the computation of [2] indicate that the error bounds revert to size $O(N^{-1/2})$.

A second limitation on the accuracy of quasi-Monte Carlo integration occurs for integration problems in which the dimension d of the integration domain is large. For large d, Morokoff and Caflisch [4] found that the discrepancy of the quasirandom sequence is of size $O(N^{-1/2})$ for intermediate values of N, even though D_N is $O(N^{-1}\log^d N)$ for N sufficiently large. The transition from $O(N^{-1/2})$ to $O(N^{-1}\log^d N)$ behavior for D_N appears to occur approximately at $N = e^d$, although this has not been proven.

There are a range of applications, such as scattering from rough surfaces, collisionless particle transport [5, 6], and simulation of small networks [7], for which the dimension is not too large and quasi-Monte Carlo is directly applicable. On the other hand, integration in high dimensions is one of the main reasons for using Monte Carlo in the first place, and high dimensional integration occurs in many problems coming from atomic physics, neutron transport, computational chemistry and molecular biology. These results indicate that it will be difficult to adapt quasi-Monte Carlo methods to such problems.

The purpose of this paper is show how these two fundamental limitations – smoothness and dimension – can be overcome, as least in several examples. This will be accomplished through modification of standard Monte Carlo methods, by smoothing discontinuities and reducing the effective dimension of the problem.

The standard rejection method, as used to sample from a density function in importance sampling, can be phrased as an integration involving a discontinuous integrand, with the discontinuity coming from the decision to accept or reject. As shown in Section 2.2.1 below, this leads to large errors in a quasirandom method. A *smoothed rejection method* will be formulated, for which the optimal error size of nearly $O(N^{-1})$ is attained (here and in the subsequent discussion the logarithmic factors $\log^d N$ are omitted, since they cannot be reliably detected numerically). This is also true of the method of *weighted uniform sampling.*

Quasirandom simulation of a stochastic process, such as Brownian motion, requires a high dimensional quasirandom sequence, with one dimension for each time step that is simulated. This easily leads to loss of the optimal $O(N^{-1})$ convergence rate. This will be illustrated in Section 3.2 below for Feynman-Kac path integrals. Nevertheless, by a modified representation of Brownian motion, or equivalently a modified quadrature rule for the path integral, the effective dimension of the problem can be drastically reduced, so that the optimal error size of approximately $O(N^{-1})$ is attained in this problem as well.

These applications of quasi-Monte Carlo to test problems may serve as a starting point for the application of such methods to a range of problems in which the integrand is smooth and the effective dimension is small, after some modification. For example, Moskowitz [8] has applied these methods to the determination of the ground state energy for the square well potential and the hydrogen atom.

2 Integrand Smoothing

In this section quasi-Monte Carlo integration is applied to problems in which the integrand is first smoothed to remove discontinuities that would naturally appear. The first application is an illustrative example of integration of a characteristic function. Then the rejection method and smoothed versions of the rejection method are presented.

2.1 Example: Smoothing a Characteristic Function

The characteristic function χ_E of the set E has infinite variation, unless the set E is rectangular with sides parallel to the coordinate axes. Consider the characteristic function f_0 of a rotated cube, defined as follows:

Example 1 *Given* $x_1, x_2, x_3 \sim U[-\frac{1}{2}, \frac{1}{2}]$, *let*

$$
\begin{aligned}
u_1 &= \frac{\sqrt{2}\,x_1 + x_2 + x_3}{2} \\
u_2 &= \frac{-\sqrt{2}\,x_1 + x_2 + x_3}{2} \\
u_3 &= \frac{-\sqrt{2}\,x_2 + \sqrt{2}\,x_3}{2}
\end{aligned}
$$

Define discontinuous characteristic function

$$
f_0(x_1, x_2, x_3) = \prod_{k=1}^{3} \chi\left\{|u_k| \le .2\right\}
$$

where $\chi\left\{|u_k| \le .2\right\} = \begin{cases} 1 & \text{if } |u_k| \le .2 \\ 0 & \text{otherwise} \end{cases}$ *, and smoothed functions* f_δ *(*δ *is the level of smoothing)*

$$
f_\delta(x_1, x_2, x_3) = \prod_{k=1}^{3} \chi_\delta\left\{|u_k| \le .2\right\}
$$

where $\chi_\delta\left\{|u_k| \le .2\right\} = \begin{cases} 1 & \text{if } |u_k| \le .2 - \delta \\ 0 & \text{if } |u_k| \ge .2 + \delta \\ \frac{\delta - (|u_k| - .2)}{2\,\delta} & \text{otherwise} \end{cases}$

The integrals of the functions f_0 and f_δ, $0 < \delta \ll 1$, over the unit cube in three dimensions all equal .064.

The results of crude Monte Carlo estimation in three different cases are compared: First, the discontinuous case; second, the continuous case with weak smoothing ($\delta = .025$); and third, the continuous case with strong smoothing ($\delta = .1$). Results are obtained by computing the root mean square error over 75 repetitions. (See Section 2.2.4 for more details.) Log-log plots are used so that the slopes (which are presented parenthetically in the figure keys) correspond to convergence rates. Numerical results are plotted in Figure 1. As expected, the quasirandom sequences, in this example from Halton [9], perform better as the level of smoothing is increased, while pseudorandom sequences are not affected by the smoothing.

The results for this simple example show how the effectiveness of quasi-Monte Carlo integration is lost if the integrand is discontinuous, but can be regained if the integrand is smoothed without changing the value of the integral. Similar tests have been performed by Morokoff and Caflisch [2], Press and Teukolsky [10], and Berblinger [11] with consistent results.

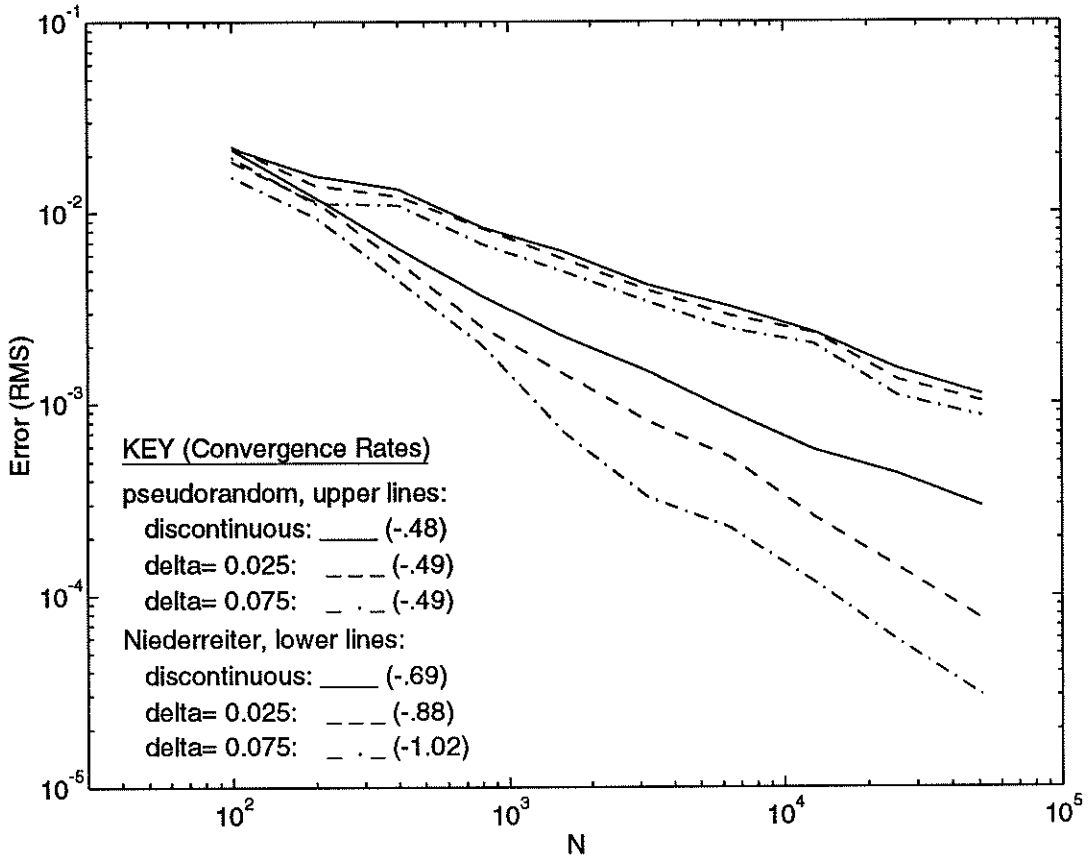


Figure 1: Results, 75 runs, Example 1.

2.2 Importance Sampling Using the Rejection Method

Variance reduction through importance sampling is one of the most commonly used Monte Carlo methods. For the integral $A = \int_D f(\boldsymbol{x})d\boldsymbol{x}$, introduce an importance function h which mimics the behavior of f over D but is either integrable analytically or easily numerically integrated. The sampling procedure is then altered to generate points distributed according to h instead of points which are uniformly distributed. Then, instead of evaluating $f(\boldsymbol{x})$ at each sample point, one evaluates $\frac{f(\boldsymbol{x})}{h(\boldsymbol{x})}$, which provides an unbiased estimate of the integral, since

$$E_h\left(\frac{f}{h}\right) = \int_D \frac{f(\boldsymbol{x})}{h(\boldsymbol{x})} \cdot h(\boldsymbol{x})\, d\boldsymbol{x} = \int_D f(\boldsymbol{x})\, d\boldsymbol{x} = A$$

The Importance Sampled Monte Carlo estimate can be written as follows:

$$\tilde{A}_N = \frac{1}{N}\sum_{i=1}^{N} \frac{f(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i)} \quad , \quad \boldsymbol{x}_i \sim h(\boldsymbol{x}) \tag{2}$$

where $\boldsymbol{x}_i \sim h(\boldsymbol{x})$ indicates that the sample points, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ have a probability density given by h (note that h should be normalized so that $\int_D h\, d\boldsymbol{x} = 1$). The variance of this estimate is substantially reduced if f/h has a smaller variance than f.

For a limited class of density functions h, the sample points may be generated by a transformation of the uniform distribution. For more general h, some other generation procedure must be used, such as the rejection method.

2.2.1 Standard Rejection Method

Here is the basic algorithm for the **Rejection Method** when the original sample is uniformly distributed in the $d-$dimensional unit cube. Note that one additional dimension is added for the extra variable y which acts as a "decision-maker":

1. Select $\gamma \geq \sup_{\boldsymbol{x}\in D} h(\boldsymbol{x})$.
2. Repeat until N points have been accepted:
   (a) Sample $(\boldsymbol{x}_t, y_t)$ from $U([0,1]^{d+1})$.
   (b) If $y_t < \gamma^{-1}\, h(\boldsymbol{x}_t)$, accept trial point $\boldsymbol{x}_t$.
   Otherwise, reject the trial point.

This algorithm produces a sequence of accepted points in d dimensions which are distributed according to h, as required for Importance Sampling.

The sum (2), in which the points $\boldsymbol{x}_i$ are chosen using the rejection method, can be reinterpreted as a Monte Carlo evaluation of the following integral

$$A = \gamma \int_D \int_0^1 \frac{f(\boldsymbol{x})}{h(\boldsymbol{x})} \chi(y < \gamma^{-1} h(\boldsymbol{x})) dy d\boldsymbol{x}. \tag{3}$$

The direct Monte Carlo estimation of this integral is

$$\tilde{A}_N = \frac{\gamma}{N} \sum_{i=1}^{N} \chi \left\{ y_i < \frac{h(\boldsymbol{x}_i)}{\gamma} \right\} \cdot \frac{f(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i)} \tag{4}$$

The variance of this integral is reduced by replacing the total number of points N by the number of accepted points to get the following Monte Carlo estimator

$$\tilde{A}_N = \frac{1}{\sum_{i=1}^{N} \chi \left\{ y_i < \frac{h(\boldsymbol{x}_i)}{\gamma} \right\}} \sum_{i=1}^{N} \chi \left\{ y_i < \frac{h(\boldsymbol{x}_i)}{\gamma} \right\} \cdot \frac{f(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i)} \tag{5}$$

One difference between the sum (5) and a straightforward Monte Carlo evaluation of (3), is that in (5) the number N is taken to be the number of accepted points rather than the total number of trial points. This accounts for the factor γ in (4).

Quasi-Monte Carlo cannot be effectively applied to the integral (3) because the integrand contains a characteristic function, corresponding to the decision to accept or reject. This will be demonstrated in computational examples below.

2.2.2 Smoothed Rejection

The Rejection Method is discontinuous as a result of the inherently discontinuous nature of the binary decision: accept or reject. This can be eliminated by allowing sample points to have associated "acceptance" weights as described below. A **Smoothed Rejection** method that retains the flexibility and advantages of the Rejection method but does not involve a discontinuous integrand will now be formulated. First, replace the integral in (3) by an equivalent smooth integral, as in

$$A = \gamma \int_D \int_0^1 \frac{f(\boldsymbol{x})}{h(\boldsymbol{x})} \chi_\delta(y, \gamma^{-1} h(\boldsymbol{x})) dy d\boldsymbol{x}. \tag{6}$$

in which the smooth function χ_δ satisfies

$$\int_0^1 \chi_\delta(y, \gamma^{-1} h(\boldsymbol{x})) dy = \gamma^{-1} h(\boldsymbol{x}). \tag{7}$$

The function χ_δ, which will also be referred to as the weight w, will be chosen to be piecewise linear below.

The corresponding Monte Carlo approximation corresponds to the following Smoothed Rejection procedure:

1. Select $\gamma \geq \sup_{\boldsymbol{x} \in D} h(\boldsymbol{x})$ and $0 < \delta \ll 1$.
2. Repeat until weight of accepted points is within one unit of N:
   (a) Sample $(\boldsymbol{x}_t, y_t)$ from $U([0,1]^{d+1})$.
   (b) If $y_t < \frac{h(\boldsymbol{x}_t)}{\gamma} - \frac{1}{2}\delta$, then acceptance weight is $w = 1$.
   Else if $y_t > \frac{h(\boldsymbol{x}_t)}{\gamma} + \frac{1}{2}\delta$, then $w = 0$.
   Else $w = \frac{1}{\delta}\left(\frac{h(\boldsymbol{x}_t)}{\gamma} + \frac{1}{2}\delta - y_t\right)$.

The density of accepted points $\boldsymbol{x}$ is $f_{accept}(\boldsymbol{x})$, given by (with $w = \chi_\delta$)

$$\begin{aligned} f_{accept}(\boldsymbol{x}) &= \frac{1 \cdot \int_0^1 w(\boldsymbol{x}, y)dy}{\int_{I^d}[1 \cdot \int_0^1 w(\boldsymbol{\xi}, y)dy]d\boldsymbol{\xi}} \\ &= \frac{h(\boldsymbol{x})/\gamma}{1/\gamma} \\ f_{accept}(\boldsymbol{x}) &= h(\boldsymbol{x}) \end{aligned}$$

which shows that the density function h is correctly sampled.

There is some extra work required by Smoothed Rejection as compared to ordinary Rejection. First, there is the work associated with assigning each sample point a weight, and storing and using these weights. This is typically small enough to be considered insignificant. Second, there is additional work which comes from the acceptance of points with weights of less than 1; i.e., in order to reach a total acceptance weight of size N, more than N evaluations of f are required. This extra work can be minimized by setting the constant δ sufficiently small. On the other hand, if δ is made too small, the advantages of continuity will be effectively lost.

2.2.3 Weighted Uniform Sampling

Another alternative to the Rejection method can be formulated by eliminating the accept/reject decision entirely, and instead assigning each sample point a weight equal to its acceptance probability given by $\gamma^{-1}h(\boldsymbol{x}_i)$.

A new Monte Carlo estimate is obtained as follows:

$$\check{A}_N = \frac{\sum_{i=1}^N \frac{f(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i)} \cdot \frac{h(\boldsymbol{x}_i)}{\gamma}}{\sum_{i=1}^N \frac{h(\boldsymbol{x}_i)}{\gamma}}$$

$$= \frac{\sum_{i=1}^{N} f(\boldsymbol{x}_i)}{\sum_{i=1}^{N} h(\boldsymbol{x}_i)}$$

This sum is essentially the ratio of two Crude Monte Carlo integration estimates – an estimate of the original function, f, in the numerator, and an estimate of the importance function, h, in the denominator. The resulting estimate is biased. Nevertheless, Powell and Swann [12] and Spanier and Maize [13] have shown that this bias is negligible in comparison with the rms-error as $N \rightarrow \infty$. Moreover the positive correlation between f and h, when h has been well chosen to closely mimic the behavior of f, provides substantial variance reduction.

The bias and rms-error are given as follows:

$$bias(\breve{A}_N) = \frac{A\, var(h)}{N} - \frac{cov(f,g)}{N} + O(N^{-3/2}) \quad (8)$$

$$rmse(\breve{A}_N) = \frac{\sqrt{var(f) + A^2\, var(h) - 2\, A\, cov(f,h)}}{\sqrt{N}} + O(N^{-3/4}) \quad (9)$$

One advantage of Weighted Uniform Sampling, as emphasized in [12], is that sample points need not be generated for the density h. A second advantage is that the weighted uniform sampling estimate is continuous (assuming that f and h are continuous) so that quasirandom sequences may be effectively used in the sum. This will be demonstrated in computational examples below.

On the other hand, a disadvantage of Weighted Uniform Sampling is that for problems with large regions of low importance many more function evaluations will be performed within such regions than if Importance Sampling were used instead. This is reflected in a greater amount of variance reduction in the latter case, and for such problems Smoothed Rejection may be preferable to Weighted Uniform Sampling.

2.2.4 Computational Examples

The Rejection, Smoothed Rejection and Weighted Uniform Sampling methods will now be compared on several examples. Consider the following integral:

$$A = \int_{I^d} f(\boldsymbol{x})\, d\boldsymbol{x}$$

where I^d is the unit cube in d dimensions, and evaluate the following four estimates of this integral:

$$\text{Crude:} \quad \hat{A}_N^{(1)} = \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{x}_i) \;\; , \;\; \boldsymbol{x}_i \sim U(I^d)$$

$$\text{W.U.S.:} \quad \hat{A}_N^{(2)} = \frac{\sum_{i=1}^N f(\boldsymbol{x}_i)}{\sum_{i=1}^N h(\boldsymbol{x}_i)} \ , \ \boldsymbol{x}_i \sim U(I^d)$$

$$\text{Rej Meth:} \quad \hat{A}_N^{(3)} = \frac{1}{N} \sum_{i=1}^N \frac{f(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i)} \ , \ \boldsymbol{x}_i \sim h(\boldsymbol{x}), \text{ accepted point.}$$

$$\text{Smooth Rej:} \quad \hat{A}_N^{(4)} = \frac{1}{N} \sum_{i=1}^{N^*} w_i \frac{f(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i)} \ , \ (\boldsymbol{x}_i, w_i) \sim h(\boldsymbol{x})$$

For the last estimate, $(\boldsymbol{x}_i, w_i)$ represents a weighted sample point with acceptance weight w_i, and N^* is chosen such that the sum of the acceptance weights is within one unit of N.

For a given value of N, take M samples of each of these estimates, denoted by $\hat{A}_{N,k}^{(j)}$ for $1 \le k \le M$. (For quasirandom, using successive points from a single quasirandom sequence.) Two measurements of the average integration error are the empirical root mean square error and the empirical standard deviation, defined as

$$rmse(\hat{A}_N^{(j)}) \approx \sqrt{\frac{1}{M} \sum_{k=1}^M (\hat{A}_{N,k}^{(j)} - A)^2} \ , \ j = 1, \ldots, 4 \tag{10}$$

$$sd(\hat{A}_N^{(j)}) \approx \sqrt{\frac{1}{M-1} \sum_{k=1}^M (\hat{A}_{N,k}^{(j)} - \overline{\hat{A}_N^{(j)}})^2} \ , \ j = 1, \ldots, 4 \tag{11}$$

in which $\overline{\hat{A}_N^{(j)}} = \frac{1}{M} \sum_{k=1}^M \hat{A}_{N,k}^{(j)}$. Note that the formula for *rmse* error uses the exact integral A, which is known for the examples below, whereas the *sd* error uses an empirical average.

Since successive estimates $\hat{A}_{N,k}^{(j)}$, $k = 1, \ldots, M$ are not independent for quasi-Monte Carlo, there is no theoretical basis for these error expressions. This is in contrast to standard Monte Carlo for which the Central Limit Theorem implies that the $\hat{A}_{N,k}^{(j)}$ come approximately from a Gaussian distribution with variance given by the *rmse*. Nevertheless, in the examples below these two error estimates give consistent results; in the examples below, they never differ by more than 2%.

Finally, the bias of Weighted Uniform Sampling proved to be insignificant, as expected, when using either pseudorandom or quasirandom sequences of points. (Note: In Examples 2 and 3, we use the base-2 Niederreiter quasirandom sequence [14].)

Example 2 *The first example is Monte Carlo integration over* $I^5 = [0,1]^5$ *of*

the function

$$f_2(\boldsymbol{x}) = \exp\left(\sum_{i=1}^{5} a_i\, x_i^2\, \frac{2+\sin\left(\sum_{j=1,j\neq i}^{5} x_j\right)}{2}\right)$$

using the positive definite importance function:

$$h_2(\boldsymbol{x}) = \frac{1}{\eta}\exp\left(\sum_{i=1}^{5} a_i\, x_i^2\right)$$

where $\boldsymbol{a} = (1, \frac{1}{2}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$ *and* $\eta = \int_{I^5} \exp\left(\sum_{i=1}^{5} a_i\, x_i^2\right) d\boldsymbol{x}$ *so that* h_2 *is normalized (i.e. its integral is one).*

The resulting *rmse* error for Example 2 using pseudorandom and quasirandom points are presented in Figures 2 and 3, respectively. Note that η is easily computed with high accuracy as the product of five one-dimensional integrals, using for example Gaussian quadrature.

Example 3 *The second example is Monte Carlo integration over* $I^7 = [0,1]^7$ *of the function*

$$f_3(x_1,\ldots,x_7) = e^{1-(\sin^2(\frac{\pi}{2}x_1)+\sin^2(\frac{\pi}{2}x_2)+\sin^2(\frac{\pi}{2}x_3))}\ \arcsin(\sin(1)+\frac{x_1+\cdots+x_7}{200})$$

using the positive definite importance function:

$$h_3(x_1,\ldots,x_7) = \frac{1}{\eta}e^{1-(\sin^2(\frac{\pi}{2}x_1)+\sin^2(\frac{\pi}{2}x_2)+\sin^2(\frac{\pi}{2}x_3))}$$

where η *is*

$$\eta = \int_{I^7} h_3(x_1,\ldots,x_7)\, dx_1 \ldots dx_7 = e \cdot \left(\int_0^1 e^{-\sin^2(\frac{\pi}{2}x)}\, dx\right)^3$$

which is easily approximated to high accuracy as a one-dimensional integral.

The resulting *rmse* errors for Example 3 using pseudorandom and quasirandom points are presented in Figures 4 and 5, respectively.

These computational examples show that quasi-Monte Carlo works well for the Smoothed Rejection Method, but even better for Weighted Uniform Sampling.

The results confirm that Smoothed Rejection is an improvement over Ordinary Rejection when using quasirandom sequences. Furthermore, Weighted Uniform Sampling produces results that are far superior to those obtained using Importance Sampling with the Rejection Method, even with smoothing. We

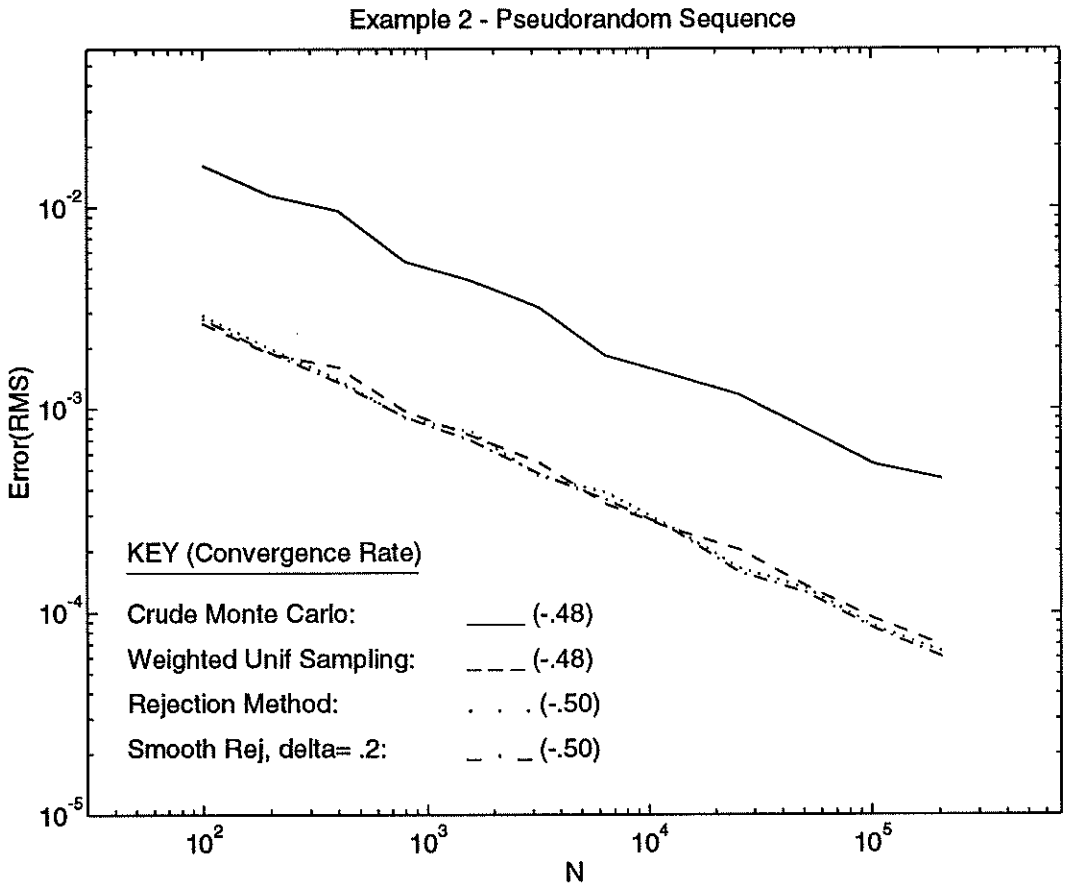


Figure 2: Pseudorandom Results, 70 runs, Example 2.

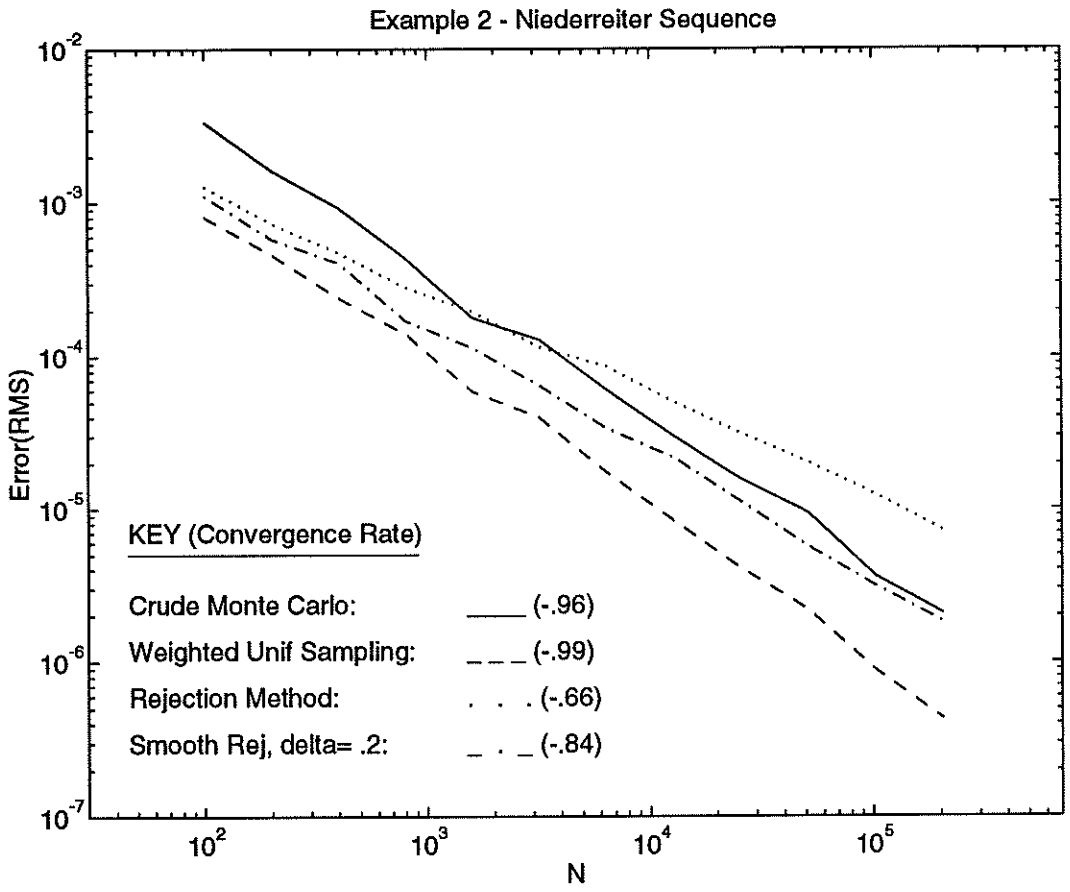


Figure 3: Quasi-Random Results, 70 runs, Example 2.

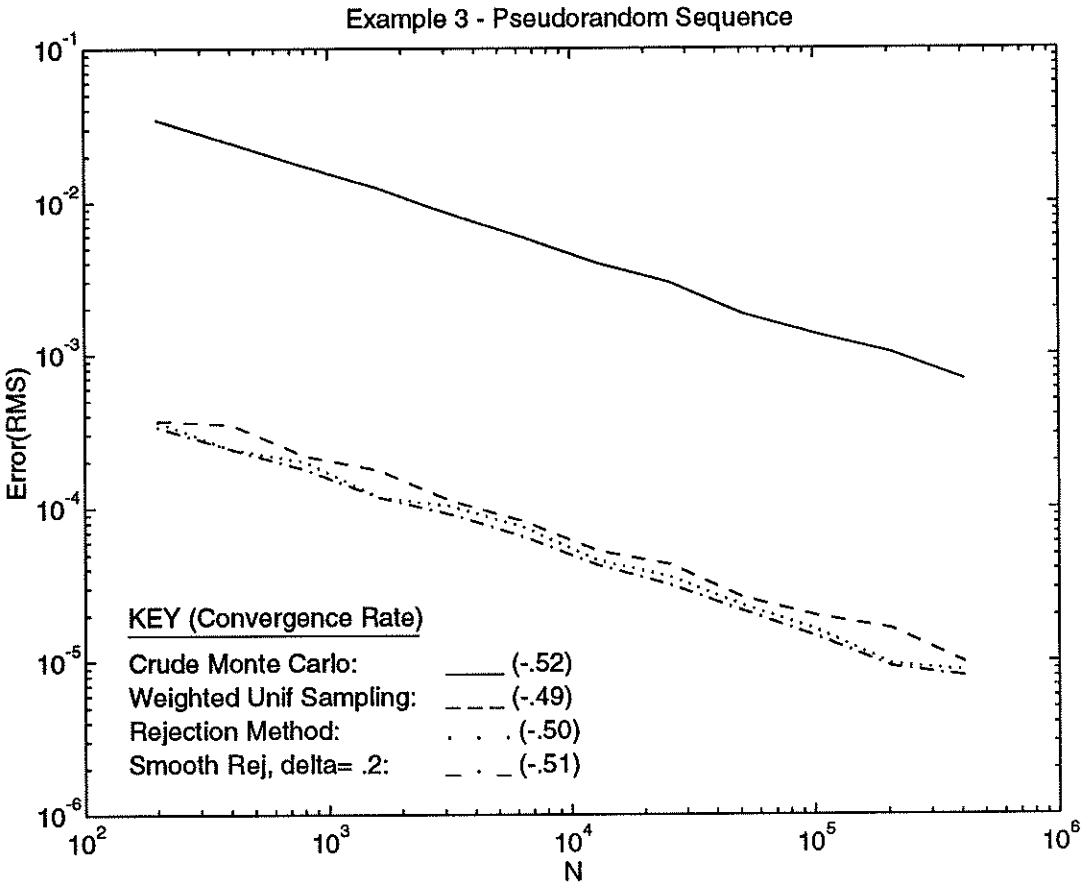


Figure 4: Pseudorandom Results, 75 runs, Example 3.

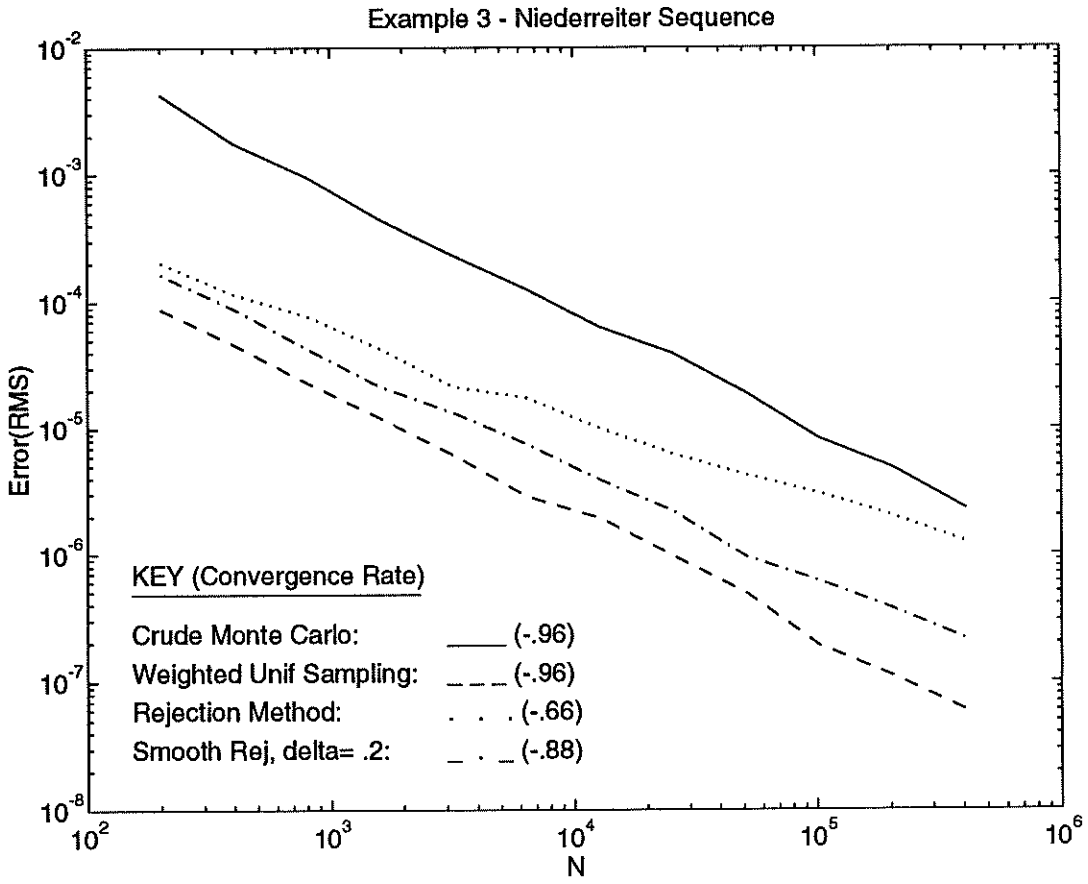


Figure 5: Quasi-Random Results, 75 runs, Example 3.

believe the reason for this to be that although the Smoothed Rejection method involves a continuous integrand, there is still a relatively sharp slope in acceptance weights in the transition from acceptance to rejection regions. This leads to higher variation and poorer quasirandom performance, in general, than for Weighted Uniform Sampling, where the sharp transition is completely eliminated. Another, admittedly minor, advantage to Weighted Uniform Sampling is that the extra 'decision' variable y is no longer needed, reducing the effective dimension by one.

Nevertheless, Smoothed Rejection may be superior to Weighted Uniform Sampling for problems in which there are large regions of low importance, as discussed at the end of Section 2.2.3.

3 Reduction of Dimensions

Unlike standard Monte Carlo which is independent of the number of dimensions, quasirandom methods tend to become less effective as the number of dimensions in the integrand increases [2, 4, 15, 16], although much less so than numerical quadrature methods such as the trapezoidal rule. Therefore, while numerical quadrature methods are best for low dimensional integrals, and pseudorandom Monte Carlo appears to be best for very high dimensional integrals, quasirandom methods can often be optimal for moderate dimensional integrals. Our experience indicates that quasirandom methods, with N in the range of 10^4 to 10^6, are effective for integrals of dimension four to thirty.

In this section, we discuss methods to reduce the effective dimension of a quasi-Monte Carlo integration problem.

3.1 A Simple Example of Dimension Reduction

The following example shows that quasirandom integration error may be reduced by rearrangement of the variables so that the principal variations of the integrand occur over the lower dimensions:

Example 4 *Compare the results of approximating these two equal integrals using pseudorandom and Halton points:*

$$A = \int \cdots \int_{I^d} 5 \cos\left(4(x_4 + x_5 + x_6) + \frac{(x_1 + x_2 + x_3)}{5}\right) dx_1 \ldots dx_6$$

$$B = \int \cdots \int_{I^d} 5 \cos\left(4(x_1 + x_2 + x_3) + \frac{(x_4 + x_5 + x_6)}{5}\right) dx_1 \ldots dx_6$$

These are equivalent, but the first integrand varies more as a function of the higher three dimensions while the second varies more as a function of the lower three dimensions. Therefore we should expect that quasirandom estimates will be more effective estimating B than A. On the other hand, using pseudorandom points, there should be little difference between the two cases. The results, shown in Figure 6, confirm this.

3.2 Dimension Reduction for the Feynman-Kac Formula

The Feynman-Kac Formula provides a connection between linear parabolic differential equations and stochastic path integrals, which is similar to the method of characteristics for solving hyperbolic differential equations. In the simple case when the stochastic paths are Brownian motion, a modification of the discretization of the Brownian path allows one to concentrate most of the variance into the first few dimensions, which significantly improves the accuracy of quasi-Monte Carlo integration.

Consider the following initial value problem:

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(x,t) + v(x,t)\,u(x,t) \quad , \quad t \geq 0 \tag{12}$$

with initial condition $u(x,0) = f(x)$. Its solution is given by the Feynman-Kac formula as

$$u(x,t) = E_{x,0}\left[f(\xi(t))\, e^{\int_0^t v(\xi(r),t-r)\,dr}\right] \tag{13}$$

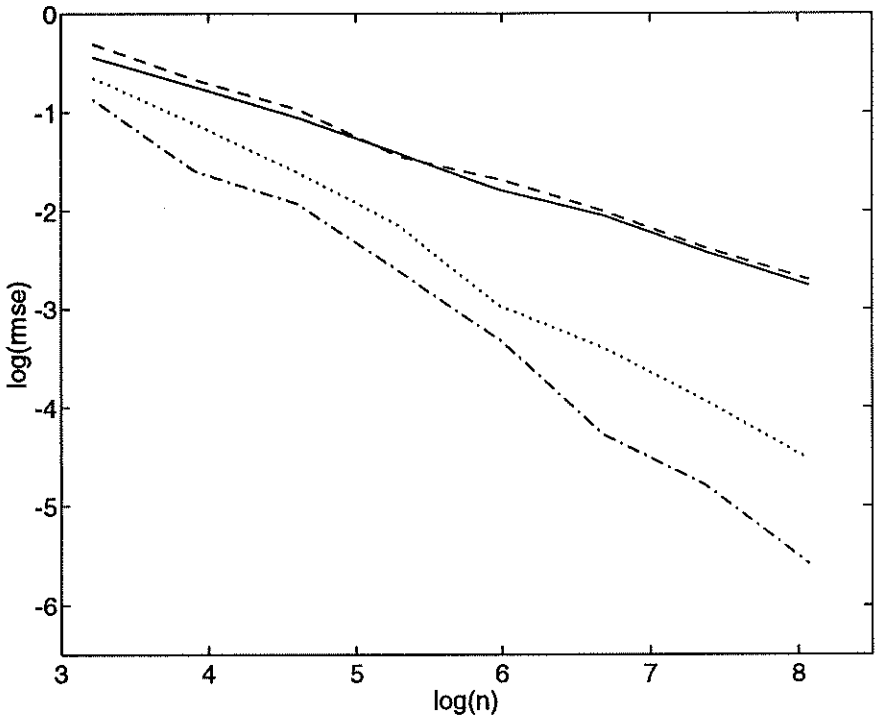
where ξ represents a Wiener path starting at $\xi(0) = x$.

This solution can be approximated by discretizing time, generating Wiener process sample paths, and averaging over the set of sample paths to obtain Monte Carlo estimates of the expectation above. The integral in the exponent is itself approximated for each sample path using the Trapezoidal rule.

3.2.1 Standard Discretization

The **Standard Discretization** of the Wiener process is defined as follows:

1. Choose (x,t) at which the solution is to be approximated.
2. Choose the number of equal time steps m and set $\Delta t = \frac{t}{m}$ and $t_i = i\,\Delta t$ for $i = 0, \ldots, m$.
3. Generate Gaussian independent variables, $\Delta W^{(i)}$ from the distribution $\mathcal{N}(0, \Delta t)$ (the normal distribution with mean 0 and variance Δt), for $i = 1, \ldots, m$.



Integrand	Sequence	Slope of *rmse* vs. N
A	pseudorandom	-0.479 (solid)
B	pseudorandom	-0.490 (dashed)
A	Halton	-0.814 (dotted)
B	Halton	-0.974 (dot-dash)

Figure 6: Convergence Results and Log-Log Plot for Example 4, 100 trials.

4. The Wiener path is sampled exactly at each t_i as the following sum:

$$\xi^{(i)} = x + \sum_{j=1}^{i} \Delta W^{(j)} \tag{14}$$

The sample paths are exact at the discrete times t_i, but an approximation is involved when the path integral of v is estimated. For example, using the Trapezoidal rule the Wiener path integral is approximated as

$$\int_0^t v(\xi(r), t-r)\, dr \approx \frac{\Delta t}{2}\left[v(x,t) + 2\sum_{i=1}^{m-1} v(\xi^{(i)}, t-t_i) + v(\xi^{(m)}, 0)\right] = \tilde{I} \tag{15}$$

The nondifferentiability of the paths leads to an error (bias) of order $O(\sqrt{\Delta t})$ for this quadrature rule. In the computations presented below, however, Δt is sufficiently small so that discretization errors are negligible in comparison to statistical errors.

Using the sample paths above, the Monte Carlo estimate of the Feynman-Kac path integral is

$$\hat{u}_N(x,t) = \frac{1}{N}\sum_{k=1}^{N} f(\xi_k^{(m)})\, e^{\tilde{I}_k} \tag{16}$$

in which $\tilde{I}_k$ and $\xi_k^{(m)}$ for $i \le k \le N$ are N realizations of $\tilde{I}$ and $\xi^{(m)}$. If the $\xi_k^{(m)}$ are normally distributed, as described above, then

$$E(\hat{u}_N(x,t)) = u(x,t) + \text{bias error}$$

in which the bias error comes from the quadrature rules and is insignificant.

For application of quasirandom Monte Carlo, this average is expressed as a multidimensional integral. First, each normal variable $\Delta W^{(j)}$ can be represented as a transformation G of a uniformly distributed variable x_j by $\Delta W^{(j)} = G(x_j)$. Then define

$$\begin{aligned} \tilde{f}(x_1, x_2, \ldots, x_m) &= f(\xi^{(m)}) \\ \tilde{I}(x_1, x_2, \ldots, x_m) &= \tilde{I} \end{aligned}$$

in which $\xi^{(i)} = x + \sum_{j=1}^{i} G(x_j)$, and x_j are uniformly distributed, e.g. pseudo-random or quasirandom points. The expectation of the Monte Carlo estimate in (16) can then be written as the following integral:

$$E(\hat{u}_N(x,t)) = \int_0^1 \cdots \int_0^1 \tilde{f}(x_1, \ldots, x_m) e^{\tilde{I}(x_1,\ldots,x_m)}\, dx_1 \ldots dx_m$$

The Monte Carlo estimate of this integral is

$$\tilde{u}_N(x,t)) = \frac{1}{N}\sum_{k=1}^{N} \tilde{f}(x_1^k, \ldots, x_m^k) e^{\tilde{I}(x_1^k,\ldots,x_m^k)} \tag{17}$$

The transformation G used here is that of Marsaglia [17], which is continuous and monotonic. By contrast, Box-Muller, which is discontinuous, gives poorer results for quasirandom sequences [3].

3.2.2 Alternative Discretization Method

When the number of dimensions, m, is moderately large, an alternative discretization method leads to significant improvements in the quasirandom integration estimates by concentrating most of the variance into the lowest dimensions.

Assume that m is a power of two, and define the alternative discretization as follows:

Alternative Discretization:

1. Choose (x, t) at which the solution is to be approximated.
2. Choose the number of equal time steps $m = 2^p$. Then let $\Delta t = \frac{t}{m}$ and $t_i = i\,\Delta t$ for $i = 0, \ldots, m$.
3. Generate Gaussian variables, $\Delta W^{(0)}$ distributed according to $\mathcal{N}(0, t)$ and $\Delta W^{(i)}$ distributed according to $\mathcal{N}(0, t/\alpha)$, for $i = 1, \ldots, m$, in which $\alpha = 4 \cdot 2^{int(\log_2 i)}$.
4. The Wiener path $\xi^{(i)}$ is sampled exactly at each t_i as follows:

$$\begin{aligned}
\xi^{(0)} &= x \\
\xi^{(m)} &= \xi^{(0)} + \Delta W^{(0)} \\
\xi^{(\frac{m}{2})} &= \frac{\xi^{(0)} + \xi^{(m)}}{2} + \Delta W^{(1)} \\
\xi^{(\frac{m}{4})} &= \frac{\xi^{(0)} + \xi^{(\frac{m}{2})}}{2} + \Delta W^{(2)} \\
\xi^{(\frac{3m}{4})} &= \frac{\xi^{(\frac{m}{2})} + \xi^{(m)}}{2} + \Delta W^{(3)} \\
&\vdots
\end{aligned} \tag{18}$$

For this discretization method, the first step is directly from 0 to t. Then the intermediate steps are filled in by taking successive subdivisions of the time intervals into halves. Each new intermediate path position is determined using the following rule:

$$\xi^{(j)} = \frac{\xi^{(j_1)} + \xi^{(j_2)}}{2} + \Delta W \;\;,\;\; \Delta W \in \mathcal{N}(0, \frac{t_{j_2} - t_{j_1}}{4}) \tag{19}$$

	Variance	
	Standard	Alternative
x_1	$1.0000t$	$3.1875t$
x_2	$0.8750t$	$1.6875t$
x_3	$0.7500t$	$.1875t$
x_4	$0.6250t$	$.1875t$
x_5	$0.5000t$	$.0625t$
x_6	$0.3750t$	$.0625t$
x_7	$0.2500t$	$.0625t$
x_8	$0.1250t$	$.0625t$

Table 1: Variances for Standard and Alternative Discretization Schemes, $m = 8$

where j_1, j_2 are the indices of the nearest prior and later time steps, respectively, for which the positions have already been determined. This representation of the Wiener path in terms of past and future positions is referred to as the Brownian bridge (also called a tied-down Wiener process).

The representation (18) can be conveniently rewritten as follows:

$$\begin{aligned}
\xi^{(0)} &= x \\
\xi^{(m)} &= x + \Delta W^{(0)} \\
\xi^{(\frac{m}{2})} &= x + \frac{\Delta W^{(0)}}{2} + \Delta W^{(1)} \\
\xi^{(\frac{m}{4})} &= x + \frac{\Delta W^{(0)}}{4} + \frac{\Delta W^{(1)}}{2} + \Delta W^{(2)} \\
\xi^{(\frac{3m}{4})} &= x + \frac{3\,\Delta W^{(0)}}{4} + \frac{\Delta W^{(1)}}{2} + \Delta W^{(3)} \\
&\vdots
\end{aligned} \tag{20}$$

The variance contributed by each dimension to the Wiener sample paths is compared for the Standard and the Alternative Discretizations in Table 1 for the case $m = 8$. Note that variance due to x_k slowly decreases for the Standard Discretization, since early values of the Wiener process affect more parts of the path than later values, but that this decay is much more rapid for the Alternative Discretization.

Figure 7 is a graphical comparison of the Standard and Alternative discretization points for $m = 4$, showing how more of the variance is concentrated in the lower dimensions for the Alternative method. The dotted horizontal lines demark one standard deviation. The solid lines show the actual discretized stochastic paths. In the upper figure, the dash/dot line indicates the additive nature of the successive steps. In the lower figure, the dashed line shows the ef-

fects of the large first step on all of the intermediate positions, while the dash/dot lines show the effects of the second step on intermediate positions. Note that the two plots show the same resultant paths for convenience only. There is no connection between the particular path generated by the same pseudorandom or quasirandom point using each of the two different methods. There is only an equivalence in terms of the expectations over all such paths.

3.2.3 Computational Example for the Feynman-Kac Formula

This example demonstrates the accuracy of Monte Carlo estimates of the solution of a simple linear parabolic differential equation using (16) at selected values of x and a fixed (small) time t. A comparison will be made of pseudorandom and quasirandom sequences with either the Standard discretization (14), or the Alternate discretization (19).

Example 5 *Consider the following linear parabolic differential equation:*

$$\frac{\partial f}{\partial t}(x,t) = \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x,t) + \left(\frac{1}{t+1} + \frac{1}{x^2+1} - \frac{4x^2}{(x^2+1)^2}\right)\frac{\partial f}{\partial x}(x,t)$$

with initial condition $f(x,0) = \frac{1}{x^2+1}$. *The exact solution is solution:* $f(x,t) = \frac{t+1}{x^2+1}$.

Estimates are computed at a fixed (small) time, T, *and at eight equally spaced positions between* $x = -3$ *and* $x = 3$ *using a discretization of time into* m *equal steps and a sample size of* N. *The estimates are then compared with the exact solution using the following* L^2 *measure of error:*

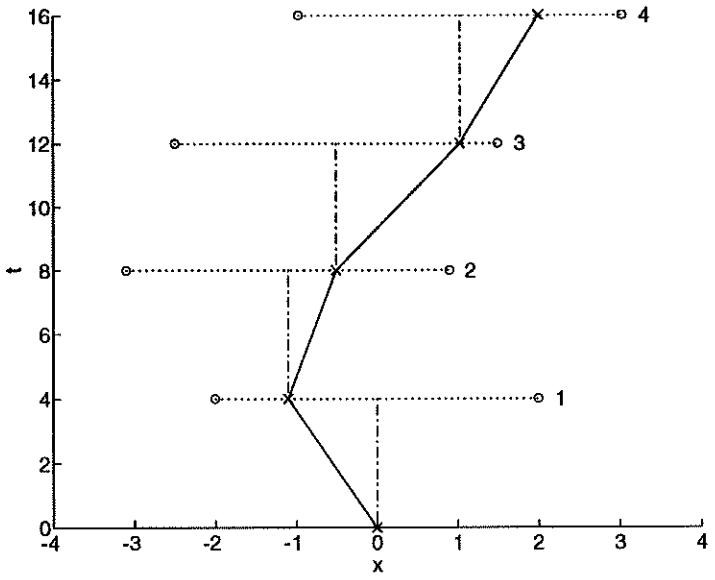
$$L2e = \sqrt{\frac{1}{8}\sum_{k=0}^{7}\left[f(x_k,T) - \hat{f}_N(x_k,T)\right]^2}$$

where $x_k = -3 + 6\,k/7$.

T is selected to be very small so that discretization errors are insignificant relative to the Monte Carlo errors.

A measure of the rate of decrease of the values of $L2e$ is obtained by a least squares fit of the function $cN^{-\alpha}$ to the calculated values of $L2e$. The resulting values of α for pseudorandom and quasirandom (Halton was used here) points used in both the standard and alternative discretizations are presented along with log-log plots of the data in Figures 8–10 for three values of T with equal time increments in each case; i.e., $T = .02$ ($m = 8$), $T = .04$ ($m = 16$) and $T = .08$ ($m = 32$). The results show that the convergence rate for pseudorandom is independent of T and of the discretization scheme. On the

Standard Scheme, 4 Steps



Alternate Scheme, 4 Steps

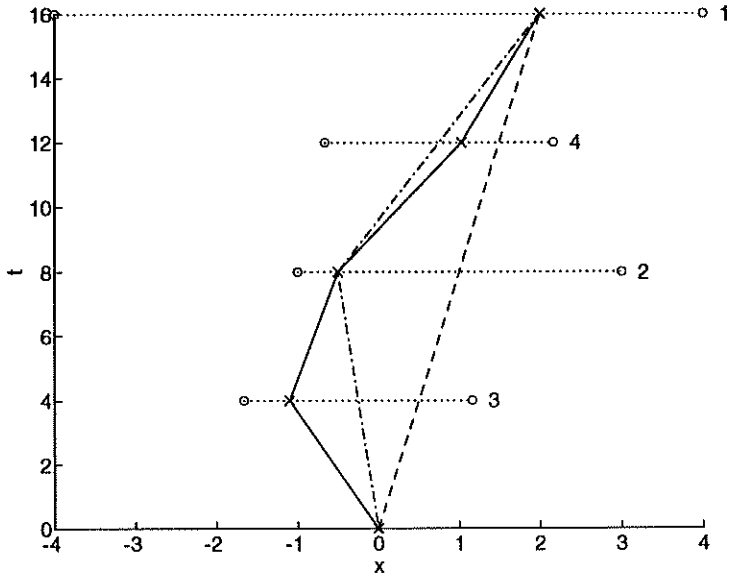


Figure 7: Graphical Comparison of Discretization Schemes.

other hand, the performance of quasirandom Monte Carlo rapidly degrades in the standard discretization scheme, but is maintained at α nearly equal to 1 in the alternative discretization. It is important to notice that the quasirandom improvement from the alternate scheme grows as the dimension increases. This is expected, as is the fact that at 32 dimensions standard quasirandom is just barely better than pseudorandom.

For standard Monte Carlo using pseudorandom sequences there is no difference between the two discretizations, since the total variance of the two methods is the same. On the other hand, the Alternative discretization method provides significantly more accurate results than the Standard discretization, when using quasirandom sequences in the Feynman-Kac formula. This improvement is due to reduction of the effective dimension of the problem.

4 Conclusions

The analysis and computations presented above demonstrate the need to modify standard Monte Carlo methods for effective implementation of quasirandom sequences. The advantages of quasirandom sequences over pseudorandom sequences are lost for problems with discontinuous integrands or large dimension. These advantages can be regained, at least for some problems, by smoothing or dimension reduction.

This has been successfully implemented for the Rejection method through a Smoothed Rejection method, and for the evaluation of Feynman-Kac integrals through an alternative discretization. In a forthcoming paper, one of the authors (BM) will show how quasirandom sequences may be also used in Diffusion Monte Carlo for computation of the ground state energy of various systems. On the other hand, we have not yet been successful in applying quasirandom methods to the Metropolis algorithm.

We consider the results presented in these examples to be an important step forward in the continuing and promising application of quasirandom sequences to Monte Carlo methods. While quasirandom sequences will probably not be useful in all Monte Carlo methods, we expect that a wide range of effective applications may be developed.

References

[1] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods.* SIAM, Philadelphia, 1992.

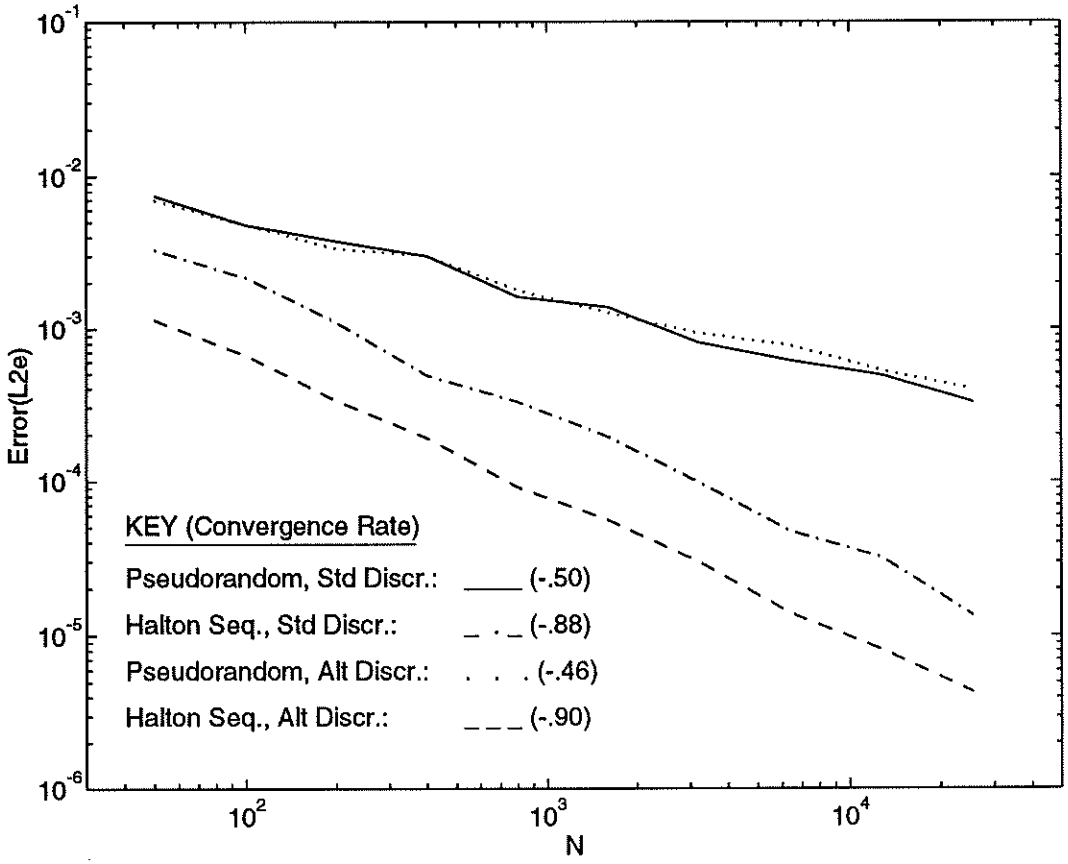


Figure 8: Log-Log Plot for Example 5, 75 runs, $T = 0.02$, $m = 8$.

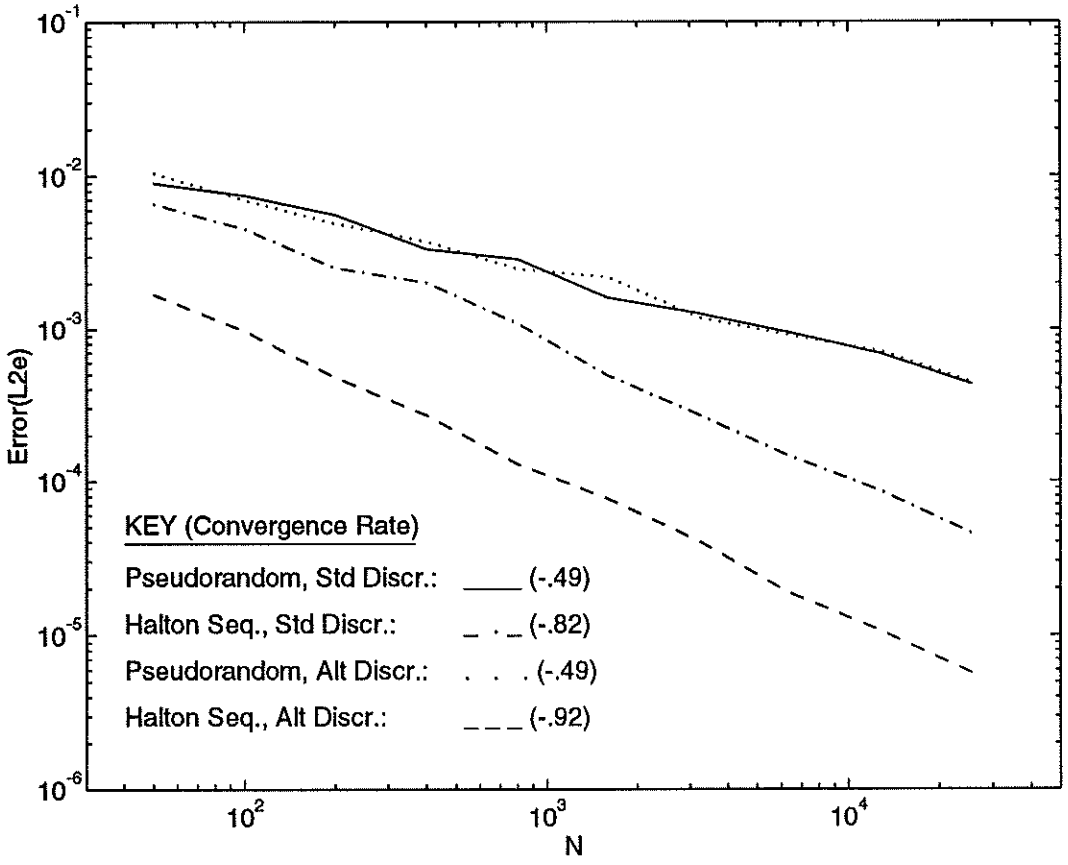


Figure 9: Log-Log Plot for Example 5, 75 runs, $T = 0.04$, $m = 16$.

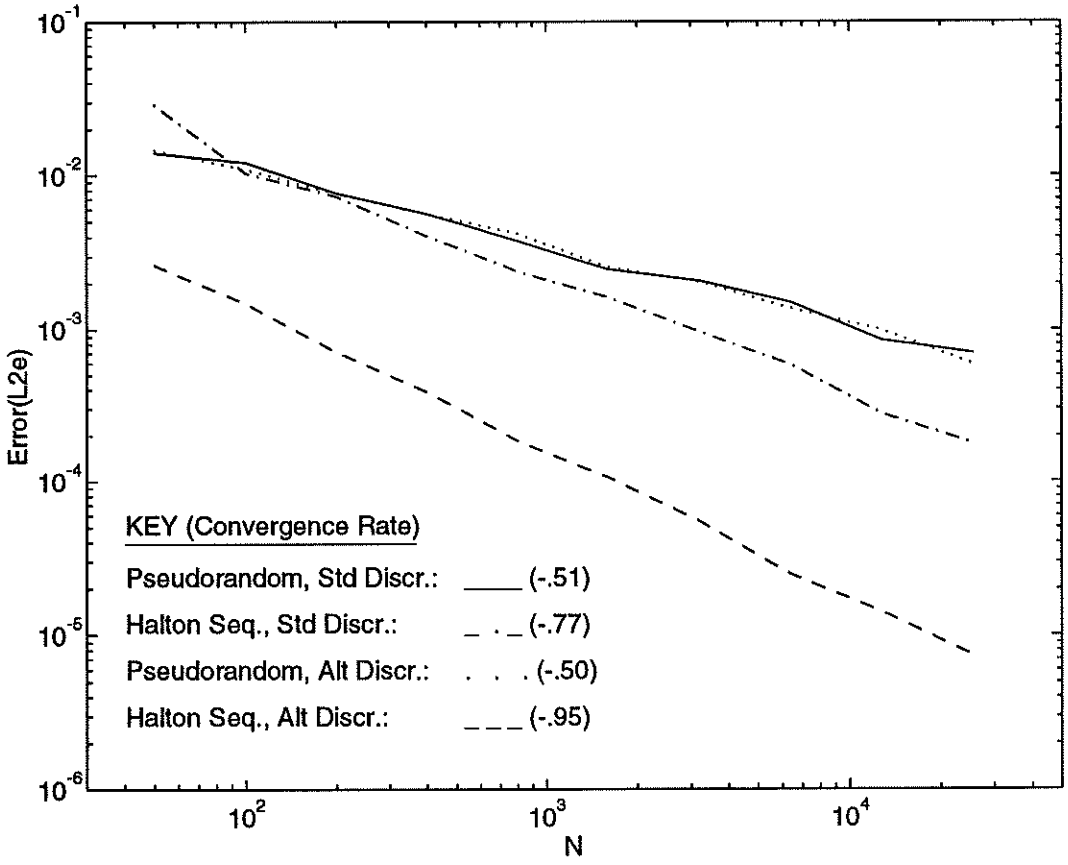


Figure 10: Log-Log Plot for Example 5, 75 runs, $T = 0.08$, $m = 32$.

[2] W. Morokoff and R.E. Caflisch. Quasi-Monte Carlo integration. *J. Comp. Phys.*, 1994.

[3] W. Morokoff and R.E. Caflisch. A Quasi-Monte Carlo approach to particle simulation of the heat equation. *SIAM Journal on Numerical Analysis*, 30:1558–1573, 1993.

[4] W. Morokoff and R.E. Caflisch. Quasi-random sequences and their discrepancies. *SIAM J. Sci. Stat. Computing*, to appear, 1993.

[5] C. Lecot. A quasi-Monte Carlo method for the Boltzmann equation. *Mathematics of Computation*, 56(194):621–644, April 1991.

[6] B.V. Shuhman. Application of quasirandom points for simulation of gamma radiation transfer. *Progress in Nuclear Energy*, 24:89–95, 1990.

[7] V. Adlakha. An empirical evaluation of anithetic variates and quasirandom points for simulating stochastic networks. *Simulation*, 58(1):23–31, January 1992.

[8] B. Moskowitz. *Application of Quasi-Random Sequences to Monte Carlo Methods.* PhD thesis, University of California, Los Angeles, September 1993.

[9] J.H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2:84–90, 1960.

[10] W.H. Press and S.A. Teukolsky. Quasi- (that is, sub-) random numbers. *Computers in Physics*, pages 76–79, Nov/Dec 1988.

[11] M. Berblinger and C. Schlier. Monte Carlo integration with quasi-random numbers: Some experience. *Computer Physics Communications*, 66:157–166, 1991.

[12] M.J.D. Powell and J. Swann. Weighted uniform sampling – a Monte Carlo technique for reducing variance. *J.Inst.Maths Applics*, 2:228–236, 1966.

[13] J. Spanier and E.H. Maize. Quasi-random methods for estimating integrals using relatively small samples. *SIAM Review*, 36:18–44, 1994.

[14] Paul Bratley, Bennett L. Fox, and Harald Niederreiter. Implementation and tests of low-discrepancy sequences. *ACM Transactions on Modeling and Computer Simulation*, 2(3):195–213, July 1992.

[15] I.M. Sobol'. Quasi-Monte Carlo methods. *Progress in Nuclear Energy*, 24:55–61, 1990.

[16] Y.L. Levitan, N.I. Markovich, S.G. Rozin, and I.M. Sobol′. Short communications on quasirandom sequences for numerical computations. *Zh. výchisl. Mat. mat. Fiz.*, 28(5):755–759, 1988.

[17] G. Marsaglia. Normal (Gaussian) random variables for supercomputers. *The Journal of Supercomputing*, 5:49–55, 1991.