

4

Very High-Dimensional Integration and Mathematical Finance

The valuation of financial instruments often requires the calculation of very high-dimensional integrals. Dimensions of 360 and higher are not unusual. What do the complexity results of Chapter 3 suggest about the numerical calculation of such integrals? We recall several results of that chapter, for the convenience of the reader.

In the worst case setting for integrands of total smoothness r , the complexity of guaranteeing an answer to within ε is of order $(1/\varepsilon)^{d/r}$, where d is the dimension of the domain of integration. Hence if $r > 0$, the problem is intractable in dimension; if $r = 0$, it is unsolvable. To make the problem computationally feasible, we must either weaken the worst case assurance or change the class of inputs.

First, we describe replacing the worst case guarantee by a stochastic assurance. In the randomized setting, we settle for the expected cost of an approximation to be at most ε , where the expectation is with respect to the distribution determining the approximation. Suppose that we use information consisting of sample points chosen as independent uniformly distributed random variables from $[0, 1]^d$. If $r = 0$, the Monte Carlo algorithm is optimal, and its complexity is of order $1/\varepsilon^2$. If $r > 0$, the complexity is of order $1/\varepsilon^\sigma$, where $\sigma < 2$. Hence in the randomized setting, integration is strongly tractable. Because of its $1/\varepsilon^2$ cost, Monte Carlo is widely used in many applications.

Can we do better than $1/\varepsilon^2$? The answer is “yes” if we’re willing to shift to the average case setting. In this setting, we guarantee that the expected error is at most ε , where the expectation is now with respect to the measure on the space of integrands. Since the average case setting is *deterministic*, we would like a procedure for obtaining the sample points. The experimental design problem of selecting sample points with good average cost has been open since Sacks & Ylvisaker [1966].

It was settled by Woźniakowski [1991], for the case of the Wiener sheet measure on the space of continuous functions on $[0, 1]^d$. Woźniakowski showed that the integrand should be evaluated at points that are related to low discrepancy points. Since discrepancy theory has been extensively studied in number theory, one could now draw on a very large existing literature.

Rather than attempting to select points from a uniform distribution, why not simply select uniform points deterministically? Uniformity is *not* sufficient, as the example of regular grids shows. Papageorgiou & Wasilkowski [1990] showed that the cost of any quadrature rule that used regular grid points is exponential in d .

What we desire is a “small” set of points in d dimensions, which is uniform. By *uniform*, we mean that the fraction of points lying within any rectangular subregion (with sides parallel to the coordinate axes) of the d -dimensional unit cube is as close as possible to the volume of that subregion. The *discrepancy* of a sequence of points is a measure of its deviation from uniformity; we therefore desire *low discrepancy*. The discrepancy can be measured in various ways; we confine ourselves here to L_2 and L_∞ discrepancy.

For $x = [x_1, \dots, x_d] \in [0, 1]^d$, define $[0, x] = [0, x_1] \times \dots \times [0, x_d]$. Let $\chi_{[0, x]}$ be the characteristic function of $[0, x]$. For $t_1, \dots, t_n \in [0, 1]^d$, define

$$R_n(x, t_1, \dots, t_n) = \frac{1}{n} \sum_{i=1}^n \chi_{[0, x]}(t_i) - \prod_{i=1}^n t_i.$$

The L_2 and L_∞ *discrepancy* are respectively defined as

$$D_{n,d}^{(2)}(t_1, \dots, t_n) \equiv D_{n,d}^{(2)} = \left(\int_{[0,1]^d} R_n(x, t_1, \dots, t_n)^2 dx \right)^{1/2}$$

and

$$D_{n,d}^{(\infty)}(t_1, \dots, t_n) \equiv D_{n,d}^{(\infty)} = \sup_{x \in [0,1]^d} |R_n(x, t_1, \dots, t_n)|.$$

For historical reasons, $D_{n,d}^{(\infty)}$ is usually written $D_{n,d}^*$, and we'll follow that usage here. See the monographs by Niederreiter [1992] and Drmota & Tichy [1997] for extensive material on discrepancy.

We next relate the average case L_2 integration error to the L_2 discrepancy. Let

$$I(f) = \int_{[0,1]^d} f(x) dx,$$

and let

$$U(f) = \frac{1}{n} \sum_{i=1}^n f(t_i),$$

for arbitrary t_1, \dots, t_n . Let w be the classical Wiener sheet measure w on the space $C([0, 1]^d)$ of continuous functions on $[0, 1]^d$. Woźniakowski [1991] proved that the L_2 integration error is given by

$$\left(\int_{C([0,1]^d)} [I(f) - U(f)]^2 w(df) \right)^{1/2} = D_{n,d}^{(2)}(z_1, \dots, z_n),$$

where

$$z_i = \mathbf{1} - t_i \quad \text{for } 1 \leq i \leq n. \quad (4.1)$$

Hence to minimize the average case L_2 integration error, we choose the sequence $t_1, \dots, t_n \in [0, 1]^d$ such that $D_{n,d}^{(2)}(z_1, \dots, z_n)$ is as small as possible. Roth ([1954], [1980]) proved that

$$\inf_{z_1, \dots, z_n \in [0,1]^d} D_{n,d}^{(2)}(z_1, \dots, z_n) = \Theta \left(n^{-1} (\log n)^{(d-1)/2} \right). \quad (4.2)$$

Furthermore, the optimal z_i are related to Hammersley points.

We turn to the L_∞ discrepancy $D_{n,d}^*$. For our present purposes, the most important property of $D_{n,d}^*$ is given by the celebrated Koksma-Hlawka inequality (Niederreiter [1992], p. 20): if f has bounded variation $V(f)$ on $[0, 1]^d$ (in the sense of Hardy and Krause), then for any points $t_1, \dots, t_n \in [0, 1]^d$, we have

$$|I(f) - U(f)| \leq V(f) D_{n,d}^*(t_1, \dots, t_n). \quad (4.3)$$

We remark that $V(f)$ is finite if f has one derivative in each coordinate direction; however, the calculation of $V(f)$ for large d may be an onerous task, as its definition contains $2^d - 1$ terms.

To minimize the upper bound on the integration error in (4.3), we want to choose t_1, \dots, t_n so that $D_{n,d}^*(t_1, \dots, t_n)$ is as small as possible. How small can that be? It is believed that

$$D_{n,d}^*(t_1, \dots, t_n) \geq B_d n^{-1} (\log n)^{d-1} \quad \forall n \geq 2,$$

where B_d depends only on the dimension d . Furthermore, there exist points $t_1^*, \dots, t_n^* \in [0, 1]^d$ such that

$$D_{n,d}^*(t_1^*, \dots, t_n^*) = O \left(n^{-1} (\log n)^{d-1} \right).$$

More generally, any sequence t_1, \dots, t_n satisfying

$$D_{n,d}^*(t_1, \dots, t_n) = O(n^{-1}(\log n)^d)$$

is said to be a *low discrepancy sequence* (LDS); see Niederreiter [1992] and Tezuka [1995]. Many examples of LDS are known, including Halton, Sobol', Hammersley, Faure, generalized Faure, and generalized Niederreiter sequences. Although many LDS share the same asymptotic behavior, their observed performance on important practical problems can differ widely, as we shall see below.

We compare and contrast the Woźniakowski and Koksma-Hlawka theorems. Woźniakowski's theorem states that to get optimal integration error averaged over a space of integrands, one should use points having minimal L_2 discrepancy. The Koksma-Hlawka theorem states that to minimize the upper bound on the integration error for any integrand of bounded variation, one should use points having minimal L_∞ discrepancy; that is, an LDS.

The motivation for looking at average behavior was to see if the Monte Carlo algorithm could be beaten. Recall that the expected error of the Monte Carlo algorithm is proportional to $n^{-1/2}$, while we've seen that there are deterministic sequences whose expected or worst case error is proportional to n^{-1} times a polylog factor. To fix ideas, let's focus on comparing $n^{-1/2}$ with $n^{-1}(\log n)^d$. What can we conclude?

- $n^{-1}(\log n)^d$ is asymptotically (in n) smaller than $n^{-1/2}$. This is why low discrepancy methods have long been of interest.
- For applications such as mathematical finance, n is modest in size while d is in the hundreds or thousands. Therefore, the asymptotic implications don't apply.

For d large and n fixed, the factor $(\log n)^d$ looks ominous. Therefore, leading experts believed that LDS should not be used for high-dimensional problems. For example, $d = 12$ was considered high by Bratley *et al.* [1992].

We'll estimate $\int_{[0,1]^d} f(x) dx$ by $n^{-1} \sum_{i=1}^n f(t_i)$. If the t_i are chosen at random, this is the Monte Carlo (MC) algorithm. If the t_i are chosen from a (deterministic) LDS, this is a quasi-Monte Carlo (QMC) algorithm. We believe that the term "quasi-Monte Carlo algorithm" is somewhat misleading because these algorithms are completely deterministic; however, since this term is so widely used, we will follow the general usage.

In 1992, one of us (JFT) and a then Ph.D. student (S. Paskov) followed a suggestion of I. Vanderhoof to test the efficacy of QMC algorithms for

the valuation of financial derivatives. A *financial derivative* is a financial instrument whose value is derived from an underlying asset. At the time of this writing (1998) it is estimated that there are some ten to twenty trillion dollars in assets covered by financial derivatives. The valuation of financial derivatives is therefore of considerable interest to the financial community and a fascinating problem for the computational scientist.

The valuation of financial derivatives often requires very high-dimensional integration. Boyle [1977] suggested the use of MC, which became a major computational tool in the financial community.

The model problem chosen to compare the efficacy of QMC with MC was a 30-year Collateralized Mortgage Obligation (CMO). The particular CMO chosen required the computation of ten 360-dimensional integrals (360 being the number of months in thirty years). Since the model problem required some 10^5 floating point operations per sample point, it was important to use as few points as possible. See Paskov [1997] for a description of the CMO.

Software construction and testing of QMC methods for financial computations were begun in Fall, 1992. The first published announcement about the empirical results was in Traub & Woźniakowski [1994]. Details were reported in Paskov & Traub [1995], Papageorgiou & Traub [1996], and Paskov [1997]. For a popular account, see Cipra [1996].

We mention here a few of the empirical results from Paskov & Traub [1995]. Two QMC algorithms based on Halton and Sobol' points were compared with the Monte Carlo algorithm.

- Both QMC algorithms outperformed the MC algorithm.
- The convergence of the QMC algorithms was much smoother than that of the MC algorithm. This makes automatic termination of QMC easier and more reliable than MC.
- MC is very sensitive to the initial seed.

We next summarize empirical results of Papageorgiou & Traub [1996]. They compared QMC using the generalized Faure sequence (see Tezuka [1995]) and using a modified Sobol' sequence. It must be stressed that the results reported below are for the *modified* Sobol' sequence. Published Sobol' sequences (such as in Press *et al.* [1992]) will not lead to such results. We refer to the two QMC algorithms as QMC-GF and QMC-MS, respectively.

The conclusions regarding the valuation of the CMO model problem can be divided into three groups. Similar results hold for other derivatives, such as Asian options.

(i) QMC and MC Algorithms

Both QMC algorithms beat the MC algorithm by a wide margin.

In particular:

- Both the QMC-GF and the QMC-MS algorithms converge significantly faster than the MC algorithm.
- The QMC-GF algorithm always converges at least as fast as the QMC-MS algorithm, and frequently faster.
- The MC algorithm is sensitive to the initial seed.

(ii) Small Number of Sample Points

QMC algorithms outperform the MC algorithm for a small number of sample points.

In particular:

- QMC algorithms attain small error with a small number of points.
- For the hardest of the ten integrals required for the CMO valuation, the QMC-GF algorithm achieves accuracy 10^{-2} with just 170 points, while QMC-MS uses 600 points. On the other hand, the MC algorithm requires 2700 points for the same accuracy.
- The MC algorithm tends to waste points due to clustering, which severely compromises its performance when the sample size is small.

(iii) Speedup

The advantage of QMC algorithms over the MC algorithm is further amplified as the accuracy demands grow.

In particular:

- QMC algorithms are 20 to 50 times faster than the MC algorithm with even moderate sample sizes (2000 deterministic points or more).
- When high accuracy is desired, QMC algorithms can be 1000 times faster than MC.

We amplify a number of these points. The fact that the QMC-GF algorithm achieves accuracy 10^{-2} with just 170 points is particularly important for financial computations. Since the interest and prepayment functions have considerable uncertainty, people in the financial community find valuations whose accuracy is one part in a hundred to be sufficient. Furthermore, very rapid valuations are important because a financial institution may have a large book of instruments which have to be valued on a regular basis.

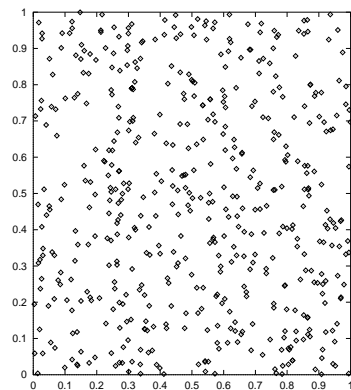


Fig. 4.1 512 pseudorandom points

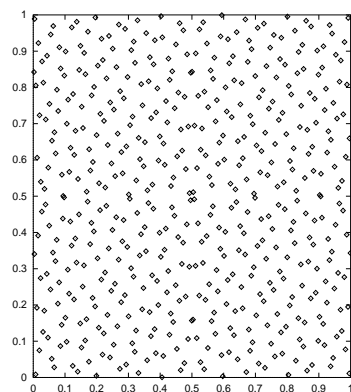


Fig. 4.2 512 low discrepancy points

Thus performance with a small number of points is particularly important for this class of applications. Contrast Figure 4.1, which exhibits 512 pseudorandom points, with Figure 4.2, which exhibits 512 low discrepancy points.

A software system called FINDER for computing high-dimensional integrals has been built at Columbia University. FINDER has modules for generating generalized Faure points and modified Sobol' points. As further improvements in low discrepancy methods are found, they will be added to the software. FINDER may be obtained from Columbia University.

Tests by other researchers, including Joy *et al.* [1996] and Ninomiya

& Tezuka [1996], lead to similar conclusions for the high-dimensional problems of mathematical finance.

The excellent results reported above are empirical. A number of hypotheses have been advanced to explain the observed results. For example, Caffisch *et al.* [1997] use a Brownian bridge argument to infer that the effective dimension of the CMO is much lower than its stated dimension. Another concept of effective dimension may be found in Paskov [1997].

We will discuss a very recent paper by Sloan & Woźniakowski [1998b], which might provide an explanation. Their paper is based on the observation that many problems of mathematical finance are highly non-isotropic. Assume that the various dimensions have “weights” $\gamma_1, \dots, \gamma_d$, where

$$1 = \gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_d \geq 0.$$

For any such sequence γ , they define a class F_γ of integrands. (Note that this is an instance of changing the class of inputs, which was discussed in Chapter 3.)

Sloan and Woźniakowski prove the following theorem. Let $n(d, \gamma, \varepsilon)$ be the minimal number of sample points needed to reduce the initial error by a factor of ε , for any integrand in F_γ . If

$$s(d, \gamma) = \sum_{j=1}^d \gamma_j$$

is finite and “small,” there exists a QMC method such that

$$n(d, \gamma, \varepsilon) \leq C\varepsilon^{-p} \quad \text{for all } d, \quad (4.4)$$

where C and p are constants, with $p \leq 2$. Note that unlike the stochastic guarantee of MC, this theorem offers a worst case guarantee.

Remark: For d finite and fixed, $s(d, \gamma)$ is, of course, finite. However, d is permitted to be arbitrarily large and, indeed, to go to infinity. \square

The Sloan-Woźniakowski paper leaves a number of issues unresolved. The theorem of Sloan and Woźniakowski is about a class of integrands. To show that this explains why QMC methods are so effective for problems in finance, it has to be established that these problems have integrands belonging to F_γ .

The proof of (4.4) is nonconstructive. It is of a great practical importance to construct sample points for which (4.4) holds. It is hoped that known low discrepancy points satisfy (4.4). Furthermore, it is known

that $p \leq 2$. If it could be shown that $p < 2$, then QMC, with a worst case guarantee, would beat MC. If it could be shown that p is close to 1, this would explain the empirical results from mathematical finance.

We summarize the discussion above by

Open Problems:

- ▷ Which financial problems have integrands belonging to F_γ ?
- ▷ Provide a *deterministic* construction of a strongly tractable QMC method for F_γ .
- ▷ What is the best value of the exponent p in (4.4)?

Research into why QMC performs so well for high-dimensional integrals of mathematical finance is currently a very active area.

QMC looks promising for problems besides those occurring in mathematical finance. For example, Papageorgiou & Traub [1997] reported test results on the model integration problem (suggested by Keister [1996])

$$\left(\frac{1}{2\pi}\right)^{d/2} \int_{\mathbb{R}^d} \cos(\|x\|) e^{-\|x\|^2} dx, \quad (4.5)$$

where $\|\cdot\|$ denotes the Euclidean norm and $d = 25$. This problem is of particular interest, since it is isotropic, as opposed to the integrands of the finance problems, which are non-isotropic. The performance of the QMC-GF algorithm is very impressive. It achieved error 10^{-2} using fewer than 500 points. Its error over the range tested (up to 10^6 points) was $C \cdot n^{-1}$, with $C < 30$. By comparison, the error of the MC algorithm was proportional to $n^{-1/2}$. Although Keister [1996] stopped at $d = 25$, Papageorgiou & Traub [1997] tested (4.5) for dimension as high as 100, with results similar to those for $d = 25$.

There exists a transformation that reduces (4.5) to a univariate problem. The empirical convergence rate of QMC is n^{-1} , as if it sees that this is really a one-dimensional problem. In contrast, the empirical convergence rate of MC remains proportional to $n^{-1/2}$; it does not see that the problem is really one-dimensional. An analysis by Papageorgiou [1998] shows that an upper bound on the convergence rate of QMC is of order $n^{-1}(\log n)^{1/2}$.

These test results suggest that non-isotropy is not the only condition under which QMC is superior to MC. Our current belief is that there are a number of classes of integrands for which QMC is superior to MC.

We end this chapter by summarizing its main points. We began with the Koksma-Hlawka and Woźniakowski theorems, which state that

(modulo a polylog factor) QMC algorithms converge as n^{-1} , while MC converges as $n^{-1/2}$. These theorems make QMC appear to be a very promising alternative to MC. However, these theorems are for the asymptotic case; they say nothing about the kinds of problems one must solve in areas such as mathematical finance, where d is very large and the typical values of n are very much not “asymptotic.” But the empirical results indicate that QMC is markedly superior to MC for many problems of mathematical finance. The theory should be enriched to explain the test results. We pose the following

Open Problem: Characterize classes of integrands for which QMC is superior to MC.