Chapter 1

RECENT ADVANCES IN RANDOMIZED QUASI-MONTE CARLO METHODS

Pierre L'Ecuyer
Département d'Informatique et de Recherche Opérationnelle
Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, H3C 3J7, CANADA
lecuyer@iro.umontreal.ca

Christiane Lemieux
Department of Mathematics and Statistics
University of Calgary, 2500 University Drive N.W., Calgary, T2N 1N4, CANADA
lemieux@math.ucalgary.ca

Abstract We survey some of the recent developments on quasi-Monte Carlo (QMC) methods, which, in their basic form, are a deterministic counterpart to the Monte Carlo (MC) method. Our main focus is the applicability of these methods to practical problems that involve the estimation of a high-dimensional integral. We review several QMC constructions and different randomizations that have been proposed to provide unbiased estimators and for error estimation. Randomizing QMC methods allows us to view them as variance reduction techniques. New and old results on this topic are used to explain how these methods can improve over the MC method in practice. We also discuss how this methodology can be coupled with clever transformations of the integrand in order to reduce the variance further. Additional topics included in this survey are the description of figures of merit used to measure the quality of the constructions underlying these methods, and other related techniques for multidimensional integration.

1. Introduction

To approximate the integral of a real-valued function f defined over the unit hypercube $[0,1)^s$, given by

$$\mu = \int_{[0,1)^s} f(\mathbf{u})d\mathbf{u}, \tag{1.1}$$

a frequently-used approach is to choose a point set $P_n = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subset [0,1)^s$ and then take the average value of f over P_n,

$$Q_n = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{u}_i), \tag{1.2}$$

as an approximation of μ. Many problems can be formulated as in (1.1), e.g., when simulation is used to estimate an expectation and each simulation run requires s calls to a pseudorandom number generator that outputs numbers between 0 and 1; see Section 2 for more details.

If f is very smooth and s is small, the product of one-dimensional integration rules such as the rectangular or trapezoidal rules can be used to define P_n [20]. When these conditions are not met, the *Monte Carlo method* (MC) is usually more appropriate. It amounts to choosing P_n as a set of n i.i.d. uniformly distributed vectors $\mathbf{u}_i$ over $[0,1)^s$. With this method, Q_n is an unbiased estimator of μ whose error can be approximated via the central limit theorem, and whose variance is given by

$$\frac{\sigma^2}{n} = \frac{1}{n}\left(\int_{[0,1)^s} f^2(\mathbf{u})d\mathbf{u} - \mu^2\right),$$

assuming that $\sigma^2 < \infty$ (i.e., f is square-integrable). This means that the error $|Q_n - \mu|$ associated with the MC method has probabilistic order $O_p(n^{-1/2})$.

Quasi-Monte Carlo (QMC) methods can be seen as a deterministic counterpart to the MC method. They are based on the idea of using more regularly distributed point sets P_n to construct the approximation (1.2) than the random point set associated with MC. The fact that QMC methods are deterministic suggests that one has to make assumptions on the integrand f in order to guarantee a certain level of accuracy for Q_n. In other words, the improved regularity of P_n comes with worst-case functions for which the QMC approximation Q_n is bad. For this reason, the usual way to analyze QMC methods consists in choosing a set $\mathcal{F}$ of functions and a definition of *discrepancy* $D(P_n)$ to measure, in some way, how far from the uniform distribution on $[0,1)^s$ is the empirical distribution induced by P_n. Once $\mathcal{F}$ and $D(P_n)$ are determined, one can usually derive upper bounds on the deterministic error, of the following form [86, 43]:

$$|Q_n - \mu| \leq D(P_n)V(f), \text{ for any function } f \in \mathcal{F}, \tag{1.3}$$

where $V(f)$ is a measure of the variability of f such that $V(f) < \infty$ for all $f \in \mathcal{F}$. A well-known special case of (1.3) is the Koksma-Hlawka inequality [50], in which $\mathcal{F}$ is the set of functions having bounded variation in the sense of Hardy and Krause, and $D(P_n)$ is the *rectangular-star discrepancy*. To compute this particular definition of $D(P_n)$, one considers all rectangular boxes in $[0,1)^s$ aligned with the axes and with a "corner" at the origin, and then take the supremum, over all these boxes, of the absolute difference between the volume of a box and the fraction of points of P_n that fall in it. The requirement that $V(f) < \infty$ in this case implies in particular that f has continuous partial derivatives; see [86, page 19] for a precise definition.

It is clear from (1.3) that a small value of $D(P_n)$ is desirable for the set P_n. This leads to the notion of *low-discrepancy sequences*, which are sequences P_∞ of points such that if P_n is constructed by taking the first n points of P_∞, then $D(P_n)$ is significantly smaller than the discrepancy of a typical set of n i.i.d. uniform points. The term *low-discrepancy point set* usually refers to a set obtained by taking the first n points of a low-discrepancy sequence, although it is sometimes used in a looser way to describe a point set that has a better uniformity than an independent and uniform point set.

In the case where $D(P_n)$ is the rectangular-star discrepancy, it is common to say that P_∞ is a low-discrepancy sequence if $D(P_n) = O(n^{-1}\log^s n)$. Following this, the usual argument supporting the superiority of QMC methods over MC is to say that if Q_n is obtained using a low-discrepancy point set, then the error bound (1.3) is in $O(n^{-1}\log^s n)$, which for a fixed dimension s is a better asymptotic rate than the $n^{-1/2}$ rate associated with MC. For this reason, one expects QMC methods to approximate μ with a smaller error than MC if n is sufficiently large. However, the dimension s does not need to be very large in order to have $n^{-1}\log^s n > n^{-1/2}$ for large values of n. For example, if $s = 10$, one must have $n \geq 1.2 \times 10^{39}$ to ensure that $n^{-1}\log^s n \leq n^{-1/2}$, and thus the superiority of the convergence rate of the QMC bound over MC is meaningful only for values of n that are much too large for practical purposes.

Nevertheless, this does not mean than QMC methods cannot improve upon MC in practice, even for problems of large dimension. Arguments supporting this are that first, the upper bound given in (1.3) is a worst-case bound for the whole set $\mathcal{F}$. It does not necessarily reflect the behavior of Q_n on a given function in this set. Second, it happens often in practice that even if the dimension s is large, the integrand f can be well approximated (in a sense to be specified in the next section) by a sum of low-dimensional functions. In that case, a good approximation Q_n for μ can be obtained by simply making sure that the corresponding projections of P_n on these low-dimensional subspaces are well distributed. These observations have recently led many researchers to turn to other tools than the setup that goes with (1.3) for analyzing and improving the application of QMC methods to practical problems, where the dimension s is typically large, or even infinite (i.e., there is no *a priori* bound on s). In connection with these new tools, the idea of randomizing QMC point sets has been an important contribution that has extended the practical use of these methods. The purpose of this chapter is to give a survey of these recent findings, with an emphasis on the theoretical results that appear most useful in practice. Along with explanations describing why these methods work, our goal is to provide, to the reader, tools for applying QMC methods to his own specific problems.

Our choice of topics is subjective. We do not cover all the recent developments regarding QMC methods. We refer the readers to, e.g., [33, 41, 86, 109] for other viewpoints. Also, the fact that we chose not to talk more about inequalities like (1.3) does not mean that they are useless. In fact, the concept of discrepancy turns out to be useful for defining selection criteria on which exhaustive or random searches to find "good" sets P_n can be based, as we will see later. Furthermore, we think it is important to be aware of the discouraging order of magnitude for n required for the rate $n^{-1}\log^s n$ to be better than $n^{-1/2}$, and to understand that this problem is simply a consequence of the fact that placing points uniformly in $[0,1)^s$ is harder and harder as s increases because the space to fill becomes too large. This suggests that the success of QMC methods in practice is due to a clever choice of point sets exploiting the features of the functions that are likely to be encountered, rather than to an unexplainable way of breaking the "curse of dimensionality".

Highly-uniform point sets can also be used for estimating the *minimum* of a function instead of its integral, sometimes in a context where function evaluations are noisy. This is discussed in Chapter 6 of [86], Chapter 13 of [33], and was also the subject of collaborative work between Sid Yakowitz and the first author [122].

This chapter is organized as follows. In Section 2, we give some insight on how point sets P_n should be constructed by using an ANOVA decomposition of the integrand over low-dimensional subspaces. Section 3 recalls the definition of different families of low-discrepancy point sets.

In Section 4, we present measures of quality (or selection criteria) for low-discrepancy point sets that take into account the properties of the decomposition discussed in Section 2. Various randomizations that have been proposed for QMC methods are described in Section 5. Results on the error and variance of approximations based on (randomized) QMC methods are presented in Section 6. The purpose of Section 7 is to briefly review different classes of transformations that can be applied to the integrand f for reducing the variance further by exploiting, or not, the structure of the point set P_n. Integration methods that are somewhere between MC and QMC but that exploit specific properties of the integrand more directly are discussed in Section 8. Conclusions and ideas for further research are given in Section 9.

2. A Closer Look at Low-Dimensional Projections

We mentioned earlier that as the dimension s increases, it becomes difficult to cover the unit hypercube $[0,1)^s$ very well with a fixed number n of points. However, if instead our goal is to make sure that for some chosen subsets $I \subset \{1,\ldots,s\}$, the projections $P_n(I)$ over the subspace of $[0,1)^s$ indexed by the coordinates in I are evenly distributed, the task is easier. By doing this, one can get a small integration error if the chosen subsets I match the most important terms in the functional ANOVA decomposition of f, which we now explain.

The functional ANOVA decomposition [23, 52, 96] writes a square-integrable function f as a sum of orthogonal functions as follows:

$$f(\mathbf{u}) = \sum_{I \subseteq \{1,\ldots,s\}} f_I(\mathbf{u}),$$

where $f_I(\mathbf{u}) = f_I(u_1,\ldots,u_s)$ corresponds to the part of f that depends only on the variables in $\{u_j, j \in I\}$. Moreover, this decomposition is such that $f_\emptyset(\mathbf{u}) = \mu$,

$$\int_{[0,1)^s} f_I(\mathbf{u}) d\mathbf{u} = 0$$

if I is non-empty, and

$$\int_{[0,1)^s} f_I(\mathbf{u}) f_J(\mathbf{u}) d\mathbf{u} = 0,$$

if $I \neq J$. Defining $\sigma_I^2 = \mathrm{Var}(f_I(\mathbf{u}))$, we then have

$$\sigma^2 = \sum_{\emptyset \neq I \subseteq S} \sigma_I^2.$$

The best mean-square approximation of $f(\cdot)$ by a sum of d-dimensional (or less) functions is $\sum_{I:|I|\le d} f_I(\cdot)$. Also, the relative importance σ_I^2/σ^2 of each component f_I indicates which variables or which subsets of variables are the most important [44].

A function f has *effective dimension* d (in the superposition sense) if $\sum_{I:|I|\le d}\sigma_I^2 \approx \sigma^2$ [11]. Functions defined over many variables but having a low effective dimension often arise in practical applications [11, 70]. The concept of effective dimension has actually been introduced (in a different form than above) by Paskov and Traub [99] in the context of financial pricing to explain the much smaller error obtained with QMC methods compared with MC, for a problem defined over a 360-dimensional space.

A broad class of problems that are likely to have low effective dimension (relative to s) are those arising from simulation applications. To see this, note that simulation is typically used to estimate the expectation of some measure of performance defined over a stochastic system, and proceeds by transforming in a more or less complicated way a sequence of numbers between 0 and 1 produced by a pseudorandom generator into an observation of the measure of interest. Hence it fits the framework of equation (1.1), with s equal to the number of uniforms required for each simulation, and f taken as the mapping that transforms a point in $[0,1)^s$ into an observation of the measure of performance. In that context, it is frequent that the uniform numbers that are generated close to each other in the simulation (i.e., corresponding to dimensions that are close together) are associated to random variables that interact more together. In other words, for these applications it is often the subsets I containing nearby indices and not too many of them that are the most important in the ANOVA decomposition. This suggests that to design point sets P_n that will work well for this type of problems, one should consider the quality of the projections $P_n(I)$ corresponding to these "important" subsets. We present, in Section 4, measures of quality defined on this basis.

We conclude this section by recalling two important properties related to the projections of a point set P_n in $[0,1)^s$. Firstly, we say that P_n is *fully projection-regular* [102, 65] if each of its projections $P_n(I)$ over a non-empty subset of dimensions $I \subseteq \{1,\ldots,s\}$ contains n distinct points. Such a property is certainly desirable, for the lack of it means that some of the ANOVA components of f are integrated by less than n points even if n evaluations of f have been done. Secondly, we say that P_n is *dimension-stationary* [65] if $P_n(I) = P_n(I+j)$ for any $I = \{i_1,\ldots,i_d\}$, $1 \le d, j < s$, such that $1 \le i_1 < \ldots < i_d < i_d + j \le s$; that is, only the spacings between the indices in I are relevant in the definition of the projections $P_n(I)$ of a dimension-stationary point set, and not their individual values. Hence not all $2^s - 1$ non-empty projections of P_n need to be considered when measuring the quality of P_n since many are the same. Another advantage of dimension-stationary point sets is that because the quality of their projections does not deteriorate as the first index i_1 increases, they can be used to integrate functions that have important ANOVA components associated with subsets I having a large value of i_1. Therefore, when working with those point sets it is not necessary to try rewriting f so that the important ANOVA components are associated with subsets I having a small first index i_1 (as is often done; see, e.g., [33]). We underline that not all types of QMC point sets have these properties.

3. Main Constructions

In this section, we present constructions for low-discrepancy point sets that are often used in practice. We first introduce *lattice rules* [102], and a special case of this construction called *Korobov rules* [55], which turn out to fit in another type of construction based on successive overlapping s-tuples produced by a recurrence defined over a finite ring. This type of construction is also used to define pseudorandom number generators (PRNGs) with huge period length when the underlying ring has a very large cardinality (e.g., $\geq 2^{100}$); low-discrepancy point sets are, in contrast, obtained by using a ring with a small cardinality (e.g., between 10^2 and 10^6). For this reason, we refer to this type of construction as *small PRNGs*, and discuss it after having introduced *digital nets* [86], which form an important family of low-discrepancy point sets that also provides examples of small PRNGs. Various digital net constructions are described. We also recall the Halton sequence [37], and discuss a method by which the number of points in a Korobov rule can be increased sequentially, thus offering an alternative to *digital sequences*. Additional references regarding the implementation of QMC methods are provided at the end of the section.

3.1 Lattice Rules

The general construction for lattice rules has been introduced by Sloan and his collaborators (see [102] and the references therein) by building upon ideas developed by Korobov [55, 56], Bakhlavov [5], and Hlawka [51]. The following definition is taken from the expository book of Sloan and Joe [102].

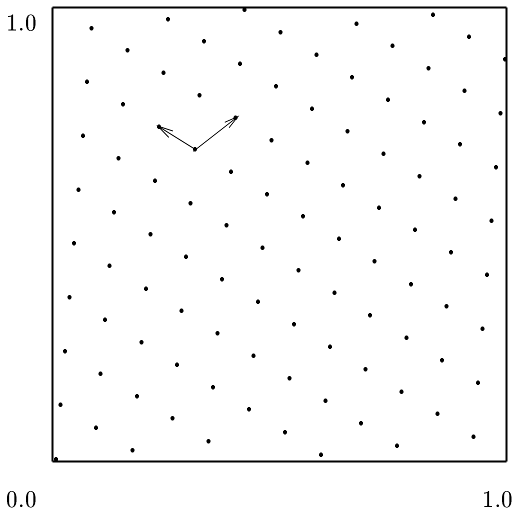

Figure 1.1. Korobov lattice point set with $n = 101$ and $a = 12$; the two vectors shown correspond to the basis $\{(-8/101, 5/101), (9/101, 7/101)\}$.

Definition: *Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_s\}$ be a set of s-dimensional vectors linearly independent over $\mathbb{R}^s$, and with coordinates in $[0, 1)$. Define*

$$L_s = \{\mathbf{v} = \sum_{j=1}^{s} z_j \mathbf{v}_j \textit{ such that each } z_j \in \mathbb{Z}\}, \tag{1.4}$$

and assume that $\mathbb{Z}^s \subseteq L_s$. The approximation Q_n based on the set $P_n = L_s \cap [0,1)^s$ is a lattice rule. *The number of points n in the rule is equal to the inverse of the absolute value of the determinant of the matrix $\mathbf{V}$ whose rows are the vectors $\mathbf{v}_j$, $j = 1, \ldots, s$. This number n is called the* order *of the rule.*

Note that the basis for L_s is not unique, but the determinant of the matrix $\mathbf{V}$ remains constant for all choices of basis.

Figure 1.1 gives an example of a point set P_n that corresponds to a two-dimensional lattice rule, with $n = 101$. Here, the two vectors shown in the figure, $\mathbf{v}_1 = (-8/101, 5/101)$ and $\mathbf{v}_2 = (9/101, 7/101)$, form a basis for the lattice L_2. Another basis for the same lattice is formed by $\mathbf{v}_1 = (1/101, 12/101)$ and $\mathbf{v}_2 = (0, 1)$.

These 101 points cover the unit square quite uniformly. They are also placed very regularly on equidistant parallel lines, for several families of lines. For example, any of the vectors $\mathbf{v}_1$ or $\mathbf{v}_2$ given above determines one family of lines that are parallel to this vector. This regularity property stems from the lattice structure and it holds for *any* lattice rule—in more than two dimensions, the lines are simply replaced by equidistant parallel hyperplanes [54, 15]. For P_n to cover the unit hypercube quite well, the successive hyperplanes should never be too far apart (to avoid wide uncovered gaps), for any choice of family of parallel hyperplanes. Selection criteria for lattice rules, based on the idea of minimizing the distance between successive hyperplanes for the "worst-case" family, are discussed in Section 4.1.

From now on, we refer to point sets P_n giving rise to lattice rules as *lattice point sets*. Each lattice point set P_n has a *rank* associated with it, which can be defined as the smallest integer r such that P_n can be obtained by taking all integer combinations, modulo 1, of r vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ independent over $\mathbb{R}^s$. Alternatively, the rank can be defined as the smallest number of cyclic groups whose direct sum yields P_n [102]. For example, if $a_1, \ldots, a_s$ are positive integers such that $\gcd(a_j, n) = 1$ for at least one j, then the lattice point set

$$P_n = \left\{ \frac{i}{n}(a_1, \ldots, a_s) \bmod 1,\ i = 0, \ldots, n-1 \right\} \tag{1.5}$$

has rank 1 and contains n distinct points. It can also be obtained by taking $P_n = L_s \cap [0,1)^s$ with $\mathbf{v}_1 = (a_1, \ldots, a_s)/n$ and $\mathbf{v}_j = \mathbf{e}_j$ for $j = 2, \ldots, s$ in (1.4), where $\mathbf{e}_j$ is a vector of zeros with a one in the j^{th} position. The condition $\gcd(a_j, n) = 1$ for all j is necessary and sufficient for a rank-1 lattice point set P_n to be fully projection-regular.

A *Korobov rule* is obtained by choosing an integer $a \in \{1, \ldots, n-1\}$, and taking $a_j = a^{j-1} \bmod n$ in (1.5), for all $j = 1, \ldots, s$. In this case, having n and a relatively prime is a necessary and sufficient condition for P_n to be both fully projection-regular and dimension-stationary [65]. The integer a is usually referred to as the *generator* of the lattice point set P_n. For instance, the point set given on Figure 1.1 has a generator a equal to 12. In Section 3.3, we describe an efficient way of constructing P_n for Korobov rules when n is prime and a is a primitive element modulo n.

In the definition of a lattice rule, we assumed $\mathbb{Z}^s \subseteq L_s$. This implies that L_s has a period of 1 in each dimension. A lattice L_s with this property is called an *integration lattice* [102]. A necessary and sufficient condition for L_s to be an integration lattice is that the inverse $\mathbf{V}^{-1}$ of the matrix $\mathbf{V}$ has only integer entries. In this case, it can be shown that if the determinant of $\mathbf{V}$ is $1/n$, then there exists a basis for L_s with coordinates of the form a/n, where $0 \le a \le n$. We assume from now on that all lattices considered are integration lattices.

The columns of the inverse matrix $\mathbf{V}^{-1}$ form a basis of the *dual lattice* of L_s, defined as

$$L_s^* = \{\mathbf{h} \in \mathbb{Z}^s : \mathbf{h} \cdot \mathbf{v} \in \mathbb{Z},\ \text{for all } \mathbf{v} \in L_s\}.$$

If the determinant of $\mathbf{V}^{-1}$ is n, then L_s^* contains n times less points per unit of volume than $\mathbb{Z}^s$. Also, L_s^* is periodic with a period of n in each dimension. As we will see in Sections 4.1 and 6.1, this dual lattice plays an important role in the error and variance analysis for lattice rules, and in the definition of selection criteria.

3.2 Digital Nets

We first recall the general definition of a digital net in base b, a concept that was first introduced by Sobol' [104] in base 2, and subsequently generalized by Faure [26], Niederreiter [83], and Tezuka [113]. The following definition is from Niederreiter [86], with the convention from Tezuka [113] that the *generating matrices* $\mathbf{C}^j$ contain an infinite number of rows (although often, only a finite number of these rows are nonzero).

Definition 1 *Let $s \geq 1$ and $k \geq 1$ be integers. Choose*

1. *a commutative ring R with identity and with cardinality b (usually $\mathbb{Z}_b$);*
2. *bijections $\psi_r : \mathbb{Z}_b \to R$ for $0 \leq r \leq k-1$;*
3. *bijections $\eta_{jl} : R \to \mathbb{Z}_b$ for $1 \leq j \leq s$ and $1 \leq l \leq k$;*
4. *Generating matrices $\mathbf{C}^1, \ldots, \mathbf{C}^s$ of dimension $\infty \times k$ over R.*

For $i = 0, \ldots, b^k - 1$ let

$$i = \sum_{r=0}^{k-1} a_r b^r,$$

with $a_r \in \mathbb{Z}_b$, be the digit expansion of i in base b. Consider the vector

$$\mathbf{y} = (\psi_0(a_0), \ldots, \psi_{k-1}(a_{k-1}))^T \in R^k$$

and compute

$$(b_{j,1}, b_{j,2}, \ldots)^T = \mathbf{C}^j \cdot \mathbf{y},$$

where each element $b_{j,l}$ is in R. For $j = 1, \ldots, s$, let

$$u_{ij} = \frac{\eta_{j1}(b_{j1})}{b} + \frac{\eta_{j,2}(b_{j,2})}{b^2} + \ldots \tag{1.6}$$

Then $P_n = \{\mathbf{u}_i = (u_{i1}, \ldots, u_{is}),\ i = 0, \ldots, n-1\}$, with $n = b^k$, is a digital net over R in base b.

This scheme has been used to construct point sets having a low-discrepancy property that can be described by introducing the notion of

$(q_1, \ldots, q_s)$*-equidistribution* [69]. Let $q = q_1 + \ldots + q_s$, where the q_j are non-negative integers, and consider the b^q *b-ary boxes* obtained by partitioning $[0,1)^s$ into b^{q_j} equal intervals along the jth axis. If each of these b^q boxes contains exactly b^{k-q} points from a point set P_n, where $n = b^k$, then P_n is said to be $(q_1, \ldots, q_s)$*-equidistributed.* If a digital net P_n is $(q_1, \ldots, q_s)$-equidistributed whenever $q \leq k - t$, for some integer $t \geq 0$, it is called a (t, k, s)*-net* [86]. The smallest integer t having this property is a widely-used measure of uniformity for digital nets and we call it the *t-value* of P_n. Note that the t-value is meaningless if $k \leq t$, and that the smaller t is, the better is the quality of P_n. Criteria for measuring the equidistribution of digital nets are discussed in more details in Section 4.2.

Figure 1.2 shows an example of a two-dimensional point set with $n = 3^4 = 81$ points in base 3, having the best possible equidistribution; that is, its t-value is 0 and thus, any partition of the unit square into ternary boxes of size 3^{-4} is such that exactly one point is in each box. The figure shows two examples of such a partition, into rectangles of sizes $3^{-2} \times 3^{-2}$ and $3^{-3} \times 3^{-1}$, respectively. The other partitions (not shown here) are into rectangles of sizes 1×3^{-4}, $3^{-1} \times 3^{-3}$, and $3^{-4} \times 1$. For all of these partitions, each rectangle contains one point of P_n. This point set contains the first 81 points of the two-dimensional Faure sequence in base 3. In this case, in the definition above, $R = \mathbb{Z}_3$, all bijections are the identity function over $\mathbb{Z}_3$, the generating matrix $\mathbf{C}^1$ for the first dimension is the identity matrix, and $\mathbf{C}^2$ is given by

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 3 & 3 & 1 \end{pmatrix}.$$

The general definition of this type of construction is given below.

From now on, b is assumed to be prime power, and in all the constructions described below, R is taken equal to $\mathbb{F}_b$. Also, we assume that a bijection from $\mathbb{F}_b$ to $\mathbb{Z}_b$ has been chosen to identify the elements of $\mathbb{F}_b$ with the "digits" $\{0, \ldots, b-1\}$, and all bijections ψ_r and η_{jl} are defined according to this bijection. In particular, when b is prime, these bijections are equal to the identity function over $\mathbb{Z}_b$, and all operations are performed modulo b. The base b and the generating matrices $\mathbf{C}^j$ therefore completely describe these constructions, because the b-ary expansion of a given coordinate u_{ij} of P_n is obtained by simply multiplying $\mathbf{C}^j$ with the digit expansion of i in base b. The goal is then to choose the generating matrices so that the equidistribution property mentioned above holds for b-ary boxes that are as small as possible. In terms of

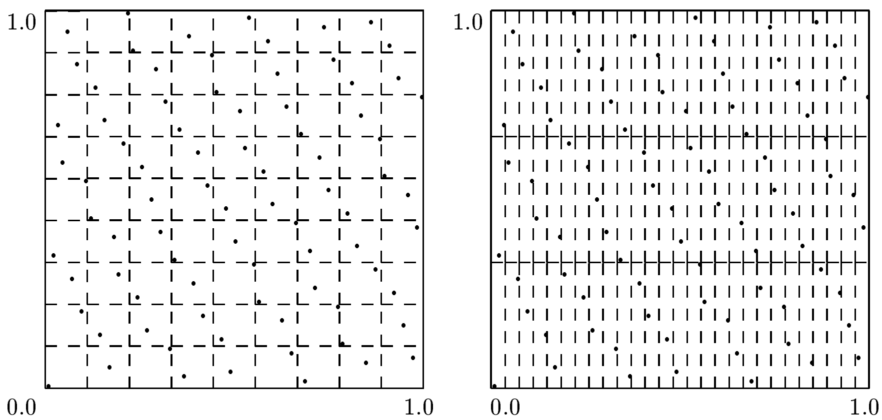

Figure 1.2. Digital net P_n in base 3 with $n = 3^4$ points obtained from a Faure sequence. Left: Each of the 81 squares of size $3^{-2} \times 3^{-2}$ contains one point from P_n; Right: Each of the 81 rectangles of size $3^{-3} \times 3^{-1}$ contains one point from P_n.

these matrices, this roughly means that we want them to have a large number of rows that are linearly independent. For example, if for each j the first k rows of the matrix $\mathbf{C}^j$ are linearly independent, then each b-ary box obtained by partitioning the jth axis into n equal intervals of length b^{-k} has one point from P_n. In particular, this implies that P_n is fully projection-regular.

Digital sequences in base b (see, e.g., [57, 86]) are infinite sequences obtained in the same way as digital nets except that the generating matrices $\mathbf{C}^j$ have an infinite number of columns; the first b^k points of the sequence thus form a digital net for each $k \geq 1$. For example, Sobol', Generalized Faure, and Niederreiter sequences (to be discussed below) are all defined as digital sequences since the "recipe" to add extra columns in the generating matrices $\mathbf{C}^j$ is given by the method.

We now describe specific well-known constructions of digital nets with a good equidistribution. We do not discuss recent constructions proposed by Niederreiter and Xing (e.g., [88, 89]), as they require the introduction of many concepts from algebraic function fields that go well beyond the scope of this chapter. These sequences are built so as to optimize the asymptotic behavior of their t-value as a function of the dimension, for a fixed base b. See [100] for a definition of these sequences and a description of a software implementation.

Sobol' Sequences. Here the base is $b = 2$ and the specification of each generating matrix $\mathbf{C}^j$ requires a primitive polynomial $f_j(z)$ over $\mathbb{F}_2$, and integers $m_{j,q}$, for $1 \leq q \leq k = \deg(f_j(z))$, to initialize a recurrence based on $f_j(z)$ that generates the *direction numbers* defining $\mathbf{C}^j$. The method specifies that the polynomial $f_j(z)$ should be the jth one in the list of primitive polynomials over $\mathbb{F}_2$ sorted by increasing degree (within each degree, Sobol' specifies a certain order which is given in the code of Bratley and Fox [8] for $j \leq 40$). There remains the parameters $m_{j,q}$ to control the quality of the point set.

Assume $f_j(z) = z^k + a_{j,1}z^{k-1} + \ldots + a_{j,k}$, where $a_{j,l} \in \mathbb{F}_2$ for each j, l. The direction numbers $v_{j,1}, v_{j,2}, \ldots$ are rationals of the form

$$v_{j,q} = \frac{m_{j,q}}{2^q} = \sum_{l=1}^{q} v_{j,q,l} 2^{-l},$$

where $m_{j,q}$ is an odd integer smaller than 2^q, for $q \geq 1$. The first k values $v_{j,1}, \ldots, v_{j,k}$ must be (carefully) chosen, and the following ones are obtained through the recurrence

$$v_{j,q} = a_{j,1}v_{j,q-1} \oplus \ldots \oplus a_{j,k-1}v_{j,q-k+1} \oplus v_{j,q-k} \oplus (v_{j,q-k} \gg k),$$

where $\oplus$ denotes a bit-by-bit exclusive-or operation, and $v_{j,q-k} \gg k$ means that the binary expansion of $v_{j,q-k}$ is shifted by k positions to the right (i.e., $v_{j,q-k}$ is divided by 2^k). These direction numbers are then used to define $\mathbf{C}^j$, whose entry in the l^{th} row and q^{th} column is given by $v_{j,q,l}$.

A good choice of the initial values $v_{j,1}, \ldots, v_{j,k}$ (or $m_{j,1}, \ldots, m_{j,k}$) in each dimension j is important for the success of this method, especially when n is small. The implementation of Bratley and Fox [8] uses the initial values given in [107] for $m_{j,q}$, with $q \leq k = \deg(f_j(z))$ and $j \leq 40$. More details on how to choose these initial values to optimize certain quality criteria are given in Section 4.2.

Generalized Faure Sequences. The Faure sequences were introduced in [26] and generalized by Tezuka [113]. For this type of sequence, the base b is the smallest prime power larger or equal to the dimension s (which means that these sequences are practical only for small values of s). An important feature of this construction is that their t-value has the best possible value ($t = 0$), provided n is of C the form $n = \lambda b^d$, for some integers $\lambda, d \geq 1$. Assuming $n = b^k$, the matrices $\mathbf{C}^j$ take the form:

$$\mathbf{C}^j = \mathbf{A}_j (\mathbf{P}^T)^{j-1},$$

where $\mathbf{P}$ is the $\infty \times k$ Pascal's matrix (i.e., with entries $\mathbf{P}_{i,j} = \binom{i-1}{j-1} \in \mathbb{F}_b$), and $\mathbf{A}_j$ is an arbitrary $\infty \times \infty$ non-singular lower-triangular matrix. The original Faure sequence is obtained by taking b prime and $\mathbf{A}_j$ as the identity matrix for all $j = 1, \ldots, s$. By allowing the matrices $\mathbf{A}_j$ to be different from the identity matrix, point sets that avoid some of the defects observed on the original Faure sequence [81, 6, 110] can be built. For instance, the Generalized Faure sequence of Tezuka and Tokuyama [115] amounts to take $\mathbf{A}_j = \mathbf{P}^{j-1}$. Recently, Faure suggested another form of Generalized Faure sequence that consists in taking $\mathbf{A}_j$ equal to the lower-triangular matrix with all nonzero entries equal to 1, for all j [27].

Niederreiter Sequences. This construction has been proposed in 1988 [84]. We first describe a special case [86, Section 4.5], and then briefly explain how it can be generalized. The base b is assumed to be a prime power, and for this special case the generating matrices are defined by distinct monic irreducible polynomials $p_1(z), \ldots, p_s(z)$ in $\mathbb{F}_b[z]$. Let $d_j = \deg(p_j(z))$ for $j = 1, \ldots, s$. In what follows, $\mathbb{F}_b((z^{-1}))$ represents the field of formal Laurent series over $\mathbb{F}_b$; that is,

$$\mathbb{F}_b((z^{-1})) = \left\{ \sum_{l=w}^{\infty} d_l z^{-l} \text{ such that } w \in \mathbb{Z} \text{ and } d_l \in \mathbb{F}_b \text{ for each } l \right\}.$$

To find the element on the lth row and qth column of $\mathbf{C}^j$, for $1 \le j \le s$, $l \ge 1$, $1 \le q \le k$, consider the expansion

$$\frac{z^r}{p_j(z)^l} = \sum_{w=0}^{\infty} a_{l,r,w}^{(j)} z^{-w-1}, \tag{1.7}$$

where $r = r(l,q)$ is the unique integer satisfying $0 \le r < d_j$ and $l - 1 = \alpha d_j + r$, and α is also uniquely determined. The element on the lth row and qth column of $\mathbf{C}^j$ is then given by $a_{\alpha+1,r,q}^{(j)}$. For these sequences, the t-value is given by $\sum_{j=1}^{s}(d_j - 1)$, which suggests that to minimize t, the $p_j(z)$'s should be taken equal to the s monic irreducible polynomials of smallest degree over $\mathbb{F}_b$. In the general definition of Niederreiter sequence [84], the numerator z^r in (1.7) can be multiplied by a different polynomial for each pair (j,l) of dimension and row index, and the $p_j(z)$'s just need to be pairwise relatively prime polynomials. Tezuka [113, Section 6.1.2] generalizes this concept a step further by removing more restrictions on this numerator.

Polynomial Lattice Rules. This construction can be seen as a lattice rule in which the ring of integers is replaced by a ring of polynomials over a finite field. As we explain below, it generalizes the construction discussed in [85], [86, Section 4.4] and [58], and is studied in more details in [69].

Definition: *Let $\{\mathbf{v}_1(z), \ldots, \mathbf{v}_s(z)\}$ be a set of s-dimensional vectors of formal Laurent series over $\mathbb{F}_b$, linearly independent over $(\mathbb{F}_b((z^{-1})))^s$, where b is a prime power. Define*

$$\mathcal{L}_s = \left\{ \mathbf{v}(z) = \sum_{j=1}^{s} q_j(z)\mathbf{v}_j(z) \text{ such that each } q_j(z) \in \mathbb{F}_b[z] \right\}, \qquad (1.8)$$

and assume that $(\mathbb{F}_b[z])^s \subseteq \mathcal{L}_s$. Let $\varphi : (\mathbb{F}_b((z^{-1})))^s \to [0,1)^s$ be the mapping that evaluates each component of a vector $\mathbf{v}(z)$ in $(\mathbb{F}_b((z^{-1})))^s$ at $z = 2$. The approximation Q_n based on the set $P_n = \varphi(\mathcal{L}_s) \cap [0,1)^s$ is a polynomial lattice rule. *The number of points n in the rule is called the* order *of the rule and is equal to b^k, where k is the degree of the inverse of the determinant of the matrix $\mathbf{V}$ whose rows are the vectors $\mathbf{v}_j(z)$.*

Most definitions and results that we mentioned for lattice rules, which from now on are referred to as *standard* lattice rules, have their counterpart for polynomial lattice rules, as we now explain. First, we refer to point sets P_n that define polynomial lattice rules as *polynomial lattice point sets*, whose rank r is equal to the smallest number of basis vectors $\mathbf{v}_1(z), \ldots, \mathbf{v}_r(z)$ required to write P_n as

$$P_n = \left\{ \varphi(\mathbf{v}(z)) : \mathbf{v}(z) = \sum_{j=1}^{r} q_j(z)\mathbf{v}_j(z) \bmod \mathbb{F}_b[z],\ q_j(z) \in \mathbb{F}_b[z] \right\}.$$

In the expression above, "mod $\mathbb{F}_b[z]$" represents the operation by which all non-negative powers in $\mathbf{v}(z)$ are dropped. The construction introduced in [85], and discussed in [86, Section 4.4] and [58], and sometimes referred to as the "method of optimal polynomials", is a polynomial lattice point set of rank 1, since it can be obtained as

$$\begin{aligned} P_n \quad = \quad & \left\{ \varphi(\mathbf{v}(z)) : \mathbf{v}(z) = \frac{q(z)}{P(z)}(g_1(z), \ldots, g_s(z)) \bmod \mathbb{F}_b[z], \right. \\ & \qquad\qquad\qquad\qquad \left. q(z) \in \mathbb{F}_b[z]/(P) \right\}, \end{aligned} \qquad (1.9)$$

where $n = b^k$, $P(z) = z^k + a_1 z^{k-1} + \ldots + a_k$ is a polynomial of degree k over $\mathbb{F}_b$, and the $g_j(z)$ are polynomials in $\mathbb{F}_b[z]/(P)$, the ring of polynomials over $\mathbb{F}_b$ modulo $P(z)$. A *Korobov polynomial lattice point set* is obtained by taking $g_j(z) = g(z)^{j-1} \bmod P(z)$ in (1.9) for some $g(z) \in \mathbb{F}_b[z]/(P)$. As in the standard case, the condition $\gcd(g(z), P(z))$ is necessary and sufficient to guarantee that a Korobov polynomial lattice point set is dimension-stationary and fully projection-regular [69].

An efficient way of generating this type of point set when $P(z)$ is a primitive polynomial is described in the next subsection.

The condition that $(\mathbb{F}_b[z])^s \subseteq \mathcal{L}_s$ implies that $\varphi(\mathcal{L}_s)$ has a period of 1 in each dimension, and we call $\mathcal{L}_s$ a *polynomial integration lattice* in this case. If $\mathbf{V}$ represents a matrix whose rows are formed by basis vectors of $\mathcal{L}_s$, then $\mathcal{L}_s$ is a polynomial integration lattice if and only if all entries in $\mathbf{V}^{-1}$ are in $\mathbb{F}_b[z]$. In that case, a basis for $\mathcal{L}_s$ with vectors having coordinates of the form $q(z)/P(z)$ can be found, where $P(z)$ is the determinant of $\mathbf{V}^{-1}$ and $q(z)$ is in $\mathbb{F}_b[z]/(P)$ or equal to $P(z)$. Finally, the columns of $\mathbf{V}^{-1}$ form a basis of the *dual lattice* of $\mathcal{L}_s$, which is defined by

$$\mathcal{L}_s^* = \{\mathbf{h}(z) \in (\mathbb{F}_b[z])^s : \mathbf{h}(z) \cdot \mathbf{v}(z) \in \mathbb{F}_b[z], \text{ for all } \mathbf{v}(z) \in \mathcal{L}_s\},$$

where the scalar product $\cdot$ is defined as $\mathbf{h}(z) \cdot \mathbf{v}(z) = \sum_{j=1}^{s} h_j(z) v_j(z)$.

This construction is a special case of digital nets in which the generating matrices are defined by

$$\mathbf{C}_{l,q}^{j} = v_{ij}(z)\big|_{l+q-1},$$

where $v_{ij}(z)\big|_m$ denotes the coefficient of z^{-m} in the formal series $v_{ij}(z)$, the jth coordinate of $\mathbf{v}_i(z)$, and the index i is determined by q and the structure of $\mathcal{L}_s$ as follows: i has the property that for some non-zero polynomials $h_{11}(z), \ldots, h_{ss}(z)$ coming from a triangular basis of $\mathcal{L}_s^*$ and such that $\prod_{m=1}^{s} h_{m,m}(z) = P(z)$ (see [69, Lemma A.2] for more details), we have

$$\deg(h_{11}(z) + \ldots + h_{i-1,i-1}(z)) < q \le \deg(h_{11}(z) + \ldots + h_{i,i}(z)).$$

In other words, the $\deg(h_{11}(z))$ first columns of each matrix $\mathbf{C}^j$ contain the coefficients associated with the first basis vector $\mathbf{v}_1(z)$; the $\deg(h_{22}(z))$ next columns are associated with the second basis vector $\mathbf{v}_2(z)$, and so on.

For polynomial lattice point sets of rank 1, the corresponding generating matrices can be described a bit more easily [86, Section 4.4], as we now explain. For each dimension $j = 1, \ldots, s$, consider the formal Laurent series expansion

$$\frac{g_j(z)}{P(z)} = \sum_{l=1}^{\infty} w_{j,l} z^{-l}.$$

The first k coefficients $w_{j,1}, \ldots, w_{j,k}$ satisfy

$$\mathbf{A} \begin{pmatrix} w_{j,1} \\ \vdots \\ w_{j,k} \end{pmatrix} = \begin{pmatrix} g_{j,1} \\ \vdots \\ g_{j,k} \end{pmatrix},$$

where $\mathbf{A}$ is defined as

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ a_1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{k-1} & \dots & a_1 & 1 \end{pmatrix}$$

(see, e.g., [69]), and the coefficients $g_{j,l}$ are such that $g_j(z) = \sum_{l=1}^{k} g_{j,l} z^{k-l}$. The coefficients $w_{j,l}$ for $l > k$ satisfy the recurrence determined by $P(z)$, i.e.,

$$w_{j,l} = -a_1 w_{j,l-1} - \dots - a_k w_{j,l-k},$$

where the minus sign represents the subtraction in $\mathbb{F}_b$. The entries of the generating matrices $\mathbf{C}^1, \dots, \mathbf{C}^s$ are defined by

$$\mathbf{C}^j_{l,q} = w_{j,l+q-1}, \text{ for } l \geq 1, 1 \leq q \leq k.$$

Note that in the definition given in [86, 101], the matrices $\mathbf{C}^j$ are restricted to k rows.

3.3 Constructions Based on Small PRNGs

Consider a PRNG based on a recurrence over a finite ring R and a specific output function from R to $[0,1)$. The idea proposed here is to define P_n as the set of all overlapping s-tuples in $[0,1)^s$ than can be obtained by running the PRNG from all possible initial states in R. More precisely, let $\phi : R \to R$ be a bijection over R called the *transition function* of the PRNG, and define

$$r_i = \phi(r_{i-1}),$$

for $i \geq 0$. The sequence $r_0, r_1, \dots$ obtained from any seed $r_0 \in R$ has a period of at most $|R|$. Now let $g : R \to [0,1)$ be an *output function*, and define

$$u_i = g(r_i)$$

for $i \geq 0$. We call the point set

$$P_n = \{(u_0, \dots, u_{s-1}),\ r_0 \in R\}, \tag{1.10}$$

where $n = |R|$, a *recurrence-based* point set. The requirement that ϕ be a bijection guarantees that P_n is dimension-stationary, and therefore fully-projection regular [65]. In concrete realizations, it is often the case that the recurrence $r_0, r_1, \dots$ has two main cycles, one of period length $|R| - 1$, and the other of period length 1 and which contains 0, the zero of R. In this case, P_n can be generated very efficiently (provided the function ϕ is easy to compute) by running the PRNG for $n+s-3$ steps to obtain the point set

$$\{(u_i, \ldots, u_{i+s-1}),\ i = 0, \ldots, n-2\}, \tag{1.11}$$

and adding the vector $(g(0), \ldots, g(0))$ to this set.

An important advantage of this construction is that it can be used easily on problems for which f depends on a random an unbounded number of variables. For this type of function, one needs a point set whose defining parameters are independent of s, and for which points of arbitrary size can be generated. Recurrence-based point sets satisfy these requirements since they are determined by the functions ϕ, g and the set R, and those are independent of the dimension. Also, points of any size can be obtained by running the PRNG as long as required. Note that when $s > n$, the coordinates of each point in P_n have a periodic structure but by randomizing P_n as in Section 5, this periodicity disappears and the coordinates of each point become mutually independent, so one can have $s \gg n$ without any problem.

The link between PRNG constructions and QMC point sets was pointed out by Niederreiter [82]. Not only similar constructions have been proposed in both fields but in addition, some of the selection criteria used in the two settings have interesting connections. Criteria measuring the uniformity of the s-dimensional point set P_n given in (1.10) are also required in the PRNG context because in this case, this point set can be seen as the sampling space for the PRNG when it is used on a problem requiring s uniform numbers per run. See [61] for details on this aspect of PRNGs and more.

We now describe two particular cases of this type of construction that provide an alternative way of generating Korobov-type lattice point sets (either standard or polynomial).

Example 1 Let $R = \mathbb{Z}_n$ for some prime n, define

$$r_i = \phi(r_{i-1}) \equiv a r_{i-1} \bmod n, \tag{1.12}$$

for some nonzero $a \in \mathbb{Z}_n$, and let $g(r_i) = r_i/n$. This type of PRNG is a *linear congruential generator* (LCG) [54, 61], and the recurrence-based point set P_n is a Korobov lattice point set. When a is a *primitive element modulo n* (see [71] for the definition), the recurrence (1.12) has the maximal period of $n-1$ for any nonzero seed, and P_n can thus be generated efficiently using (1.12) and (1.11).

Example 2 Let $P(z) = z^k + a_1 z^{k-1} + \ldots + a_k$ be a *primitive polynomial* over $\mathbb{F}_2$ (see [71] for the definition), ν be a positive integer,

$R = \mathbb{F}_2[z]/(P)$, and

$$r_i(z) = \phi(r_{i-1}(z)) \equiv z^\nu r_{i-1}(z) \bmod P(z). \tag{1.13}$$

Let $g : \mathbb{F}_2[z]/(P) \to [0,1)$ be given by the composition

$$g(r_i(z)) = \psi(r_i(z)/P(z)),$$

where $\psi : \mathbb{F}_2((z^{-1})) \to [0,1)$ is defined as the evaluation of a formal Laurent series over $\mathbb{F}_2$ at $z = 2$; that is,

$$\psi\left(\sum_{l=w}^{\infty} d_l z^{-l}\right) = \sum_{l=0}^{\infty} d_l 2^{-l}.$$

This type of PRNG is called a *polynomial LCG* [61, 113], and the recurrence-based point set P_n is a Korobov polynomial lattice point set. When ν and $2^k - 1$ are relatively prime, the recurrence (1.13) has maximal period length $2^k - 1$. Polynomial LCGs can also be obtained via Tausworthe-type linear feedback shift-register sequences [111]: the idea is to use the recurrence

$$x_i = (a_1 x_{i-1} + \ldots + a_k x_{i-k}) \bmod 2$$

over $\mathbb{F}_2$, and to define the output at step i as

$$u_i = \sum_{l=1}^{\infty} x_{i\nu+l-1} 2^{-l},$$

which in practice is typically truncated to the word-length of the computer. Tezuka and L'Ecuyer give an efficient algorithm to generate the output u_i under some specific conditions [114, 62].

Combining a few Tausworthe generators to define the output u_i can greatly help improving the quality of the associated set P_n, as explained in [62, 114, 119]. Another way of enhancing the quality of point sets based on linear recurrences modulo 2 is to use *tempering transformations* [77, 78, 66]. Note that these transformations generally destroy the lattice structure of P_n [69]. However the point set obtained is still a digital net and therefore, it can be studied under this more general setting. Conditions that preserve the dimension-stationarity of P_n under these transformations are given in [69]. The idea of combining different constructions to build sets P_n with better equidistribution properties is also discussed in [87] in the more general setting of digital nets.

3.4 Halton sequence

This sequence was introduced in 1960 by Halton [37] for constructing point sets of arbitrary length, and is a generalization of the one-dimensional van der Corput sequence [86]. Although it is not a digital sequence, it uses similar ideas and can thus be thought of as an ancestor of those sequences. For $i \geq 0$, the i^{th} point in the sequence is given by

$$\mathbf{u}_i = (\phi_{b_1}(i), \ldots, \phi_{b_s}(i)),$$

where the integers $b_1, \ldots, b_s$ are typically chosen as the s first prime numbers sorted in increasing order, and $\phi_b(i)$ is the *radical-inverse function* in base b, defined by

$$\phi_b(i) = \sum_{r=0}^{\infty} a_r b^{-r-1},$$

where the integers $a_r \in \mathbb{Z}_b$ are the coefficients in the b-ary expansion of i, i.e., $i = \sum_{r=0}^{\infty} a_r b^r$, as in the digital net definition.

3.5 Sequences of Korobov rules

With (infinite) digital sequences, one can always add new points to P_n until the estimation error is deemed small enough. The lattice point sets that we have discussed so far, on the other hand, contain only a fixed number of points. We now consider a method discussed in [104, 75, 47, 48] for generating an infinite sequence of point sets $\{P_{2^k}, k \geq 1\}$ in $[0,1)^s$ such that for $k \geq 1$, P_{2^k} is a Korobov (polynomial) lattice point set and $P_{2^k} \subseteq P_{2^{k+1}}$.

Sequences based on nested Korobov lattice point sets can be constructed by choosing a nonzero odd integer a and defining

$$P_{2^k} = \left\{ \frac{i}{n}(1, a_k, \ldots, a_k^{s-1}) \bmod 1,\ i = 0, \ldots, n-1 \right\},$$

where $a_k = a \bmod 2^k$. Hickernell et al. [48] give tables of generators a, to build these sequences, that were chosen with respect to different selection criteria to be explained in Section 4.1.

One way of constructing a sequence based on Korobov polynomial lattice point sets is to choose a polynomial $p_1(z)$ of degree 1 in $\mathbb{F}_2[z]$ (i.e., $p_1(z) = z$ or $p_1(z) = z + 1$), and a generating polynomial $g(z) \in \mathbb{F}_2[z]$ such that $\gcd(g(z), p_1(z)) = 1$. Then define

$$\begin{aligned} P_{2^k} \quad = \quad & \left\{ \varphi(\mathbf{v}(z)) : \mathbf{v}(z) = \frac{q(z)}{P_k(z)}(1, g_k(z), \ldots, g_k^{s-1}(z)) \bmod \mathbb{F}_2[z], \right. \\ & \left. \qquad\qquad q(z) \in \mathbb{F}_2[z]/(P_k) \right\}, \end{aligned}$$

where φ is defined as on page 15, $P_k(z) = p_1^k(z)$, and $g_k(z) = g(z) \bmod P_k(z)$. In this case, the sequence $\{P_{2^k}, k = 1, 2, \ldots\}$ turns out to be a special case of a digital sequence; see also [57, page 187] for the particular case where $p_1(z) = z$, and for a more general setting based on irrational formal Laurent series.

3.6 Implementations

The points of a digital net P_n in base b can be generated efficiently using a Gray code. This idea was first suggested in [3] for Sobol' sequence, and for other constructions in, e.g., [53, 101, 113, 9]. Assuming $n = b^k$, the idea is to modify the order in which the points are generated by replacing the digits $(a_0, \ldots, a_{k-1})$ from the b-ary expansion of i by the Gray code $(g_0, \ldots, g_{k-1})$ for i, which satisfies the following property: the Gray code for i and $i+1$ only differ in one position; if c is the smallest index such that $a_c \neq b-1$, then g_c is the digit whose value changes, and it becomes $g_c + 1$ in the Gray code for $i+1$ [113, Theorem 6.6]. This reduces the generation time because only a dot product of two vectors has to be performed in each dimension to generate a point.

It is sometimes suggested in the literature [9, 76, 1, 32, 81] that for most low-discrepancy sequences, an initial portion of the sequence should be discarded because of the so-called "leading-zeros phenomenon". For sequences such as Sobol's that require initial parameters, this problem can be avoided (at least partially) by choosing these parameters carefully. Using a sufficiently large number of points and randomizing the point set can also help alleviate this problem. We refer the reader to the papers mentioned above for more details.

The FORTRAN code of Bratley and Fox [8] can be used to generate Sobol' sequence for $s \leq 40$, and is available from the Collective Algorithms of the ACM at `www.acm.org/calgo`, where Fox's code [32] for generating the Faure sequence can also be found, as well as a code for Niederreiter's sequence [10]. A code to generate Generalized Faure sequences (provided the matrices $\mathbf{C}^j$ have been precomputed) is given in [113]. Owen has written code to generate scrambled nets; it is available at `www-stat.stanford.edu/~owen/scramblednets`, and is free for noncommercial uses. Recently, a C++ library called *libseq* has been developed by Friedel and Keller [34], in which they use efficient algorithms to generate *scrambled* digital sequences, Halton sequences, and other techniques such as Latin Hypercube and Supercube Sampling [79, 96]. This library can be found at `www.multires.caltech.edu/software/libseq/index.html`. There are also a few commercial software packages to generate different QMC point sets (e.g., *QR Streams* at www.mathdirect.com/products/qrn/, and the *FinDer* software [99] at www.cs.columbia.edu/~ap/html/finder.html).

4. Measures of Quality

In this section, we present a number of criteria that have been proposed in the literature for measuring the uniformity (or non-uniformity) of a point set P_n in the unit hypercube $[0,1)^s$, i.e., for measuring the *discrepancy* between the distribution of the points of P_n and the uniform distribution, in the context of QMC integration.

In one dimension (i.e., $s = 1$), several such measures are widely used in statistics for testing the goodness of fit of a data set with the uniform distribution; e.g., the Kolmogorov-Smirnov (KS), Anderson-Darling, and chi-square test statistics. Chi-square tests also apply in more than one dimension, but their efficiency vanishes quickly as the dimension increases. The rectangular-star discrepancy discussed earlier turns out to be one possible multivariate generalization of the KS test statistic. Other variants of this measure are discussed in [86], and connections between discrepancy measures and goodness-of-fit tests used in statistics are studied in [45].

The asymptotic behavior of quality measures like the rectangular-star discrepancy is known for many constructions. For example, the Halton sequence has a rectangular-star discrepancy in $O(n^{-1}\log^s n)$, but with a hidden constant that grows superexponentially with s. This is also true for Sobol' sequence, but with a smaller hidden constant. By contrast, Faure and Niederreiter-Xing sequences are built so that this hidden constant goes to zero exponentially fast as the dimension s goes to infinity.

In practice however, as soon as the number of dimensions exceeds a few units, these general measures of discrepancy are unsatisfactory because they are too hard to compute. A more practical and less general approach is to define measures of uniformity that exploit the structure of a given class of highly structured point sets. Here we concentrate our discussion on these types of criteria, starting with criteria for standard lattice rules, then covering those for digital nets.

4.1 Criteria for standard lattice rules

The criteria discussed here all relate to the dual lattice L_s^*, defined in Section 3. The first criterion we introduce has a nice geometrical interpretation, and is often used to measure the quality of LCGs through the so-called *spectral test* [31, 30, 54, 63]. It was introduced by Coveyou and MacPherson in 1967 [18] and was first used in [25] for choosing lattice point sets. It amounts to computing the euclidean length l_s of the shortest nonzero vector in L_s^*, i.e.,

$$l_s = \min_{\mathbf{0} \neq \mathbf{h} \in L_s^*} \|\mathbf{h}\|_2.$$

The quantity l_s turns out to be equal to the inverse of the distance d_s between the successive hyperplanes in the family of most distant parallel hyperplanes on which the points of L_s lie. This distance d_s should be as small as possible so there are no wide gaps in $[0,1)^s$ without any point from P_n. Equivalently, l_s should be as large as possible. Algorithms for computing l_s are discussed in, e.g., [21, 29, 54, 64]. For instance, the dual of the basis shown in Figure 1.1 contains the shortest vector in L_s^*, given by $\mathbf{h} = (5, 8)$, so $d_2 = 1/\sqrt{5^2 + 8^2} = 1/\sqrt{89}$ in this case.

This test can be applied not only to P_n, but also to any projection $P_n(I)$ of P_n. More precisely, assume $|I| = t$ and let L_I be the t-dimensional integration lattice such that $P_n(I) = L_I \cap [0,1)^t$. The idea is to compute

$$l_I = \min_{\mathbf{0} \neq \mathbf{h} \in L_I^*} \|\mathbf{h}\|_2,$$

where $L_I^* = \{\mathbf{h} \in \mathbb{Z}^t : \mathbf{h} \cdot \mathbf{v} \in \mathbb{Z} \text{ for all } \mathbf{v} \in L_I\}$. To define a criterion in which l_I is computed for several subsets I, it is convenient to normalize l_I so that the same scale is used to compare the different projections. This can be achieved by using upper bounds l_t^* derived from the "best" possible lattice in t dimensions (not necessarily an integration lattice). Such bounds can be found in, e.g., [15, 63]. Criteria using these ideas have been used for measuring LCGs [31, 30, 63], usually for subsets I containing successive indices, i.e., of the form $I = \{1, 2, \ldots, t\}$. The following criterion is more general and has been used to provide tables of "good" Korobov lattice point sets in [65]:

$$M_{t_1,\ldots,t_d} = \min\left(\min_{I=\{1,\ldots,t\}, t \leq t_1} l_I/l_t^*, \min_{2 \leq j \leq d} \min_{I \in S(j,t_j)} l_I/l_j^* \right), \tag{1.14}$$

where $S(j, t_j) = \{\{1, i_2, \ldots, i_j\}, 1 < i_2 < \ldots < i_j \leq t_j\}$. Let the *range* of a subset $I = \{i_1, \ldots, i_t\}$ be defined as $i_t - i_1 + 1$. The criterion $M_{t_1,\ldots,t_d}$ considers projections over successive indices whose range is at most t_1; over pairs of indices whose range is at most t_2; over triplets of indices whose range is at most t_3, etc. Note that for dimension-stationary point sets, it is sufficient to do as in the definition of $M_{t_1,\ldots,t_d}$ and to only consider subsets I having 1 as their first index.

The next criterion is called the *Babenko-Zaremba index*, and is similar to l_s except that a different norm is used for measuring the vectors $\mathbf{h}$ in

L_s^*. It is defined as follows [86]:

$$\rho_s = \min_{\mathbf{0} \neq \mathbf{h} \in L_s^*} \|\mathbf{h}\|,$$

where $\|\mathbf{h}\| = \prod_{j=1}^{s} \max(1, |h_j|)$. It has been used in [74] to provide tables of "good" Korobov rules, but its computation is typically much more time-consuming than computing l_s. Both l_s and ρ_s can be seen as special cases of more general criteria such as general discrepancy measures defined by Hickernell in e.g., [44, Theorem 3.8], or the *generalized spectral test* of Hellekalek [40, Definition 5.2]. These criteria use a general norm to measure $\mathbf{h}$ and apply to point sets that do not necessarily come from a lattice.

The following criterion, called *p-alpha*, uses the same norm as the Babenko-Zaremba index, but it sums a fixed power of the length of *all* vectors in the dual lattice L_s^* instead of only considering the smallest one. For an arbitrary integer $\alpha > 1$, it is defined as [102, 44]

$$\mathcal{P}_\alpha = \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} \|\mathbf{h}\|^{-\alpha}.$$

When α is even, it simplifies to [102, 44]

$$\mathcal{P}_\alpha = -1 + \frac{1}{n} \sum_{\mathbf{u} \in P_n} \prod_{j=1}^{s} \left[1 - \frac{(-1)^{\alpha/2} (2\pi)^\alpha}{\alpha!} B_\alpha(u_j) \right],$$

where $B_\alpha(\cdot)$ is the Bernoulli polynomial of degree α [102], and it can then be computed in $O(ns)$ time. This criterion has been used in, e.g., [36, 103, 102] to choose lattice point sets.

The definition of $\mathcal{P}_\alpha$ has been generalized by Hickernell [44] by the introduction of weights β_I that take into account the relative importance of each subset I (e.g., with respect to the ANOVA decomposition of f). The generalization is defined as

$$\tilde{\mathcal{P}}_\alpha = \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} \beta_{I(\mathbf{h})}^2 \|\mathbf{h}\|^{-\alpha},$$

where $I(\mathbf{h}) = \{j : h_j \neq 0\}$ is the set of nonzero indices of $\mathbf{h}$. Assuming that α is even and the β_I are product-type weights, i.e., that

$$\beta_I = \beta_0 \prod_{j \in I} \beta_j^\alpha, \tag{1.15}$$

this "weighted" $\mathcal{P}_\alpha$ becomes

$$\tilde{\mathcal{P}}_\alpha = \beta_0^2 \left\{ -1 + \frac{1}{n} \sum_{\mathbf{u} \in P_n} \prod_{j=1}^{s} \left[1 - \frac{(-1)^{\alpha/2} (2\pi\beta_j)^\alpha}{\alpha!} B_\alpha(u_j) \right] \right\},$$

which can still be computed in $O(ns)$ time. By letting $\beta_j = 1$ for $j = 0, \ldots, s$, we recover the criterion $\mathcal{P}_\alpha$. A normalized version of this criterion and the quality measure M_{t_1} described above have been used for selecting generators a defining sequences of nested Korobov lattice point sets in [48]. Note that $\tilde{\mathcal{P}}_\alpha$ can be considered as a special case of the *weighted spectral test* of Hellekalek [40, Definition 6.1]. Various other measures of quality for lattice rules can be found in [86, 102, 44] and the references therein.

4.2 Criteria for digital nets

As we mentioned in Section 3.2, the t-value of a digital net is often used to assess its quality. To compute this value, one has to find $q^* = k - t$, which is the largest integer q such that for any integers $q_j \geq 0$ satisfying $q_1 + \cdots + q_s \leq q$, the vectors $\{\mathbf{c}_r^j,\ r = 1, \ldots, q_j,\ j = 1, \ldots, s\}$ are linearly independent, where $\mathbf{c}_r^j$ denote the r^{th} row of the generating matrix $\mathbf{C}^j$ of P_n, for $j = 1, \ldots, s$. Hence, computing t is typically quite time-consuming since for a given integer q, there are $\binom{q+s-1}{q}$ different vectors $(q_1, \ldots, q_s)$ satisfying $q_j \geq 0$ and $q_1 + \cdots + q_s = q$ [101].

If we define t_I as the value of the t-value for the projection $P_n(I)$, an equivalent definition of t is

$$t = \max_{\emptyset \neq I \subseteq \{1, \ldots, s\}} t_I.$$

That is, t measures the regularity of all projections $P_n(I)$ and returns the worst case. Inside this definition of t, we can also normalize the value of t_I so that projections over subspaces of different dimensions are judged more equitably, in the same way as the value l_t^* is used to normalize l_I in the criterion $M_{t_1, \ldots, t_d}$. To do so, we can use the lower bound for t in s dimensions, given by [89]

$$t_s^* = \max\left(k,\ \frac{s-1}{b} - \log_b \frac{(b-1)(s-1) + b + 1}{2}\right),$$

and define,

$$\tilde{t} = \max_{\emptyset \neq I \subseteq \{1, \ldots, s\}} t_{|I|}^* / t_I, \tag{1.16}$$

for example. The idea behind this is that a large value of t_I for a low-dimensional subset I is usually worse than when I is high-dimensional and therefore it should be more penalized.

The definition of the t-value that we used so far is of a geometrical nature, similarly to the interpretation of l_s as being the inverse of the distance d_s between the hyperplanes of a lattice. Interestingly, just like

l_s the t-value can also be related to the length of a shortest vector in a certain dual space [113, 59, 87, 16] as we now explain. Our presentation follows [87].

Let $\mathbf{C}^1, \ldots, \mathbf{C}^s$ be the $\infty \times k$ generating matrices associated with a digital net in base b with $n = b^k$ points. Let $\mathbf{C}$ be the $k \times sk$ matrix obtained by concatening the transposed of the first k rows of each $\mathbf{C}^j$; that is, if $\tilde{\mathbf{C}}^j$ denotes the matrix containing the first k rows of $\mathbf{C}^j$, then

$$\mathbf{C} = ((\tilde{\mathbf{C}}^1)^T | \ldots | (\tilde{\mathbf{C}}^s)^T).$$

The analysis in [87] assumes that the generating matrices are $k \times k$, but we will explain shortly why it remains valid even if we start with $\infty \times k$ matrices and truncate them.

Let $\mathcal{C}_s^*$ be the null space of the row space of $\mathbf{C}$, i.e,

$$\mathcal{C}_s^* = \{\mathbf{h} \in \mathbb{F}_b^{sk} : \mathbf{C}\mathbf{h} = \mathbf{0}\}. \tag{1.17}$$

We refer to $\mathcal{C}_s^*$ as the *dual space* of the digital net P_n from now on. Define the following norm on $\mathbb{F}_b^k$: for any nonzero $h_j = (h_{j,1}, \ldots, h_{j,k}) \in \mathbb{F}_b^k$, let $v(h_j) = \max\{1 \leq l \leq k : h_{j,l} \neq 0\}$, and $v(0) = 0$. Define the norm of a vector $\mathbf{h} = (h_1, \ldots, h_s) \in \mathbb{F}_b^{sk}$ by

$$V(\mathbf{h}) = \sum_{j=1}^{s} v(h_j). \tag{1.18}$$

The following result about the t-value is proved in [87]:

$$t = k + 1 - \min_{\mathbf{0} \neq \mathbf{h} \in \mathcal{C}_s^*} V(\mathbf{h}). \tag{1.19}$$

We now explain why this result is valid even if the matrices $\mathbf{C}^j$ have been truncated to their first k rows. Let $\mathcal{C}_{s,\mathrm{tot}}^*$ denote the dual space that would be obtained without the truncation. Observe that by definition,

$$\min_{\mathbf{0} \neq \mathbf{h} \in \mathcal{C}_{s,\mathrm{tot}}^* \setminus \mathcal{C}_s^*} V(\mathbf{h}) \geq k + 1.$$

Also, using Proposition 1 in [87] and the fact that the dimension of the row space of $\mathbf{C}$ is not larger than k, we have that

$$\min_{\mathbf{0} \neq \mathbf{h} \in \mathcal{C}_s^*} V(\mathbf{h}) \leq k + 1.$$

Therefore,

$$\min_{\mathbf{0} \neq \mathbf{h} \in \mathcal{C}_{s,\mathrm{tot}}^*} V(\mathbf{h}) = \min\left(\min_{\mathbf{0} \neq \mathbf{h} \in \mathcal{C}_{s,\mathrm{tot}}^* \setminus \mathcal{C}_s^*} V(\mathbf{h}),\ \min_{\mathbf{0} \neq \mathbf{h} \in \mathcal{C}_s^*} V(\mathbf{h})\right) = \min_{\mathbf{0} \neq \mathbf{h} \in \mathcal{C}_s^*} V(\mathbf{h}),$$

which means that (1.19) is also true if we replace $\mathcal{C}_s^*$ by $\mathcal{C}_{s,\mathrm{tot}}^*$. From now on, we assume that $\mathcal{C}_s^*$ actually represents the dual space obtained without truncating the generating matrices $\mathbf{C}^j$. Also, we view $\mathcal{C}_s^*$ as a subspace of $(\mathbb{F}_b[z])^s$, which means that each element in $\mathcal{C}_s^*$ is represented by a vector of polynomials $\mathbf{h} = (h_1(z), \ldots, h_s(z))$ over $\mathbb{F}_b[z]$ and the norm $v(\cdot)$ is defined by $v(h_j(z)) = \deg(h_j(z)) + 1$, with the convention that $\deg(0) = -1$.

In the special case where P_n is a polynomial lattice point set, the dual space $\mathcal{C}_s^*$ corresponds to the dual lattice $\mathcal{L}_s^*$. If we define the norm

$$\|\mathbf{h}\| = \max_{1 \le j \le s} v(h_j(z)), \tag{1.20}$$

then

$$\ell_s = -1 + \min_{\mathbf{0} \ne \mathbf{h} \in \mathcal{L}_s^*} \|\mathbf{h}\|,$$

where ℓ_s is the *resolution* of the polynomial lattice point set. This result is discussed in [17, 16, 69, 113]. The resolution is often used for measuring the quality of PRNGs based on linear recurrences modulo 2 such as Tausworthe generators [116, 62]. From a geometrical point of view, the resolution is the largest integer l for which P_n is $(l, \ldots, l)$-equidistributed. Obviously, $\ell_s \le \lfloor k/s \rfloor$ if $n = b^k$.

This concept can be extended from polynomial lattice point sets to general digital nets by replacing $\mathcal{L}_s^*$ by $\mathcal{C}_s^*$ above. More precisely, we have:

Proposition 1 *Let P_n be a digital net in base b, and let $\mathcal{C}_s^*$ be the dual space of P_n. The resolution ℓ_s of P_n satisfies:*

$$\ell_s = -1 + \min_{\mathbf{0} \ne \mathbf{h} \in \mathcal{C}_s^*} \|\mathbf{h}\|,$$

where $\|\mathbf{h}\|$ is defined as in (1.20).

Proof: The proof of this proposition requires results given in the forthcoming sections, and it can be found in the appendix.

The resolution can be computed for any projection of P_n: for $I = \{i_1, \ldots, i_t\}$, let

$$\ell_I = \min_{\mathbf{0} \ne \mathbf{h} \in \mathcal{C}_I^*} \|\mathbf{h}\|,$$

where $\mathcal{C}_I^*$ is the dual of the row space of the $k \times tk$ matrix $(\tilde{\mathbf{C}}_{i_1}^T | \ldots | \tilde{\mathbf{C}}_{i_t}^T)$. The following criterion has been proposed to select polynomial lattice point sets [69], and it could also be used for any digital net:

$$\Delta_{t_1,\ldots,t_d} = \max\left(\max_{I=\{1,\ldots,t\}, t \le t_1} (\ell_t^* - \ell_I),\ \max_{2 \le j \le d} \max_{I \in S(j,t_j)} (\ell_j^* - \ell_I) \right), \tag{1.21}$$

where the set $S(j, t_j)$ has the same meaning as in the definition of the criterion $M_{t_1,\ldots,t_d}$ in (1.14).

Another criterion is the digital version of the quality measure $\tilde{\mathcal{P}}_\alpha$. It is closely related to the *dyadic dyaphony* [42] and the weighted spectral test [40], and was introduced in [69] for polynomial lattice point sets in base 2. It uses a norm $W(\mathbf{h})$ defined as

$$W(\mathbf{h}) = \sum_{\substack{1 \le j \le s \\ h_j(z) \neq 0}} \deg(h_j(z)),$$

and is defined for any integer $\alpha > 1$ and weights β_I, as

$$\tilde{\mathcal{D}}_\alpha = \sum_{\mathbf{0} \neq \mathbf{h} \in \mathcal{C}_s^*} \beta_{I(\mathbf{h})}^2 W(\mathbf{h}),$$

where $I(\mathbf{h}) = \{j : h_j(z) \neq 0\}$. In the special case where $b = 2$, α is even, and the weights β_I are product-type weights as in (1.15), we have that

$$\tilde{\mathcal{D}}_\alpha = \frac{\beta_0^2}{n} \sum_{i=0}^{n-1} \tilde{\xi}(\mathbf{u}_i),$$

where

$$\tilde{\xi}(\mathbf{u}) = -1 + \prod_{j=1}^{t} \left[1 + 2\beta_j^2 \left(1 - 3 \cdot 2^{\lfloor \log_2 u_j \rfloor}\right)\right],$$

and it is assumed that $2^{\lfloor \log_2 u_j \rfloor} = 0$ when $u_j = 0$.

We conclude this subsection by giving references where numerical values of the previous criteria are given for specific point sets. The value of the t-value for Sobol' sequence can be found in [89] for dimensions $s \le 20$; these values are compared in that paper with those obtained from the improved base-2 sequences proposed by Niederreiter and Xing [88, 89]. The "Salzburg Tables" given in [101] list optimal polynomial pairs $(g(z),\, P(z))$ and their associated t-value, to build Korobov polynomial rules. Generalized Faure sequences are built so that $t = 0$, but with the drawback that the base b must be at least as large as the dimension. Hence only small powers k of the bases are typically used in practice, and the quality of P_n is measured only for k-dimensional (or less) projections, where k is small. This illustrates the fact that for comparisons to be fair between different constructions, the value of the t-value should be considered in conjunction with the base b. Algorithms to compute t are discussed in [101].

The resolution has been used by Sobol' for finding optimal values to initialize recurrences defining the direction numbers in his construction

[104, 106]. More precisely, his *Property A* means that the first 2^s points of the sequence have the maximal resolution of 1, and his *Property A'* means that the first 2^{2s} points have the maximal resolution of 2.

Following ideas from [81, 13], a criterion related to $\Delta_{t_1,\ldots,t_d}$ is used to find initial direction numbers for Sobol' sequence in dimensions $s > 40$ in the forthcoming `RandQMC` library [68]; the maximum in (1.21) is taken over all two-dimensional subsets I of the form $I = \{i_1, j\}$, where $j > 40$ is the dimension for which we want to find initial direction numbers, and $j - 8 \leq i_1 \leq j - 1$. Examples of parameters for polynomial lattice point sets chosen with respect to $\Delta_{t_1,\ldots,t_4}$ for different values of $t_1, \ldots, t_4$ are given in [69].

5. Randomizations

Once a construction is chosen for P_n and the approximation Q_n given by (1.2) is computed, it is usually important to have an estimate of the error $|Q_n - \mu|$. For that purpose, upper bounds of the form (1.3) are not very useful since they are usually much too conservative, in addition to being hard to compute and restricted to a possibly small set of functions. Instead, one can randomize the set P_n so that: 1) each point in the randomized point set $\tilde{P}_n$ has a uniform distribution over $[0,1)^s$; 2) the regularity (or low-discrepancy property) of P_n, as measured by a specific quality criterion, is preserved under the randomization. The first property guarantees that the approximation

$$\hat{\mu}_{\mathrm{RQMC}} = \frac{1}{n} \sum_{\mathbf{u}_i \in \tilde{P}_n} f(\mathbf{u}_i)$$

is an unbiased estimator of μ. When the second property holds, the variance of the estimator $\hat{\mu}_{\mathrm{RQMC}}$ can usually by expressed in a way that establishes a relation between the optimization of the criterion whose value is preserved under the randomization, and the minimization of the estimator's variance. Hence these two properties help viewing randomized QMC methods as variance reduction techniques that preserve the unbiasedness of the MC estimator. In practice, the variance of $\hat{\mu}_{\mathrm{RQMC}}$ can be estimated by generating i.i.d. copies of $\hat{\mu}_{\mathrm{RQMC}}$ through independent replications of the randomization. This estimator can then be compared with the estimated variance of the MC estimator to assess the effectiveness of QMC for any particular problem.

We now describe some randomizations having these two properties and that have been proposed in the literature for the constructions presented in the preceding section.

5.1 Random shift modulo 1

The following randomization has originally been proposed by Cranley and Patterson for standard lattice rules [19]. Some authors suggested that it could also be used for other low-discrepancy point sets [117, 80].

Let $P_n = \{\mathbf{u}_i, i = 0, \ldots, n-1\}$ be a given point set in $[0,1)^s$, and Δ an s-dimensional random vector uniformly distributed over $[0,1)^s$. The *randomly shifted estimator* based on P_n is defined as

$$\hat{\mu}_{\mathrm{sh}} = \frac{1}{n} \sum_{i=0}^{n-1} f((\mathbf{u}_i + \Delta) \bmod 1). \tag{1.22}$$

When P_n is a lattice point set, the length l_I of the shortest vector associated with any projection $P_n(I)$ is preserved under this randomization. An explicit expression for the variance of $\hat{\mu}_{\mathrm{sh}}$ in that case will be given in Section 6.1.

With this randomization, each shifted point $\tilde{\mathbf{u}}_i = ((\mathbf{u}_i + \Delta) \bmod 1)$ is uniformly distributed over $[0,1)^s$. Therefore, even if the dimension s is much larger than the number of points n and if many coordinates are equal within a given point $\mathbf{u}_i$ (for instance, when P_n comes from a LCG with a small period, as in Section 3.3), these coordinates become mutually independent *after* the randomization. Hence each point has the same distribution as in the MC method; the difference with MC is that the n points of the shifted lattice are not independent. See [65], Section 10.3, for a concrete numerical example with $n = 1021$ and $s = 10000$. These properties also hold for the other randomizations described below.

5.2 Digital b-ary shift

When P_n is a digital net in base b, the counterpart of the previous method is to consider the b-ary expansion of the random vector Δ, and to add it to each point of P_n using operations over $\mathbb{F}_b$. More precisely, if $\Delta = (\Delta_1, \ldots, \Delta_s)$ and

$$\Delta_j = \sum_{l=1}^{\infty} d_{j,l} b^{-l}, \qquad u_{ij} = \sum_{l=1}^{\infty} u_{ijl} b^{-l},$$

we compute

$$\mathbf{u}_i \oplus \Delta = (\tilde{u}_{i1}, \ldots, \tilde{u}_{is}), \tag{1.23}$$

where

$$\tilde{u}_{ij} = \sum_{l=1}^{\infty} (u_{i,j,l} + d_{j,l})\, b^{-l}, \tag{1.24}$$

and let $\tilde{P}_n = \{\tilde{\mathbf{u}}_i, i = 0, \ldots, n-1\}$.

This randomization was suggested to us by Raymond Couture for point sets based on linear recurrences modulo 2 (see also [69]). It is also used in an arbitrary base (along with other more time-consuming randomizations) in [53, 76] as we will see in Section 5.4. It is best suited for digital nets in base b, and its application preserves the resolution and the t-value of any projection.

5.3 Scrambling

This randomization has been proposed by Owen [94], and it also preserves the t-value of a digital net and its resolution, for any projection. It works as follows: in each dimension $j = 1, \ldots, s$, partition the interval $[0, 1)$ in b equal parts and permute them uniformly and randomly; then, partition each of these sub-intervals into b equal parts and permute them uniformly and randomly; etc. More precisely, to scramble L digits one needs to randomly and uniformly generate several independent permutations $\pi^j_{i_1,\ldots,i_t}$ of the integers $[0 \ldots b-1]$ (assuming a specific bijection has been chosen to identify the elements in $\mathbb{F}_b$ with those in $\mathbb{Z}_b$, if b is not prime), where $i_1 = 0$, $i_t \in \{0, \ldots, b-1\}$, $1 \le t \le L$, $1 \le j \le s$, and compute

$$\tilde{u}_{i,j} = \sum_{l=1}^{\infty} \tilde{u}_{i,j,l} b^{-l},$$

where

$$\tilde{u}_{i,j,l} = \begin{cases} \pi^j_0(u_{i,j,1}) & \text{for } l = 1, \\ \pi^j_{0,u_{i,j,1},\ldots,u_{i,j,l-1}}(u_{i,j,l}) & \text{for } 1 < l \le L, \\ u_{i,j,l} & \text{for } l > L. \end{cases}$$

In practice, L may be chosen equal to the word-length of the computer, and the digits $u_{i,j,l}$ for $l > L$ are then dropped. However, as Matoušek points out [76], if $n = b^k$ and no two points have the same first k digits in each dimension (i.e., for each j, the unidimensional projection $P_n(\{j\})$ has a maximal resolution of k), then the permutations after level k are independent for each point and therefore, the random digits $\tilde{u}_{i,j,l}$ for $l > k$ can be generated uniformly and independently over $\mathbb{Z}_b$. Hence in this case we do not need to generate any permutation after level k.

When b and s are large, the amount of memory required for storing all the permutations becomes very large, and only a *partial scrambling* might then be feasible; that is, scramble $m < k$ digits, and generate the remaining ones randomly and uniformly [33], or use the permutations for the mth digit [110]. A clever way of avoiding storage problems is discussed by Matoušek [76], and a related idea is used in Morohosi's code (which can be found at `www.misojiro.t.u-tokyo.ac.jp/~morohosi`) for scrambling Faure sequences. The idea is to avoid storing all the permutations by reinitializing appropriately the underlying PRNG so that the permutations can be regenerated as they are needed. This is especially useful when the base b is large, which happens when Faure sequences are used in large or even moderate dimensions.

5.4 Random Linear Scrambling

Matoušek [76] proposes an alternative scrambling method that does not require as much randomness and storage. It borrows ideas from the scrambling technique and transformations proposed by Faure and Tezuka [28, 113]. This method is also discussed in [53], where it is called "Owen-Scrambling"; our presentation follows theirs but we prefer the name used by Matoušek to avoid any confusion with the actual scrambling proposed by Owen. The idea is to generate s nonsingular lower-triangular $k \times \infty$ matrices $\mathbf{L}_1, \dots, \mathbf{L}_s$ with elements chosen randomly, independently, and uniformly over $\mathbb{F}_b$ (the elements on the main diagonal of each $\mathbf{L}_j$ are chosen over $\mathbb{F}_b \backslash \{0\}$), where k is such that $n = b^k$, and s ∞-dimensional vectors $\mathbf{d}_1, \dots, \mathbf{d}_s$ with entries independently and uniformly distributed over $\mathbb{F}_b$. The digits $\tilde{u}_{i,j,1}, \tilde{u}_{i,j,2}, \dots$ of a randomized coordinate $\tilde{u}_{i,j}$ are then obtained as

$$\begin{pmatrix} \tilde{u}_{i,j,1} \\ \tilde{u}_{i,j,2} \\ \vdots \end{pmatrix} = \mathbf{L}_j \begin{pmatrix} u_{i,j,1} \\ u_{i,j,2} \\ \vdots \end{pmatrix} + \mathbf{d}_j,$$

where all operations are performed in $\mathbb{F}_b$.

5.5 Others

We briefly mention some other ideas that can be used to randomize QMC point sets. In addition to the linear scrambling, Matoušek [76] proposes randomization techniques for digital sequences that are easier to generate than the scrambling method, while retaining enough randomness for the purpose of some specific theoretical analyses. Hong and Hickernell suggest another form of linear scrambling that incorporates transformations proposed by Faure and Tezuka [28]. Randomizations that use permutations in each dimension to reorder the Halton sequence are discussed in [7, 81]. Wang and Hickernell [120] propose to randomize this sequence by randomly generating its starting point.

Some authors [90, 108] suggest partitioning the set of dimensions $\{1, \dots, s\}$ into two subsets (typically, of successive indices, i.e., $\{1, \dots, t\}$ and $\{t+1, \dots, s\}$), and then to use a QMC method (randomized or not) on one subset and MC on the other one. One of the justifications for this approach is that some digital nets (e.g., Sobol' sequence) are known to have projections $P_n(I)$ with better properties when I contains small indices; this suggests using QMC on the first few dimensions and MC on the remaining ones. However, this argument becomes irrelevant if a dimension-stationary point set is used. More importantly, if the QMC point set is randomized and can be shown (or presumed) to do no worse than MC in terms of its variance, then there is no advantage or "safety net" gained by using MC on one part of the problem. Estimators obtained by "padding" randomized QMC point sets with MC are analyzed in [96, 33]. Owen [96] discusses some other padding techniques, as well as a method called *Latin Supercube Sampling* to handle very high-dimensional problems.

6. Error and Variance Analysis

In this section, we study the error for the approximations Q_n based on low-discrepancy point sets, and the variance of estimators obtained by randomizing these sets. All the results mentioned here are obtained by using a particular basis for $\mathcal{L}^2([0,1)^s)$, the set of square-integrable functions over $[0,1)^s$, to expand the integrand f. In each case, the basis is carefully chosen, depending on the structure of P_n, so that the behavior of the approximation Q_n on the basis functions is easy to analyze.

The results presented here describe, for a given square-integrable function, the variance of different randomized QMC estimators. This is one possible approach for analyzing the performance of these estimators; different aspects of randomized QMC methods have been studied in the literature. In particular, Hickernell and Woźniakowski [49] have recently studied the behavior of different types of error on Hilbert spaces $\mathcal{F}$ of integrands in a general setting that includes MC, scrambled nets, and randomly shifted lattices. They consider the "worst-case error" (error on the worst integrand for a given QMC approximation), the "random-case error" (expected error, e.g., in the root mean square sense, of a randomized QMC estimator on a given integrand), and the "average-case error" (average error over $\mathcal{F}$), and they give explicit formulas for these errors. This type of analysis provides useful connections between variance expressions such as those studied here (which corresponds to the random-case setting), and the discrepancy measures discussed in the previous section (which corresponds to the worst-case setting). Specific results for scrambled-nets estimators are given in [39].

6.1 Standard Lattices and Fourier Expansion

For standard lattice rules, the following result suggests that expanding f in a Fourier series is appropriate for error and variance analysis. Recall that the Fourier basis for $\mathcal{L}^2([0,1)^s)$ is orthonormal and given by $\{e^{2\pi\iota\mathbf{h}\cdot\mathbf{u}},\ \mathbf{h} = (h_1,\ldots,h_s) \in \mathbb{Z}^s\}$, where $\iota = \sqrt{-1}$, and $\mathbf{h}\cdot\mathbf{u} = \sum_{j=1}^{s} h_j u_j$.

Lemma 1 *([102, Lemma 2.7]) If $P_n = \{\mathbf{u}_i,\ i = 0,\ldots,n-1\}$ is a lattice point set, then for any $\mathbf{h} \in \mathbb{Z}^s$,*

$$\frac{1}{n}\sum_{i=0}^{n-1} e^{2\pi\iota\mathbf{h}\cdot\mathbf{u}_i} = \begin{cases} 1 & \text{if } \mathbf{h} \in L_s^*, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the lattice rule integrates a basis function $e^{2\pi\iota\mathbf{h}\cdot\mathbf{u}}$ with no error when $\mathbf{h} \notin L_s^*$ or $\mathbf{h} = \mathbf{0}$, and with error 1 otherwise. Using this, we get the following result:

Proposition 2 *Let P_n be a lattice point set, $\hat{\mu}_{\mathrm{sh}}$ be defined as in (1.22), and*

$$\hat{f}(\mathbf{h}) = \int_{[0,1)^s} f(\mathbf{u}) e^{-2\pi\iota\mathbf{h}\cdot\mathbf{u}} d\mathbf{u}$$

be the Fourier coefficient of f evaluated in $\mathbf{h}$. (From [102]) If f has an absolutely convergent Fourier series, then

$$Q_n - \mu = \sum_{\mathbf{0}\neq\mathbf{h}\in L_s^*} |\hat{f}(\mathbf{h})|.$$

(From [65]) If $f \in \mathcal{L}^2([0,1)^s)$, then

$$\mathrm{Var}(\hat{\mu}_{\mathrm{sh}}) = \sum_{\mathbf{0}\neq\mathbf{h}\in L_s^*} |\hat{f}(\mathbf{h})|^2, \tag{1.25}$$

whereas for the MC estimator $\hat{\mu}_{\mathrm{MC}}$ based on n points,

$$\mathrm{Var}(\hat{\mu}_{\mathrm{MC}}) = \frac{\sigma^2}{n} = \frac{1}{n}\sum_{\mathbf{0}\neq\mathbf{h}\in\mathbb{Z}^s} |\hat{f}(\mathbf{h})|^2.$$

The result (1.25) was proved independently in [118], but under the stronger assumption that f has an absolutely convergent Fourier series. Notice that by contrast with the MC estimator, there is no factor of $1/n$ that multiplies the sum of squared Fourier coefficients for the randomly shifted lattice rule estimator $\hat{\mu}_{\mathrm{sh}}$. Hence in the worst case, the variance of $\hat{\mu}_{\mathrm{sh}}$ could be n times as large as the MC estimator's variance. This worst case corresponds to an extremely unlucky pairing of function and point set for which $\hat{f}(\mathbf{h}) = 0$ for all $\mathbf{h} \notin L_s^*$. However, in the expression for the variance of $\hat{\mu}_{\mathrm{sh}}$ the coefficients are summed only over the dual lattice L_s^*, which contains n times less points than the set $\mathbb{Z}^s$ over which the sum is taken in the MC case. Therefore, if the dual lattice L_s^* is such that the squared Fourier coefficients are smaller "on average" over L_s^* than over $\mathbb{Z}^s$, then the variance of $\hat{\mu}_{\mathrm{sh}}$ will be smaller than the variance of $\hat{\mu}_{\mathrm{MC}}$.

From the results given in the previous proposition, different bounds on the error and variance can be obtained by making additional assumptions on the integrand f [102, 44, 46]. Most of these bounds involve the quality measures $\mathcal{P}_\alpha$ or $\tilde{\mathcal{P}}_\alpha$. Hence a point set P_n that minimizes one of these two criteria minimizes a bound on the error or variance for the class of functions for which those bounds hold. Such analyses often provide arguments in favor of these criteria. A different type of analysis, based on the belief that the largest squared Fourier coefficients tend to be associated with "short vectors" $\mathbf{h}$, corresponding to the low frequency terms of f, suggests that the lattice point set should be chosen so that L_s^* does not contain those "short" vectors. From this point of view, a criterion like $M_{t_1,\ldots,t_d}$ seems appropriate since it makes sure that L_s^* does not contain vectors with a small euclidean length. This criterion also has the advantage of being usually much faster to compute than $\mathcal{P}_\alpha$ or ρ_s [25, 48].

6.2 Digital Nets and Haar or Walsh Expansions

Recall that digital nets are usually built so as to satisfy different equidistribution properties with respect to partitions of the unit hypercube $[0,1)^s$ into b-ary boxes. For this reason, it is convenient to use a basis consisting of step functions that are constant over b-ary boxes for studying their associated error and variance. Both Walsh and Haar basis functions have this property. In addition, the Walsh functions form an orthonormal basis of $\mathcal{L}^2([0,1)^s)$.

Scrambled-type estimators. We first define the Haar basis functions in base b, following the presentation in [95, 38]. Let $I \subseteq \{1,\ldots,s\}$, $\kappa_I = (\kappa_j)_{j\in I}$ be a vector of positive integers, $\tau = (\tau_j)_{j\in I}$ be a vector such that $0 \le \tau_j < b^{\kappa_j}$, and $\gamma = (\gamma_j)_{j\in I}$ be such that $0 \le \gamma_j < b$. A *multivariate Haar wavelet basis function* is defined as

$$\psi_{\kappa_I,\tau,\gamma}(\mathbf{u}) = b^{(|\kappa_I|-|I|)/2} \prod_{j\in I} \left(b\mathbf{1}_{\lfloor b^{\kappa_j+1} u_j \rfloor = b\tau_j+\gamma_j} - \mathbf{1}_{\lfloor b^{\kappa_j} u_j \rfloor = \tau_j} \right),$$

where $|\kappa_I| = \sum_{j \in I} \kappa_j$. Now consider the part of the Haar expansion of f that depends on the basis functions associated with a given vector κ_I, i.e., let

$$\nu_{\kappa_I}(\mathbf{u}) = \sum_{\tau} \sum_{\gamma} \langle \psi_{\kappa_I,\tau,\gamma}(\mathbf{u}), f \rangle \, \psi_{\kappa_I,\tau,\gamma}(\mathbf{u}),$$

where

$$\langle \psi_{\kappa_I,\tau,\gamma}(\mathbf{u}), f \rangle = \int_{[0,1)^s} f(\mathbf{u}) \, \psi_{\kappa_I,\tau,\gamma}(\mathbf{u}) d\mathbf{u}.$$

The function $\nu_{\kappa_I}(\mathbf{u})$ is also a step function, which is constant within the b-ary boxes obtained by partitioning $[0,1)^s$ into b^{κ_j} equal intervals along the jth axis, for each $j \in I$. Owen [95] shows that the variance of the estimator $\hat{\mu}_{\mathrm{scr}}$ based on a scrambled digital net with n points is given by

$$\mathrm{Var}(\hat{\mu}_{\mathrm{scr}}) = \frac{1}{n} \sum_{I} \sum_{\kappa_I} \Gamma_{\kappa_I} \sigma^2_{\kappa_I},$$

where

$$\sigma^2_{\kappa_I} = \mathrm{Var}(\nu_{\kappa_I}(\mathbf{u})), \tag{1.26}$$

and Γ_{κ_I} depends on the equidistribution properties of the digital net and the definition of the scrambling. Assuming that the t-value of the net is t and $n = b^k$, we have [97]:

$$\Gamma_{\kappa_I} \begin{cases} = 0 & \text{if } |\kappa_I| + |I| \leq k - t, \\ \leq b^t \left(\frac{b^{|I|} + (b-2)^{|I|}}{2(b-1)^{|I|}} \right) & \text{if } |\kappa_I| > k - t, \\ \leq b^t \left(\frac{b+1}{b-1} \right)^{|I|} & \text{otherwise.} \end{cases} \tag{1.27}$$

Using this, Owen obtains the following bound on the variance of the scrambled-net estimator:

Proposition 3 *[97, Theorem 1] Let $\hat{\mu}_{\mathrm{scr}}$ be the estimator constructed from a scrambled digital net with $n = b^k$ points. For any square-integrable function,*

$$\mathrm{Var}(\hat{\mu}_{\mathrm{scr}}) \leq \frac{b^t}{n} \left(\frac{b+1}{b-1} \right)^s \sigma^2. \tag{1.28}$$

Hence the variance of the scrambled-net estimator cannot be larger than the MC estimator's variance, up to a certain constant (independent of n but growing exponentially with s for a fixed base b). For a $(0,k,s)$-net, which implies that $b \geq s$, this constant can be bounded by $e \approx 2.718$

[95]. In the case where f satisfies certain smoothness properties (its mixed partial derivatives satisfy a Lipschitz condition), Owen shows that

$$\sigma^2_{\kappa_I} = O(b^{-2|\kappa_I|}).$$

Under this assumption, he obtains a bound in $O(n^{-3}(\log n)^s)$ for the variance of the scrambled-net estimator. Other results on the asymptotic properties of the scrambled-net estimator, that use Haar wavelets, can be found in [38, 39] and the references cited there. Haar series are also considered in the context of QMC integration in e.g., [24, 105].

An important point to mention is that the scrambling is not the only randomization for which (1.27) holds; the result is valid for any randomization satisfying the following properties [53, 76]:

1 each point $\tilde{\mathbf{u}}_i$, $i = 0, \ldots, n-1$ in the randomized point set $\tilde{P}_n$ is uniformly distributed over $[0,1)^s$;

2 for $0 \le i, r < n$ and $1 \le j \le s$, if $u_{i,j,l} = u_{r,j,l}$ for $l = 1, \ldots, k$ but $u_{i,j,k+1} \neq u_{r,j,k+1}$, then

   (a) $\tilde{u}_{i,j,l} = \tilde{u}_{r,j,l}$ for $l = 1, \ldots, k$;

   (b) $(\tilde{u}_{i,j,k+1}, \tilde{u}_{r,j,k+1})$ is uniformly distributed over $\{(a_1, a_2) \in \mathbb{F}_b^2 : a_1 \neq a_2\}$;

   (c) $(\tilde{u}_{i,j,p}, \tilde{u}_{i,j,q})$ are uncorrelated, for any $p, q > k+1$.

The random linear scrambling mentioned in Section 5.4 is shown to satisfy these properties in [53], and therefore the bound (1.28) given in Proposition 3 holds for linearly scrambled estimators as well. This is interesting since this method has a faster implementation than the scrambling of Section 5.3. Note that the digital b-ary shift does not satisfy 2(b) since the digits $\tilde{u}_{i,j,k+1}, \tilde{u}_{r,j,k+1}$ are such that $\tilde{u}_{r,j,k+1} - \tilde{u}_{i,j,k+1} = u_{r,j,k+1} - u_{i,j,k+1} \in \mathbb{F}_b$.

Digitally b-ary shifted estimators. To study the variance of a b-ary shifted digital net we use a Walsh expansion for f. Walsh series have also been used to analyze the error produced by (non-randomized) digital nets by Larcher and his collaborators, e.g., in [59, 58, 60]. In the presentation of the forthcoming results, the vector $\mathbf{h}$ will be used both to represent elements of $\mathbb{N}_0^s$, where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and elements of the ring $(\mathbb{F}_b[z])^s$. When required, we will use the bijection $\zeta : \mathbb{N}_0^s \to (\mathbb{F}_b[z])^s$ defined by $\zeta(h_1, \ldots, h_s) = (\sum_{l=1}^{\infty} h_{1,l} z^{l-1}, \ldots, \sum_{l=1}^{\infty} h_{s,l} z^{l-1})$, where $h_j = \sum_{l=1}^{\infty} h_{j,l} b^{l-1}$ for each j, to go back and forth between these two spaces.

For any $\mathbf{h} \in \mathbb{N}_0^s$, the Walsh basis function in $\mathbf{h}$ is defined as

$$\phi_{\mathbf{h}}(\mathbf{u}) = e^{2\pi\iota\mathbf{h}\cdot\mathbf{u}/b},$$

where $\mathbf{h} \cdot \mathbf{u} = \sum_{j=1}^s h_j \cdot u_j = \sum_{j=1}^s \sum_{l=1}^\infty h_{j,l} u_{j,l}$, and the coefficients $h_{j,l}$ and $u_{j,l}$ are such that $h_j = \sum_{l=1}^\infty h_{j,l} b^{l-1}$ and $u_j = \sum_{l=1}^\infty u_{j,l} b^{-l}$, and all operations are performed in $\mathbb{F}_b$. For $\mathbf{h}_1, \mathbf{h}_2 \in \mathbb{N}_0^s$, we have that

$$\phi_{\mathbf{h}_1 \oplus \mathbf{h}_2}(\mathbf{u}) = \phi_{\mathbf{h}_1}(\mathbf{u}) \phi_{\mathbf{h}_2}(\mathbf{u}),$$

where $\mathbf{h}_1 \oplus \mathbf{h}_2$ corresponds to a digit-by-digit addition over $\mathbb{F}_b$ (as if we were adding the corresponding elements in $(\mathbb{F}_b[z])^s$). See [59, 60] for more information on generalized definitions of Walsh series in the context of QMC integration.

Let $\tilde{f}(\mathbf{h})$ denote the Walsh coefficient of f in $\mathbf{h}$, that is

$$\tilde{f}(\mathbf{h}) = \int_{[0,1)^s} f(\mathbf{u}) e^{-2\pi\iota\mathbf{h}\cdot\mathbf{u}/b} d\mathbf{u}. \tag{1.29}$$

The following result may be interpreted as the digital counterpart of the result stated in Lemma 1. Recall that $\mathcal{C}_s^*$ denotes the dual space of P_n.

Lemma 2 *Let P_n be a digital net in base b with $n = b^k$. For any $\mathbf{h} \in \mathbb{N}_0^s$, we have*

$$\frac{1}{n} \sum_{i=0}^{n-1} e^{2\pi\iota\mathbf{h}\cdot\mathbf{u}_i/b} = \begin{cases} 1 & \text{if } \zeta(\mathbf{h}) \in \mathcal{C}_s^*, \\ 0 & \text{otherwise.} \end{cases}$$

In [59] it is shown that the above sum is 0 when $\mathbf{h}$ satisfies $V(\mathbf{h}) \le k-t$.

Proof: If $\mathbf{h} \in \mathcal{C}_s^*$, then $\mathbf{h} \cdot \mathbf{u}_i = 0$ since $(u_{i,1,1}, u_{i,1,2}, \ldots, u_{i,s,1}, u_{i,s,2}, \ldots)$ is in the row space of $\mathbf{C}$ for all $i = 0, \ldots, n-1$, and the result follows easily. If $\mathbf{h} \notin \mathcal{C}_s^*$, then $\mathbf{C}\mathbf{h} = \mathbf{y} \neq \mathbf{0}$, where $\mathbf{y} \in \mathbb{F}_b^k$. We are interested in the scalar product $\mathbf{h} \cdot \mathbf{u}_i$ for $i = 0, \ldots, n-1$. Notice that $\{\mathbf{h} \cdot \mathbf{u}_i, i = 0, \ldots, n-1\} = \{\mathbf{y}^T \mathbf{x}, \mathbf{x} \in \mathbb{F}_b^k\}$, which is the image of $\mathbb{F}_b^k$ under the application of a mapping that corresponds to the multiplication by $\mathbf{y}^T$. Since $\mathbf{y} \neq \mathbf{0}$, the dimension of this image is 1 and the dimension of the kernel of this mapping is thus $k-1$. Hence each element in $\mathbb{F}_b$ has b^{k-1} pre-images in $\mathbb{F}_b^k$ under this mapping, and therefore as a multiset $\{\mathbf{h} \cdot \mathbf{u}_i, i = 0, \ldots, n-1\}$ contains b^{k-1} copies of each element of $\mathbb{F}_b$. Using this and the fact that

$$\sum_{j=0}^{b-1} e^{2\pi\iota j/b} = 0,$$

the result immediately follows. □

Using this lemma, we get the following result, which is analogous to that presented in Proposition 2. It is proved in [69] for the case where $b = 2$ and P_n is a polynomial lattice point set.

Proposition 4 *Let P_n be a digital net in base b. For any function f having an absolutely convergent Walsh series expansion, we have*

$$Q_n - \mu = \sum_{\mathbf{0} \neq \mathbf{h}:\zeta(\mathbf{h}) \in \mathcal{C}_s^*} \tilde{f}(\mathbf{h}).$$

Let $\tilde{P}_n$ be a b-ary shifted digital net in base b, and $\hat{\mu}_{\mathrm{bsh}}$ be the associated estimator. For any square-integrable function f, we have

$$\mathrm{Var}(\hat{\mu}_{\mathrm{bsh}}) = \sum_{\mathbf{0} \neq \mathbf{h}:\zeta(\mathbf{h}) \in \mathcal{C}_s^*} |\tilde{f}(\mathbf{h})|^2, \tag{1.30}$$

and the variance of the MC estimator $\hat{\mu}_{\mathrm{MC}}$ based on n points is given by

$$\mathrm{Var}(\hat{\mu}_{\mathrm{MC}}) = \frac{\sigma^2}{n} = \frac{1}{n} \sum_{\mathbf{0} \neq \mathbf{h} \in \mathbb{N}_0^s} |\tilde{f}(\mathbf{h})|^2.$$

Proof: Assume $\sum_{\mathbf{h} \in \mathbb{N}_0^s} |\tilde{f}(\mathbf{h})| < \infty$. Then we can write

$$\begin{aligned} Q_n - \mu &= \frac{1}{n} \sum_{i=0}^{n-1} f(\mathbf{u}_i) - \tilde{f}(\mathbf{0}) \\ &= \frac{1}{n} \sum_{i=0}^{n-1} \sum_{\mathbf{h} \in \mathbb{N}_0^s} \tilde{f}(\mathbf{h}) e^{2\pi\iota \mathbf{h}\cdot\mathbf{u}_i/b} - \tilde{f}(\mathbf{0}) \\ &= \frac{1}{n} \sum_{\mathbf{h} \in \mathbb{N}_0^s} \tilde{f}(\mathbf{h}) \sum_{i=0}^{n-1} e^{2\pi\iota \mathbf{h}\cdot\mathbf{u}_i/b} - \tilde{f}(\mathbf{0}) \\ &= \sum_{\mathbf{h}:\zeta(\mathbf{h}) \in \mathcal{C}_s^*} \tilde{f}(\mathbf{h}) - \tilde{f}(\mathbf{0}) = \sum_{\mathbf{0} \neq \mathbf{h}:\zeta(\mathbf{h}) \in \mathcal{C}_s^*} \tilde{f}(\mathbf{h}). \end{aligned}$$

If f is square-integrable, then the function $g : [0,1)^s \to \mathbb{R}$ defined by

$$g(\mathbf{u}) = \frac{1}{n} \sum_{i=0}^{n-1} f(\mathbf{u}_i \oplus \mathbf{u}),$$

is also square-integrable, where $\oplus$ corresponds to a digit-by-digit addition in $\mathbb{F}_b$. In addition, $\mathrm{Var}(\hat{\mu}_{\mathrm{bsh}}) = \mathrm{Var}(g(\mathbf{u})) = \sum_{\mathbf{0} \neq \mathbf{h} \in \mathbb{N}_0^s} |\tilde{g}(\mathbf{h})|^2$, because Parseval's equality holds for the Walsh series expansion (see [35], for example). Now, for any $\tilde{\mathbf{h}} \in \mathbb{N}_0^s$, we have

$$\begin{aligned}
\tilde{g}(\tilde{\mathbf{h}}) &= \int_{[0,1)^s} g(\mathbf{u}) e^{-2\pi\iota\tilde{\mathbf{h}}\cdot\mathbf{u}/b} d\mathbf{u} \\
&= \int_{[0,1)^s} \frac{1}{n}\sum_{i=0}^{n-1} f(\mathbf{u}_i \oplus \mathbf{u}) e^{-2\pi\iota\tilde{\mathbf{h}}\cdot\mathbf{u}/b} d\mathbf{u} \\
&= \frac{1}{n}\sum_{i=0}^{n-1} \int_{[0,1)^s} f(\mathbf{v}_i) e^{-2\pi\iota\tilde{\mathbf{h}}\cdot(\mathbf{v}_i\ominus\mathbf{u}_i)/b} d\mathbf{v}_i \\
&= \frac{1}{n}\sum_{i=0}^{n-1} e^{2\pi\iota\tilde{\mathbf{h}}\cdot\mathbf{u}_i/b} \int_{[0,1)^s} f(\mathbf{v}_i) e^{-2\pi\iota\tilde{\mathbf{h}}\cdot\mathbf{v}_i/b} d\mathbf{v}_i \\
&= \frac{1}{n}\sum_{i=0}^{n-1} e^{2\pi\iota\tilde{\mathbf{h}}\cdot\mathbf{u}_i/b} \tilde{f}(\tilde{\mathbf{h}}) \\
&= \begin{cases} \tilde{f}(\tilde{\mathbf{h}}) & \text{if } \zeta(\tilde{\mathbf{h}}) \in \mathcal{C}_s^*, \\ 0 & \text{otherwise.} \end{cases} \qquad (1.31)
\end{aligned}$$

In the above display, the third line is obtained by letting $\mathbf{v}_i = \mathbf{u}_i \oplus \mathbf{u}$, and thus $\mathbf{u} = \mathbf{v}_i \ominus \mathbf{u}_i$, where $\ominus$ denotes a digit-by-digit subtraction in $\mathbb{F}_b$. From (1.31), the result follows. The variance of the MC estimator is obtained by applying Parseval's equality. $\square$

To see the connection with scrambled-type estimators, we use the following result (proved for $b = 2$ in [67]), whose proof is given in the appendix. The result also makes a connection between variance expressions in terms of Haar and Walsh expansions, because $\sigma^2_{\kappa_I}$ is defined in terms of Haar coefficients.

Lemma 3 *Let κ_I be a vector of positive integers and $\sigma^2_{\kappa_I}$ be defined as in (1.26). If f is square-integrable, then*

$$\sigma^2_{\kappa_I} = \sum_{\mathbf{h}\in G(\kappa_I)} |\tilde{f}(\mathbf{h})|^2,$$

where $G(\kappa_I) = \{\mathbf{h} \in \mathbb{N}_0^s : 2^{\kappa_j - 1} \le h_j < 2^{\kappa_j}$ if $j \in I$, $h_j = 0$ otherwise$\}$.

Hence in the digital shift case, in comparison with MC, the contribution of a basis function $\phi_{\mathbf{h}}(\cdot)$ to the variance expression is either multiplied by n (if $\mathbf{h}$ is in the dual space) or by 0, whereas in the scrambled case, this contribution is multiplied by 0 for "small vectors", and by a factor that can be upper-bounded by a quantity *independent* of n otherwise. This factor being sometimes n in the digital shift case prevents us from bounding $\mathrm{Var}(\hat{\mu}_{\mathrm{bsh}})$ by a constant times σ^2. Similarly, the case of smooth functions yields a variance bound in $O(n^{-2}(\log n)^s)$ for digitally-shifted estimators, which is larger by a factor n than the order of the bound obtained for scrambled-type estimators. On the other hand, the b-ary shift is a very simple randomization easy to implement; the estimator $\hat{\mu}_{\mathrm{bsh}}$ can typically be constructed in the same (or less) time as the MC estimator based on the same number of points.

Based on the expression (1.30) for the variance of the digitally shifted estimator, the same type of heuristic arguments as those given for randomly shifted lattice rules can be used to justify selection criteria such as $\Delta_{t_1,\ldots,t_d}$ to choose digital nets. That is, if we assume that the largest Walsh coefficients are those associated with "small" vectors $\mathbf{h}$, then it is reasonable to choose P_n so that the dual space $\mathcal{C}_s^*$ does not contain those small vectors. If we use the norm $\|\mathbf{h}\|$ defined in (1.20) to measure $\mathbf{h}$, this suggests using a criterion based on the resolution such as $\Delta_{t_1,\ldots,t_d}$; if instead we use the norm $V(\mathbf{h})$ defined in (1.18), then the t-value or the variant $\tilde{t}$ defined in (1.16) should be employed. We refer the reader to [42, 112] for additional connections between Walsh expansions and nonuniformity measures (e.g., the so-called 'Walsh-spectral test' of Tezuka). Note that criteria based on the resolution are faster to compute than those based on the t-value, because the latter verifies the $(q_1,\ldots,q_s)$-equidistribution of P_n for a much larger number of vectors $(q_1,\ldots,q_s)$.

7. Transformations of the Integrand

So far, our description of how to use QMC methods can be summarized as follows: Choose a construction and a randomization; choose a selection criterion; find a good point set with respect to this criterion (or use a precomputed table of "good" point sets); randomize the point set, and compute $\hat{\mu}_{\mathrm{RQMC}}$ as an estimator for μ. If the selection criterion mimics the variance of $\hat{\mu}_{\mathrm{RQMC}}$ well enough, one should obtain a low variance estimator with this approach. Most of the selection criteria presented in Section 4 are defined so that they should imitate more or less the variance of $\hat{\mu}_{\mathrm{RQMC}}$ for a large class of functions, i.e., they provide "general-purpose" low-discrepancy point sets. However, once the problem at hand is known, the variance can sometimes be reduced further by making use of information on f in a clever way. In particular, techniques used to reduce the MC variance can also be used in combination with QMC methods. Examples of such techniques are antithetic variates, control variables, importance sampling, and conditional Monte Carlo. These methods can all be seen as transformations applied to f in order to reduce its variability; that is, one replaces f by a function g such that $\int_{[0,1)^s} g(\mathbf{u})d\mathbf{u} = \mu$ and (hopefully) $\int_{[0,1)^s} g^2(\mathbf{u})d\mathbf{u} < \int_{[0,1)^s} f^2(\mathbf{u})d\mathbf{u}$. If the function g requires more computation time for its evaluation, one should make sure that the variance reduction gained is worth the extra effort, i.e., that the *efficiency* is improved.

A second class of methods that can reduce the variance of QMC estimators for certain functions are *dimension reduction* methods. Among this class are the *Brownian bridge technique* of Caflisch and Moskovitz [12], approaches based on principal components [1, 2], and various methods discussed by Fox [33] for generating Poisson and other stochastic processes.

Typically, these methods are used when f is defined in terms of a stochastic process for which a sample path is generated using the uniform numbers $u_{i1}, \ldots, u_{is}$ provided by a point $\mathbf{u}_i$ in $\tilde{P}_n$. The goal is then to generate the sample path in a way that will decrease the effective dimension of f. This is usually achieved by using a method that gives a lot of importance to a few uniform numbers. As an illustration, we describe in the example below the Brownian bridge technique, which can be used to generate the sample paths of a Brownian motion.

Example 3 As this often happens in financial simulations (see e.g., [6, 11]), suppose we want to generate the sample path of a Brownian motion at s different times $B(t_1), \ldots, B(t_s)$, using s uniform numbers $u_1, \ldots, u_s$. For instance, this Brownian motion might be driving the price process of an asset on which an *option* has been written [22]. Instead of generating these observations sequentially (that is, by using u_j to generate the Gaussian random variable $B(t_j)/B(t_{j-1})$ given $B(t_{j-1})$), u_1 is used to generate $B(t_s)$, u_2 is used to generate $B(t_{\lfloor s/2 \rfloor})$, u_3 for $B(t_{\lfloor s/4 \rfloor})$, u_4 for $B(t_{\lfloor 3s/4 \rfloor})$, etc. This can be done easily since for $u < v < w$, the distribution of $B(v)$ given $B(u)$ and $B(w)$ is Gaussian with parameters depending only on u, v, w. By generating the Brownian motion path this way, more importance is given to the first few uniform numbers since they determine important aspects of the path such as its value at the end, middle, first quarter, etc.

Another type of transformation that can sometimes reduce the actual dimension of the problem (and not only the effective one) is the conditional Monte Carlo method; see [65, Section 10.1] for example, where this method is used with randomly shifted lattice rules. This method can also be viewed as a "smoothing technique", i.e., it replaces the integrand by a smoother function (in this case, a conditional expectation) that, e.g., satisfies conditions required to obtain variance rates that are $O(n^{-3} \log^s n)$, as in Section 6.2. We refer the reader to [109] and [33, Chapters 7 and 8] for more on smoothing techniques that can be used in combination with QMC methods.

8. Related Methods

We now discuss integration methods that are closely related to QMC methods, but that do not exactly fit the framework presented so far.

First, a natural extension for the estimator Q_n or $\hat{\mu}_{\mathrm{RQMC}}$ would be to assign weights to the different evaluation points; that is, for a point set $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$, define

$$\hat{\mu}_{\mathrm{w}} = \sum_{i=0}^{n-1} w_i f(\mathbf{u}_i),$$

with $\sum_{i=1}^{n} w_i = 1$. Hickernell [44] proved that when P_n is a lattice point set, then a certain measure of discrepancy (called the *$\mathcal{L}^2$-star discrepancy*), defined so that these weights w_i are accounted for, is minimized by setting the weights to be all equal to $1/n$. In other words, using weights is useless in this case.

However, if P_n is not restricted to have a particular form, then it can be shown that in some cases allowing different weights can bring a significant improvement [121]. For example, one can use the MC method with weights w_i defined by the *Voronoï tessellation* induced by the uniform and random points $\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}$; more precisely, define

$$w_i = \lambda(\{\mathbf{u} \in [0,1)^s : \|\mathbf{u} - \mathbf{u}_i\|_2 \leq \|\mathbf{u} - \mathbf{u}_j\|_2 \text{ for all } j \neq i\}),$$

where λ denotes the Lebesgue measure on $[0,1)^s$. This approach yields an estimator $\hat{\mu}_{\mathrm{w}}$ with variance in $O(n^{-2})$ when $s = 2$. Weighted approximations also based on Voronoï tessellations are discussed in [98].

A closely related idea is used in *stratified sampling* [14]. In this method, the unit hypercube is partitioned into n cells $W_0, \ldots, W_{n-1}$, and $\mathbf{u}_i$ is uniformly distributed over W_i, for $i = 0, \ldots, n-1$. The stratified sampling estimator is then

$$\hat{\mu}_{\mathrm{str}} = \sum_{i=0}^{n-1} w_i f(\mathbf{u}_i),$$

where $w_i = \lambda(W_i)$. It can be shown [14] that for any square-integrable function,

$$\mathrm{Var}(\hat{\mu}_{\mathrm{str}}) \leq \mathrm{Var}(\hat{\mu}_{\mathrm{MC}}).$$

The amount of variance reduction depends on the definition of the cells $W_0, \ldots, W_{n-1}$ and their interaction with the integrand f.

An integration method that is guaranteed to yield an estimator with a variance not larger than the MC estimator for monotone functions is the *Latin Hypercube Sampling* (LHS) [79]. It uses a point set whose unidimensional projections are evenly distributed (i.e., one point per interval $[j/n,(j+1)/n)$, for $j=0,\dots,n-1$). To construct this point set, one needs to generate s random, uniform, and independent permutations $\pi_1,\dots,\pi_s$ of the integers from 0 to $n-1$, and n independent shifts $\Delta_0,\dots,\Delta_{n-1}$ uniformly distributed over $[0,1/n)^s$. Then define the point set

$$\tilde{P}_n = \left\{ \left(\frac{\pi_1(i)}{n} + \Delta_{i,1}, \dots, \frac{\pi_s(i)}{n} + \Delta_{i,s} \right),\ i = 0,\dots,n-1 \right\}.$$

Additional results on this method can be found in, e.g., [4, 96] and the references therein. In particular, see [91, 73] for results showing that the LHS estimator obeys a central-limit theorem. A related method that extends the uniformity property of LHS and has close connections with digital nets is a *randomized orthogonal array design*; we refer the reader to [92, 93, 94, 72, 95] for more on this method.

9. Conclusions and Discussion

We have described various QMC constructions that can be used for multidimensional numerical integration. Measures of quality that can help selecting parameters for a given construction have been presented. We also discussed different randomizations, and provided results on the variance of estimators obtained by applying these randomizations. In particular, we gave a new result that expresses the variance of an estimator based on a digitally shifted net P_n as a sum of squared Walsh coefficients over the dual space of P_n.

In the near future, we plan to compare empirically various constructions and randomizations on practical problems, to study selection criteria and compare their effectiveness, and to investigate in more details the effect of transformations, such as those discussed in Section 7, on the variance of the randomized QMC estimators.

Appendix: Proofs

Proof of Proposition 1: The result is obtained by first generalizing Proposition 5.2 in [69] to arbitrary digital nets in base b. This can be done by using Lemma 2 from Section 6.2. More precisely, we show that P_n is $(q_1,\dots,q_s)$-equidistributed if and only if $\mathcal{C}_s^* \cap \mathcal{H}(q_1,\dots,q_s) = \{\mathbf{0}\}$, where

$$\mathcal{H}(q_1,\dots,q_s) = \{\mathbf{h} \in (\mathbb{F}_b[z])^s \text{such that } v(h_j(z)) \le q_j \text{ for each } j\}.$$

Consider the class $\mathcal{F}(q_1,\dots,q_s)$ of all real-valued functions that are constant on each of the b^q b-ary boxes in the definition of $(q_1,\dots,q_s)$-equidistribution. Clearly, P_n is

$(q_1, \ldots, q_s)$-equidistributed if and only if the corresponding point set integrates every function $f \in \mathcal{F}$ with zero error. But due to its periodic structure, each function $f \in \mathcal{F}(q_1, \ldots, q_s)$ has a Walsh expansion of the form

$$f(\mathbf{u}) = \sum_{\mathbf{h} \in H(q_1, \ldots, q_s)} \tilde{f}(\mathbf{h}) e^{2\pi\iota \mathbf{h}\cdot\mathbf{u}/b}, \tag{1.A.1}$$

where $H(q_1, \ldots, q_s) = \{\mathbf{h} \in \mathbb{N}_0^s : \zeta(\mathbf{h}) \in \mathcal{H}(q_1, \ldots, q_s)\}$, and ζ is the bijection between $\mathbb{N}_0^s$ and $(\mathbb{F}_b[z])^s$ mentioned on page 37. To see this, note that any $f \in \mathcal{F}$ can be written as

$$f(\mathbf{u}) = \sum_{v_1=0}^{b^{q_1}-1} \cdots \sum_{v_s=0}^{b^{q_s}-1} c_{v_1,\ldots,v_s} \prod_{j=1}^{s} \mathbf{1}_{v_j b^{-q_j} \le u_j < (v_j+1)b^{-q_j}}, \tag{1.A.2}$$

where the $c_{v_1,\ldots,v_s}$ are real numbers. When $\mathbf{h} \notin H(q_1, \ldots, q_s)$, there exists $j \in \{1, \ldots, s\}$ and an integer $w_j \ge q_j$ such that $h_j > 2^{w_j}$. Let $d = w_j + 1 - q_j$, and recall that $h_j \cdot u = \sum_{l=1}^{w_j+1} h_{j,l} u_l$, where the coefficients $h_{j,l}$ and u_l come from the representation $h_j = \sum_{l=1}^{w_j+1} h_{j,l} b^{l-1}$ and $u = \sum_{l=1}^{\infty} u_l b^{-l}$, respectively. When u_j goes from $v_j b^{-q_j}$ to $(v_j+1)b^{-q_j}$, l goes from 0 to $b^d - 1$ in $h_j \cdot (v_j b^{-q_j} + l b^{-w_j-1})$, and this dot product is then equal to each number between 0 and $b-1$ exactly b^{d-1} times. Hence, if we first integrate $f(\mathbf{u})$ with respect to u_j when computing $\tilde{f}(\mathbf{h})$ via (1.29), any term from the sum (1.A.2) will be 0 because

$$\begin{aligned} & \int_0^1 c_{v_1,\ldots,v_s} \mathbf{1}_{v_j b^{-q_j} \le u_j < (v_j+1)b^{-q_j}} e^{-2\pi\iota h_j \cdot u_j/b} du_j \\ = \; & c_{v_1,\ldots,v_s} \sum_{l=0}^{b^d-1} e^{-2\pi\iota h_j\cdot(v_j b^{-q_j} + l 2^{-w_j-1})/b} \; = \; 0. \end{aligned}$$

Therefore, $\tilde{f}(\mathbf{h}) = 0$ if $\mathbf{h} \notin H(q_1, \ldots, q_s)$, and (1.A.1) follows. Now, for any nonzero $\tilde{\mathbf{h}} \in H(q_1, \ldots, q_s)$, $g(\mathbf{u}) \equiv e^{-2\pi\iota\tilde{\mathbf{h}}\cdot\mathbf{u}/b}$ is in $\mathcal{F}(q_1, \ldots, q_s)$. Hence by Proposition 4, the error obtained by using P_n to integrate g is $\sum_{\mathbf{0}\ne\mathbf{h}:\zeta(\mathbf{h})\in\mathcal{C}_s^*} \tilde{g}(\mathbf{h}) = \mathbf{1}_{\zeta(\tilde{\mathbf{h}})\in\mathcal{C}_s^*}$ since the only nonzero Walsh coefficient of g is the one evaluated in $\tilde{\mathbf{h}}$ (and it is equal to 1). From this, we see that if P_n is $(q_1, \ldots, q_s)$-equidistributed, then $\mathcal{C}_s^* \cap \mathcal{H}(q_1, \ldots, q_s) = \{\mathbf{0}\}$. Hence if P_n has a resolution of ℓ_s, then it is $(\ell_s, \ldots, \ell_s)$-equidistributed and therefore $\min_{\mathbf{0}\ne\mathbf{h}\in\mathcal{C}_s^*} \|\mathbf{h}\| > \ell_s$.

We now show that $\min_{\mathbf{0}\ne\mathbf{h}\in\mathcal{C}_s^*} \|\mathbf{h}\| \le \ell_s + 1$, which will prove the result. Since the resolution is ℓ_s, it means that P_n is not $(\ell_s+1, \ldots, \ell_s+1)$-equidistributed. Therefore, the $k \times s(\ell_s+1)$ matrix $\mathbf{L}$ formed by concatenating the transposed of the first ℓ_s+1 rows of each generating matrix $\mathbf{C}^j$ has a row space whose dimension is strictly smaller than $s(\ell_s+1)$. Hence there exists a nonzero vector $\mathbf{x}$ in $\mathbb{F}_b^{s(\ell_s+1)}$ such that $\mathbf{L}\mathbf{x} = \mathbf{0}$. Furthermore, we can assume $\|\mathbf{x}\| = \ell_s + 1$ since $\|\mathbf{x}\| \le \ell_s$ would contradict our assumption that P_n has a resolution of ℓ_s. Define $\mathbf{h} = (h_1(z), \ldots, h_s(z)) \in (\mathbb{F}_b[z])^s$ by

$$h_j(z) = \sum_{l=1}^{\ell_s+1} x_{j,l} z^{l-1}.$$

Since $\mathbf{L}$ is just a truncated version of $\mathbf{C}$ and the coefficients of $h_j(z)$ for powers of z larger than $\ell_s + 1$ are zero for all j, we have that $\mathbf{C}\mathbf{h} = \mathbf{0}$, and therefore $\mathbf{h} \in \mathcal{C}_s^*$ with $\|\mathbf{h}\| = \|\mathbf{x}\| = \ell_s + 1$, which proves the result. $\square$

Proof of Lemma 3: Recall that

$$\sigma^2_{\kappa_I} = \mathrm{Var}(\nu_{\kappa_I}(\mathbf{u})),$$

where $\nu_{\kappa_I}(\mathbf{u})$ is a step function constant on the b-ary boxes obtained by partitioning the jth axis of $[0,1)^s$ in b^{κ_j} equal intervals, for each $j \in I$. Using the same notation as for the preceding proof, we have that $\nu_{\kappa_I} \in \mathcal{F}(q_1, \ldots, q_s)$, where

$$q_j = \begin{cases} \kappa_j & \text{if } j \in I, \\ 0 & \text{otherwise.} \end{cases}$$

Hence from the proof of Proposition 1, we know that $\tilde{\nu}_{\kappa_I}(\mathbf{h}) = 0$ if $\mathbf{h} \notin H(q_1, \ldots, q_s)$.

Assume $\mathbf{h} \in H(q_1, \ldots, q_s)$ and that there exists one $j \in I$ such that $h_j < 2^{\kappa_j - 1}$. We need to verify that $\tilde{\nu}_{\kappa_I}(\mathbf{h}) = 0$. Now,

$$\begin{aligned} \tilde{\nu}_{\kappa_I}(\mathbf{h}) &= \int_{[0,1)^s} \sum_\tau \sum_\gamma \langle \psi_{\kappa_I,\tau,\gamma}, f\rangle \psi_{\kappa_I,\tau,\gamma}(\mathbf{u}) e^{-2\pi\iota \mathbf{h}\cdot\mathbf{u}/b} d\mathbf{u} \\ &= \sum_\tau \sum_\gamma \langle \psi_{\kappa_I,\tau,\gamma}, f\rangle \int_{[0,1)^s} \psi_{\kappa_I,\tau,\gamma}(\mathbf{u}) e^{-2\pi\iota \mathbf{h}\cdot\mathbf{u}/b} d\mathbf{u} \\ &= \sum_\tau \sum_\gamma \langle \psi_{\kappa_I,\tau,\gamma}, f\rangle b^{(|\kappa_I - |I|)/2} \cdot \\ &\quad \left(\prod_{k \in I, k \neq j} \int_0^1 (b \mathbf{1}_{\lfloor b^{\kappa_k+1} u_j \rfloor = b\tau_k + \gamma_k} - \mathbf{1}_{\lfloor b^{\kappa_k} u_k \rfloor = \tau_k}) e^{-2\pi\iota h_k \cdot u_k/b} du_k \right) \\ &\quad \int_0^1 (b \mathbf{1}_{\lfloor b^{\kappa_j+1} u_j \rfloor = b\tau_j + \gamma_j} - \mathbf{1}_{\lfloor b^{\kappa_j} u_j \rfloor = \tau_j}) e^{-2\pi\iota h_j \cdot u_j/b} du_j. \end{aligned} \tag{1.A.3}$$

Now, since $h_j < 2^{\kappa_j - 1}$, the function $e^{-2\pi\iota h_j \cdot u_j/b}$ is constant over any interval of the form $[db^{-\kappa_j+1}, (d+1)b^{-\kappa_j+1}), 0 \le d < b^{\kappa_j - 1}$. Hence, if

$$\int_{db^{-\kappa_j+1}}^{(d+1)b^{-\kappa_j+1}} (b \mathbf{1}_{\lfloor b^{\kappa_j+1} u_j \rfloor = b\tau_j + \gamma_j} - \mathbf{1}_{\lfloor b^{\kappa_j} u_j \rfloor = \tau_j}) du_j = 0, \tag{1.A.4}$$

for any $d \in [0, \ldots, b^{\kappa_j - 1} - 1]$, then (1.A.3) is equal to zero and the result is proved. To show (1.A.4), it suffices to observe that

$$\int_{db^{-\kappa_j+1}}^{(d+1)b^{-\kappa_j+1}} (b \mathbf{1}_{\lfloor b^{\kappa_j+1} u_j \rfloor = b\tau_j + \gamma_j} - \mathbf{1}_{\lfloor b^{\kappa_j} u_j \rfloor = \tau_j}) du_j = b \cdot b^{-\kappa_j - 1} - b^{-\kappa_j} = 0,$$

for any d, which proves the result. □

Acknowledgments

This work was supported by NSERC-Canada individual grants to the two authors and by an FCAR-Québec grant to the first author. We thank Bennett Fox, Fred Hickernell, Harald Niederreiter, and Art Owen for helpful comments and suggestions.

References

[1] P. Acworth, M. Broadie, and P. Glasserman. A comparison of some Monte Carlo and quasi-Monte Carlo techniques for option pricing. In P. Hellekalek and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, number 127 in Lecture Notes in Statistics, pages 1–18. Springer-Verlag, 1997.

[2] F. Åkesson and J. P. Lehoczy. Path generation for quasi-Monte Carlo simulation of mortgage-backed securities. *Management Science*, 46:1171–1187, 2000.

[3] I. A. Antonov and V. M. Saleev. An economic method of computing LP_τ-sequences. *Zh. Vychisl. Mat. i. Mat. Fiz.*, 19:243–245, 1979. In Russian.

[4] A. N. Avramidis and J. R. Wilson. Integrated variance reduction strategies for simulation. *Operations Research*, 44:327–346, 1996.

[5] N. S. Bakhvalov. On approximate calculation of multiple integrals. *Vestnik Moskovskogo Universiteta, Seriya Matematiki, Mehaniki, Astronomi, Fiziki, Himii*, 4:3–18, 1959. In Russian.

[6] P. Boyle, M. Broadie, and P. Glasserman. Monte Carlo methods for security pricing. *Journal of Economic Dynamics & Control*, 21(8-9):1267–1321, 1997. Computational financial modelling.

[7] E. Braaten and G. Weller. An improved low-discrepancy sequence for multidimensional quasi-Monte Carlo integration. *Journal of Computational Physics*, 33:249–258, 1979.

[8] P. Bratley and B. L. Fox. Algorithm 659: Implementing Sobol's quasirandom sequence generator. *ACM Transactions on Mathematical Software*, 14(1):88–100, 1988.

[9] P. Bratley, B. L. Fox, and H. Niederreiter. Implementation and tests of low-discrepancy sequences. *ACM Transactions on Modeling and Computer Simulation*, 2:195–213, 1992.

[10] P. Bratley, B. L. Fox, and H. Niederreiter. Algorithm 738: Programs to generate Niederreiter's low-discrepancy sequences. *ACM Transactions on Mathematical Software*, 20:494–495, 1994.

[11] R. E. Caflisch, W. Morokoff, and A. Owen. Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *The Journal of Computational Finance*, 1(1):27–46, 1997.

[12] R. E. Caflisch and B. Moskowitz. Modified Monte Carlo methods using quasi-random sequences. In H. Niederreiter and P. J.-S. Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, number 106 in Lecture Notes in Statistics, pages 1–16, New York, 1995. Springer-Verlag.

[13] J. Cheng and M. J. Druzdzel. Computational investigation of low-discrepancy sequences in simulation algorithms for bayesian networks. In *Uncertainty in Artificial Intelligence Proceedings 2000*, pages 72–81, 2000.

[14] W. G. Cochran. *Sampling Techniques.* John Wiley and Sons, New York, second edition, 1977.

[15] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups.* Grundlehren der Mathematischen Wissenschaften 290. Springer-Verlag, New York, 3rd edition, 1999.

[16] R. Couture and P. L'Ecuyer. Lattice computations for random numbers. *Mathematics of Computation*, 69(230):757–765, 2000.

[17] R. Couture, P. L'Ecuyer, and S. Tezuka. On the distribution of k-dimensional vectors for simple and combined Tausworthe sequences. *Mathematics of Computation*, 60(202):749–761, S11–S16, 1993.

[18] R. R. Coveyou and R. D. MacPherson. Fourier analysis of uniform random number generators. *Journal of the ACM*, 14:100–119, 1967.

[19] R. Cranley and T. N. L. Patterson. Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis*, 13(6):904–914, 1976.

[20] P. Davis and P. Rabinowitz. *Methods of Numerical Integration.* Academic Press, New York, second edition, 1984.

[21] U. Dieter. How to calculate shortest vectors in a lattice. *Mathematics of Computation*, 29(131):827–833, 1975.

[22] D. Duffie. *Dynamic Asset Pricing Theory.* Princeton University Press, second edition, 1996.

[23] B. Efron and C. Stein. The jackknife estimator of variance. *Annals of Statistics*, 9:586–596, 1981.

[24] K. Entacher. Quasi-Monte Carlo methods for numerical integration of multivariate Haar series. *BIT*, 37:846–861, 1997.

[25] K. Entacher, P. Hellekalek, and P. L'Ecuyer. Quasi-Monte Carlo node sets from linear congruential generators. In H. Niederreiter and J. Spanier, editors, *Monte Carlo and Quasi-Monte Carlo Methods 1998*, pages 188–198, Berlin, 2000. Springer.

[26] H. Faure. Discrépance des suites associées à un système de numération. *Acta Arithmetica*, 61:337–351, 1982.

[27] H. Faure. Variations on $(0, s)$-sequences. *Journal of Complexity*, 2001. To appear.

[28] H. Faure and S. Tezuka. A new generation of $(0, s)$-sequences. To appear, 2001.

[29] U. Fincke and M. Pohst. Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. *Mathematics of Computation*, 44:463–471, 1985.

[30] G. S. Fishman. Multiplicative congruential random number generators with modulus 2^β: An exhaustive analysis for $\beta = 32$ and a partial analysis for $\beta = 48$. *Mathematics of Computation*, 54(189):331–344, Jan 1990.

[31] G. S. Fishman and L. S. Moore III. An exhaustive analysis of multiplicative congruential random number generators with modulus $2^{31} - 1$. *SIAM Journal on Scientific and Statistical Computing*, 7(1):24–45, 1986.

[32] B. L. Fox. Implementation and relative efficiency of quasirandom sequence generators. *ACM Transactions on Mathematical Software*, 12:362–376, 1986.

[33] B. L. Fox. *Strategies for Quasi-Monte Carlo.* Kluwer Academic, Boston, MA, 1999.

[34] I. Friedel and A. Keller. Fast generation of randomized low-discrepancy point sets. In K.-T. Fang, F. J. Hickernell, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2000.* Springer, 2001. To appear.

[35] B. Golubov, A. Efimov, and V. Skvortsov. *Walsh Series and Transforms: Theory and Applications*, volume 64 of *Mathematics and Applications: Soviet Series.* Kluwer Academic Publishers, Boston, 1991.

[36] S. Haber. Parameters for integrating periodic functions of several variables. *Mathematics of Computation*, 41:115–129, 1983.

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

[38] S. Heinrich, F. J. Hickernell, and R.-X. Yue. Integration of multivariate Haar wavelet series. Submitted, 2001.

[39] S. Heinrich, F. J. Hickernell, and R. X. Yue. Optimal quadrature for Haar wavelet spaces. submitted, 2001.

[40] P. Hellekalek. On the assessment of random and quasi-random point sets. In P. Hellekalek and G. Larcher, editors, *Random and Quasi-Random Point Sets*, volume 138 of *Lecture Notes in Statistics*, pages 49–108. Springer, New York, 1998.

[41] P. Hellekalek and G. Larcher, editors. *Random and Quasi-Random Point Sets*, volume 138 of *Lecture Notes in Statistics.* Springer, New York, 1998.

[42] P. Hellekalek and H. Leeb. Dyadic diaphony. *Acta Arithmetica*, 80:187–196, 1997.

[43] F. J. Hickernell. A generalized discrepancy and quadrature error bound. *Mathematics of Computation*, 67:299–322, 1998.

[44] F. J. Hickernell. Lattice rules: How well do they measure up? In P. Hellekalek and G. Larcher, editors, *Random and Quasi-Random Point Sets*, volume 138 of *Lecture Notes in Statistics*, pages 109–166. Springer, New York, 1998.

[45] F. J. Hickernell. Goodness-of-fit statistics, discrepancies and robust designs. *Statistical and Probability Letters*, 44:73–78, 1999.

[46] F. J. Hickernell. What affects accuracy of quasi-Monte Carlo quadrature? In H. Niederreiter and J. Spanier, editors, *Monte Carlo and Quasi-Monte Carlo Methods 1998*, pages 16–55, Berlin, 2000. Springer.

[47] F. J. Hickernell and H. S. Hong. Computing multivariate normal probabilities using rank-1 lattice sequences. In G. H. Golub, S. H. Lui, F. T. Luk, and R. J. Plemmons, editors, *Proceedings of the Workshop on Scientific Computing (Hong Kong)*, pages 209–215, Singapore, 1997. Springer-Verlag.

[48] F. J. Hickernell, H. S. Hong, P. L'Ecuyer, and C. Lemieux. Extensible lattice sequences for quasi-Monte Carlo quadrature. *SIAM Journal on Scientific Computing*, 22(3):1117–1138, 2001.

[49] F. J. Hickernell and H. Wożniakowski. The price of pessimism for multidimensional quadrature. *Journal of Complexity*, 17, 2001. To appear.

[50] E. Hlawka. Funktionen von beschränkter variation in der theorie der gleichverteilung. *Ann. Mat. Pura. Appl.*, 54:325–333, 1961.

[51] E. Hlawka. Zur angenäherten berechnung mehrfacher integrale. *Monatshefte für Mathematik*, 66:140–151, 1962.

[52] W. Hoeffding. A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics*, 19:293–325, 1948.

[53] H. S. Hong and F. H. Hickernell. Implementing scrambled digital sequences. Submitted for publication, 2001.

[54] D. E. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Addison-Wesley, Reading, Mass., third edition, 1998.

[55] N. M. Korobov. The approximate computation of multiple integrals. *Dokl. Akad. Nauk SSSR*, 124:1207–1210, 1959. in Russian.

[56] N. M. Korobov. Properties and calculation of optimal coefficients. *Dokl. Akad. Nauk SSSR*, 132:1009–1012, 1960. in Russian.

[57] G. Larcher. Digital point sets: Analysis and applications. In P. Hellekalek and G. Larcher, editors, *Random and Quasi-Random Point Sets*, volume 138 of *Lecture Notes in Statistics*, pages 167–222. Springer, New York, 1998.

[58] G. Larcher, A. Lauss, H. Niederreiter, and W. Ch. Schmid. Optimal polynomials for (t, m, s)-nets and numerical integration of multivariate Walsh series. *SIAM Journal on Numerical Analysis*, 33(6):2239–2253, 1996.

[59] G. Larcher, H. Niederreiter, and W. Ch. Schmid. Digital nets and sequences constructed over finite rings and their application to quasi-Monte Carlo integration. *Monatshefte für Mathematik*, 121(3):231–253, 1996.

[60] G. Larcher and G. Pirsic. Base change problems for generalized Walsh series and multivariate numerical integration. *Pacific Journal of Mathematics*, 189:75–105, 1999.

[61] P. L'Ecuyer. Uniform random number generation. *Annals of Operations Research*, 53:77–120, 1994.

[62] P. L'Ecuyer. Maximally equidistributed combined Tausworthe generators. *Mathematics of Computation*, 65(213):203–213, 1996.

[63] P. L'Ecuyer. Tables of linear congruential generators of different sizes and good lattice structure. *Mathematics of Computation*, 68(225):249–260, 1999.

[64] P. L'Ecuyer and R. Couture. An implementation of the lattice and spectral tests for multiple recursive linear random number generators. *INFORMS Journal on Computing*, 9(2):206–217, 1997.

[65] P. L'Ecuyer and C. Lemieux. Variance reduction via lattice rules. *Management Science*, 46(9):1214–1235, 2000.

[66] P. L'Ecuyer and F. Panneton. A new class of linear feedback shift register generators. In J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, editors, *Proceedings of the 2000 Winter Simulation Conference*, pages 690–696, Pistacaway, NJ, 2000. IEEE Press.

[67] C. Lemieux. *L'utilisation de règles de réseau en simulation comme technique de réduction de la variance.* PhD thesis, Université de Montréal, May 2000.

[68] C. Lemieux, M. Cieslak, and K. Luttmer. RandQMC user's guide. In preparation, 2001.

[69] C. Lemieux and P. L'Ecuyer. Selection criteria for lattice rules and other low-discrepancy point sets. *Mathematics and Computers in Simulation*, 55(1–3):139–148, 2001.

[70] C. Lemieux and A. B. Owen. Quasi-regression and the relative importance of the ANOVA components of a function. In K.-T. Fang, F. J. Hickernell, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2000.* Springer, 2001. To appear.

[71] R. Lidl and H. Niederreiter. *Introduction to Finite Fields and Their Applications.* Cambridge University Press, Cambridge, revised edition, 1994.

[72] W.-L. Loh. A combinatorial central limit theorem for randomized orthogonal array sampling designs. *Annals of Statistics*, 24:1209–1224, 1996.

[73] W.-L. Loh. On Latin hypercube sampling. *The Annals of Statistics*, 24:2058–2080, 1996.

[74] D. Maisonneuve. Recherche et utilisation des “bons treillis”, programmation et résultats numériques. In S. K. Zaremba, editor, *Applications of Number Theory to Numerical Analysis*, pages 121–201. Academic Press, New York, 1972.

[75] E. Maize. *Contributions to the theory of error reduction in quasi-Monte Carlo methods.* PhD thesis, Claremont Graduate School, Claremont, CA, 1981.

[76] J. Matoušek. On the L2-discrepancy for anchored boxes. *Journal of Complexity*, 14:527–556, 1998.

[77] M. Matsumoto and Y. Kurita. Twisted GFSR generators II. *ACM Transactions on Modeling and Computer Simulation*, 4(3):254–266, 1994.

[78] M. Matsumoto and T. Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, 1998.

[79] M. D. Mckay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, 1979.

[80] H. Morohosi and M. Fushimi. A practical approach to the error estimation of quasi-Monte Carlo integration. In H. Niederreiter and J. Spanier, editors, *Monte Carlo and Quasi-Monte Carlo Methods 1998*, pages 377–390, Berlin, 2000. Springer.

[81] W. J. Morokoff and R. E. Caflisch. Quasi-random sequences and their discrepancies. *SIAM Journal on Scientific Computing*, 15:1251–1279, 1994.

[82] H. Niederreiter. Multidimensional numerical integration using pseudorandom numbers. *Mathematical Programming Study*, 27:17–38, 1986.

[83] H. Niederreiter. Point sets and sequences with small discrepancy. *Monatshefte für Mathematik*, 104:273–337, 1987.

[84] H. Niederreiter. Low-discrepancy and low-dispersion sequences. *Journal of Number Theory*, 30:51–70, 1988.

[85] H. Niederreiter. Low-discrepancy point sets obtained by digital constructions over finite fields. *Czechoslovak Math. Journal*, 42:143–166, 1992.

[86] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1992.

[87] H. Niederreiter and G. Pirsic. Duality for digital nets and its applications. *Acta Arithmetica*, 97:173–182, 2001.

[88] H. Niederreiter and C. Xing. The algebraic-geometry approach to low-discrepancy sequences. In P. Hellekalek, G. Larcher, H. Niederreiter, and P. Zinterhof, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 127 of *Lecture Notes in Statistics*, pages 139–160, New York, 1997. Springer-Verlag.

[89] H. Niederreiter and C. Xing. Nets, (t, s)-sequences, and algebraic geometry. In P. Hellekalek and G. Larcher, editors, *Random and Quasi-Random Point Sets*, volume 138 of *Lecture Notes in Statistics*, pages 267–302. Springer, New York, 1998.

[90] G. Ökten. A probabilistic result on the discrepancy of a hybrid-Monte Carlo sequence and applications. *Monte Carlo methods and Applications*, 2:255–270, 1996.

[91] A. B. Owen. A central limit theorem for Latin hypercube sampling. *Journal of the Royal Statistical Society B*, 54(2):541–551, 1992.

[92] A. B. Owen. Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, 2:439–452, 1992.

[93] A. B. Owen. Lattice sampling revisited: Monte Carlo variance of means over randomized orthogonal arrays. *Annals of Statistics*, 22:930–945, 1994.

[94] A. B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In H. Niederreiter and P. J.-S. Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, number 106 in Lecture Notes in Statistics, pages 299–317. Springer-Verlag, 1995.

[95] A. B. Owen. Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal on Numerical Analysis*, 34(5):1884–1910, 1997.

[96] A. B. Owen. Latin supercube sampling for very high-dimensional simulations. *ACM Transactions of Modeling and Computer Simulation*, 8(1):71–102, 1998.

[97] A. B. Owen. Scrambling Sobol and Niederreiter-Xing points. *Journal of Complexity*, 14:466–489, 1998.

[98] G. Pagès. A space quantization method for numerical integration. *Journal of Computational and Applied Mathematics*, 89:1–38, 1997.

[99] S. Paskov and J. Traub. Faster valuation of financial derivatives. *Journal of Portfolio Management*, 22:113–120, 1995.

[100] G. Pirsic. A software implementation of Niederreiter-Xing sequences. In K.-T. Fang, F. J. Hickernell, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2000.* Springer, 2001. To appear.

[101] G. Pirsic and W. C. Schmid. Calculation of the quality parameter of digital nets and application to their construction. *Journal of Complexity*, 2001. To appear.

[102] I. H. Sloan and S. Joe. *Lattice Methods for Multiple Integration.* Clarendon Press, Oxford, 1994.

[103] I. H. Sloan and L. Walsh. A computer search of rank-2 lattice rules for multidimensional quadrature. *Mathematics of Computation*, 54:281–302, 1990.

[104] I. M. Sobol'. The distribution of points in a cube and the approximate evaluation of integrals. *U.S.S.R. Comput. Math. and Math. Phys.*, 7:86–112, 1967.

[105] I. M. Sobol'. *Multidimensional Quadrature Formulas and Haar Functions.* Nauka, Moskow, 1969. In Russian.

[106] I. M. Sobol'. Uniformly distributed sequences with an additional uniform property. *USSR Comput. Math. Math. Phys. Academy of Sciences*, 16:236–242, 1976.

[107] I. M. Sobol' and Y. L. Levitan. The production of points uniformly distributed in a multidimensional. Technical Report Preprint 40, Institute of Applied Mathematics, USSR Academy of Sciences, 1976. In Russian.

[108] J. Spanier. Quasi-Monte Carlo methods for particle transport problems. In H. Niederreiter and P. J.-S. Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 106 of *Lecture Notes in Statistics*, pages 121–148, New York, 1995. Springer-Verlag.

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

[110] K. S. Tan and P. P. Boyle. Applications of randomized low discrepancy sequences to the valuation of complex securities. *Journal of Economic Dynamics and Control*, 24:1747–1782, 2000.

[111] R. C. Tausworthe. Random numbers generated by linear recurrence modulo two. *Mathematics of Computation*, 19:201–209, 1965.

[112] S. Tezuka. Walsh-spectral test for gfsr pseudorandom numbers. *Communications of the ACM*, 30(8):731–735, Aug 1987.

[113] S. Tezuka. *Uniform Random Numbers: Theory and Practice.* Kluwer Academic Publishers, Norwell, Mass., 1995.

[114] S. Tezuka and P. L'Ecuyer. Efficient and portable combined Tausworthe random number generators. *ACM Transactions on Modeling and Computer Simulation*, 1(2):99–112, 1991.

[115] S. Tezuka and T. Tokuyama. A note on polynomial arithmetic analogue of Halton sequences. *ACM Transactions on Modeling and Computer Simulation*, 4:279–284, 1994.

[116] J. P. R. Tootill, W. D. Robinson, and D. J. Eagle. An asymptotically random Tausworthe sequence. *Journal of the ACM*, 20:469–481, 1973.

[117] B. Tuffin. On the use of low-discrepancy sequences in Monte Carlo methods. Technical Report No. 1060, I.R.I.S.A., Rennes, France, 1996.

[118] B. Tuffin. Variance reduction order using good lattice points in Monte Carlo methods. *Computing*, 61:371–378, 1998.

[119] D. Wang and A. Compagner. On the use of reducible polynomials as random number generators. *Mathematics of Computation*, 60:363–374, 1993.

[120] X. Wang and F. J. Hickernell. Randomized Halton sequences. *Math. Comput. Modelling*, 32:887–899, 2000.

[121] S. Yakowitz, J. E. Krimmel, and F. Szidarovszky. Weighted Monte Carlo integration. *SIAM Journal on Numerical Analysis*, 15:1289–1300, 1978.

[122] S. Yakowitz, P. L'Ecuyer, and F. Vázquez-Abad. Global stochastic optimization with low-discrepancy point sets. *Operations Research*, 48(6):939–950, 2000.