New Perspectives on $(0,s)$-Sequences

Christiane Lemieux and Henri Faure

Abstract Low-discrepancy sequences that have an optimal value of 0 for their *t-parameter* have always been of interest to both theorists and practitioners. However, in practice the Sobol' sequence often performs better than the original $(0,s)$-sequences in prime bases proposed by Faure in 1982, although the former construction does not have an optimal value of 0 for its t-parameter. In this paper, we introduce new ideas that can be used to find improved constructions for $(0,s)$-sequences in prime bases. To do so, we study them within the framework of *generalized Niederreiter sequences*, which was introduced by Tezuka in 1993. We take a closer look at the structure of the corresponding generating matrices, as this helps us to better understand the differences and analogies between the constructions that we are interested in. This study is then used to guide our search for improved $(0,s)$-sequences, which are shown to perform well on a variety of problems.

1 Introduction

Low-discrepancy sequences are the backbone of quasi-Monte Carlo methods. While several families of sequences have been proven to have good theoretical properties, it is often necessary to carefully choose the parameters that determine a given sequence so that the resulting approximations work well in practice. In particular,

Christiane Lemieux
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada

Henri Faure
Institut de Mathématiques de Luminy
Marseille, France

$(0,s)$-sequences are very attractive from a theoretical point of view, because the value of the quality parameter t for this construction is 0, which means their equidistribution is optimal. However, when used in practice—for problems of large dimensions and with a relatively small number of points—they may not work so well if their defining parameters are not chosen carefully, which is the case for Faure sequences [3]. Similarly, for the widely used Sobol' sequence, one needs to choose the so-called *direction numbers*.

Our goal in this paper is to explore a few ideas leading to concrete $(0,s)$-sequence constructions that can be used in practice. We also establish comparisons between different families of constructions, namely the Sobol' sequence, the $(0,s)$-sequences, and generalized Halton sequences. To do so, it is helpful to use the framework of *generalized Niederreiter sequences*, which was proposed by Tezuka in [24].

This paper is organized as follows. In Section 2, we provide background information on low-discrepancy constructions that are based on van der Corput sequences, and describe different families to be studied in this paper. We present generalized Niederreiter sequences in Section 3, but use the language of generating matrices to explain this construction. We believe this offers easier-to-grasp concepts to readers unfamiliar with the mathematical tools used in Tezuka's original definitions. In Section 4, we describe two ideas that each result in concrete recommendations for specific $(0,s)$-sequence constructions. Our parameters can be found on the Internet at [27]. In Section 5, we propose to use the underlying idea at the basis of $(0,s)$-sequences to construct sequences that are extensible in the dimension. We show numerical results on a few different problems in Section 6, and conclude in Section 7 with a summary of our findings.

2 Background information

We start by describing the *generalized van der Corput sequence* [2], which is obtained by choosing a sequence $\Sigma = (\sigma_r)_{r\geq 0}$ of permutations of $Z_b = \{0,1,\ldots,b-1\}$. Then, the nth term of the sequence is defined as

$$S_b^{\Sigma}(n) = \sum_{r=0}^{\infty} \sigma_r\big(a_r(n)\big) b^{-r-1},$$

where $a_r(n)$ is the rth digit of the b-adic expansion of $n-1 = \sum_{r=0}^{\infty} a_r(n)\, b^r$. If the same permutation σ is used for all digits, (i.e., if $\sigma_r = \sigma$ for all $r \geq 0$), then we use the notation S_b^{σ} to denote S_b^{Σ}. The van der Corput sequence in base b is obtained by taking $\sigma_r = I$ for all $r \geq 0$, where I stands for the identity permutation over Z_b.

Another way of enriching the van der Corput sequence is by applying a linear transformation to the digits $a_0(n), a_1(n), \ldots$ before outputting a number between 0 and 1 (see in chronological order [22], [3] and [17]). We call this a *linearly scrambled* van der Corput sequence. For a prime base b, it is obtained by choosing a matrix C with elements in $\mathbb{Z}_b$ and an infinite number of rows and columns, and then

defining the nth term of this sequence as

$$S_b^C(n) = \sum_{r=0}^{\infty} \frac{x_{n,r}}{b^{r+1}} \quad \text{in which} \quad x_{n,r} = \sum_{k=0}^{\infty} c_{r+1,k+1} a_k(n), \tag{1}$$

where $c_{r,k}$ is the kth element on the rth row of C. Note that the second summation is finite and performed in $\mathbb{Z}_b$, but the first one can be infinite and is performed in $\mathbb{R}$, with the possibility that $x_{n,r} = b-1$ for all but finitely many r.

An important particular case is the one where C is a nonsingular upper triangular (NUT) matrix, in which case the first summation is finite. Of course, we obtain the original van der Corput sequence S_b^I with the identity matrix.

Finally, a van der Corput sequence in base b—either generalized or linearly scrambled— can be *shifted* by choosing a number $v \in [0,1)$ with base b expansion

$$v = \sum_{r=0}^{\infty} v_r b^{-r-1}$$

and then adding v_r (modulo b) to $\sigma_r\big(a_r(n)\big)$, or $x_{n,r}$ in $S_b^{\Sigma}(n)$, or $S_b^C(n)$, respectively.

To construct multidimensional sequences in the unit hypercube $I^s = [0,1)^s$, we present two classical approaches based on the one-dimensional case above. The first one is to juxtapose van der Corput sequences in different bases. This is precisely the idea behind the *Halton sequence* [10], whose nth term is defined as $X_n = (S_{b_1}(n), \ldots, S_{b_s}(n))$, where the b_j's, for $j = 1, \ldots, s$, are pairwise coprime. That is, the jth coordinate is defined using S_{b_j}, the van der Corput sequence in base b_j.

A *generalized Halton sequence* is defined by choosing s sequences of permutations $\Sigma_j = (\sigma_{j,r})_{r \geq 0}$, $j = 1, \ldots, s$. The sequence's nth point $X_n \in I^s$ is given by

$$X_n = (S_{b_1}^{\Sigma_1}(n), \ldots, S_{b_s}^{\Sigma_s}(n)),\ n \geq 1, \tag{2}$$

where the b_j's are pairwise coprime bases. These b_j's are typically chosen as the first s prime numbers. In this case, we denote the jth base as p_j.

Another approach for defining multidimensional sequences is to choose a base b and also s matrices $C_1, \ldots, C_s$ with elements in $\mathbb{Z}_b$ — called *generating matrices* — and then form a sequence by juxtaposing the s linearly scrambled van der Corput sequences $S_b^{C_1}, \ldots, S_b^{C_s}$. This is a special case of the *digital sequences* proposed by Niederreiter in his general construction principles [17, p. 306 and 313], [18, p. 63 and 72]. In this paper, for simplicity we focus on prime bases.

As it was the case for the choice of bases in the Halton sequences, the matrices C_j cannot be arbitrary and must be carefully chosen in order to obtain *low-discrepancy sequences*, which are sequences for which the star discrepancy satisfies $D^*(P_N) \in O((\log N)^s)$. (Several authors have a $1/N$ factor when defining $D^*(P_N)$, for instance [14, 16], but here we use the convention from Number Theory, as in [3, 17].)

It has been proved that both the generalized Halton sequences and several classes of digital sequences are low-discrepancy sequences.

While it is possible to obtain bounds on $D^*(P_N)$, computing this quantity turns out to be extremely difficult. However, if we replace the sup norm by the L_2-norm, we obtain discrepancy measures that can be computed in practice. More precisely, in what follows we work with the *L_2–discrepancy* $T(P_N)$, introduced by Morokoff and Caflisch [16], whose square can be computed as [16, p.1263–1264]

$$T^2(P_N) = \sum_{i=1}^{N}\sum_{j=1}^{N}\prod_{k=1}^{s}(1-\max(X_{i,k},X_{j,k}))\min(X_{i,k},X_{j,k}) \\ -N2^{-s+1}\sum_{i=1}^{N}\prod_{k=1}^{s}X_{i,k}(1-X_{i,k})+N^2 12^{-s}, \tag{3}$$

where $X_{i,k}$ is the kth coordinate of the ith point $X_i \in I^s$ of P_N. Hence $T(P_N)$ can be computed in $O(N^2 s)$. Note that the L_2-discrepancy considers boxes not necessarily anchored at the origin in its definition, by contrast with the star L_2-discrepancy (T^* in [16], but still denoted T by many authors), for which a formula is given in [26].

We now introduce the fundamental quality parameter t that measures the quality of a given low-discrepancy sequence. This is achieved by the general concept of (t,s)-sequences in base b, originally introduced by Niederreiter in [17] to give a general framework for various constructions of s-dimensional low-discrepancy sequences and to obtain further ones. In this framework, smaller values of the integer $t \geq 0$ give smaller discrepancies. In order to adapt to important new constructions, Tezuka [24] and, a bit later, Niederreiter and Xing [19] have been led to generalize the original definition by using the *truncation operator*, defined as follows (see [19, p.271] and the examples below, after Prop.1, where the truncation is required).

Truncation: Let $x = \sum_{i=1}^{\infty} x_i b^{-i}$ be a b–adic expansion of $x \in [0,1]$, with the possibility that $x_i = b-1$ for all but finitely many i. For every integer $m \geq 1$, define $[x]_{b,m} = \sum_{i=1}^{m} x_i b^{-i}$ (depending on x via its expansion).

An *elementary interval in base b* is an interval of the form $[ab^{-d},(a+1)b^{-d})$ with integers a,d such that $d \geq 0$ and $0 \leq a < b^d$.

We first consider the one-dimensional case, where we say that a sequence $(X_n)_{n\geq 1}$ (with prescribed b-adic expansions for each X_n) is a $(t,1)$-*sequence in base b* if for all integers $l \geq 0$, $m \geq t$, every elementary interval E with $\lambda(E) = b^{t-m}$ contains exactly b^t points of the point set $\{[X_n]_{b,m}\,; lb^m+1 \leq n \leq (l+1)b^m\}$, where the notation $\lambda(\cdot)$ refers to the Lebesgue measure. The original definition of $(t,1)$-sequences was the same with X_n instead of $[X_n]_{b,m}$ in the definition of the point set above. These sequences are sometimes called $(t,1)$-sequences in the *narrow sense* and the others just $(t,1)$-sequences [19, p. 271]. In the following proposition, we deal only with the most interesting case of $(0,1)$-sequences. The proof can be found in [7].

Proposition 1. *The two generalizations of van der Corput sequences defined above, S_b^Σ and S_b^C—where C is such that every left upper $m \times m$ submatrix is nonsingular for all $m \geq 1$—are $(0,1)$-sequences in base b.*

Here, the truncation is required for sequences S_b^Σ when $\sigma_r(0) = b-1$ for all sufficiently large r and for sequences S_b^C when the matrix C gives digits $x_{n,r} = b-1$ for all sufficiently large r in (1). See [7] for the details.

A typical way to ensure that the condition in the preceding proposition holds is to take C to be an NUT matrix. The corresponding sequence is called an *NUT digital* $(0,1)$*-sequence*. Note that in this case, it is not necessary to resort to the truncation since all summations are finite in the definition.

For multidimensional sequences in base b, the same idea is applied, but now to elementary intervals of the form $E = [a_1/b^{d_1},(a_1+1)/b^{d_1}) \times \cdots \times [a_s/b^{d_s},(a_s+1)/b^{d_s})$. A sequence $(X_n)_{n\geq 1}$ (with prescribed b-adic expansions for each coordinate of X_n) is a (t,s)*–sequence in base* b (in the broad sense) if for all integers $l \geq 0$ and $m \geq t$, every elementary interval E with $\lambda(E) = b^{t-m}$ contains exactly b^t points of the point set $\{[X_n]_{b,m}\,; lb^m + 1 \leq n \leq (l+1)b^m\}$.

For multidimensional sequences defined over different bases, similar definitions exist (see [4]) with sets of bases $B = (b_1,\ldots,b_s)$ and sets of parameters $T = (t_1,\ldots,t_s)$. But unfortunately, so far the only realization of "low-discrepancy" (T,s)-sequences in bases B are generalized Halton sequences for which $T = (0,\ldots,0)$ and bases from B are pairwise coprime. A blend of Halton and $(0,s)$-sequences was also proposed in [4] to obtain extensible sequences, but without further investigation. New interesting work in this area has been done by Hofer et al. [11, 12] in a more general setting. In particular, they give conditions under which their new sequences are uniformly distributed.

For digital sequences in a given base b, the first construction that was defined so that $t = 0$ was given by Faure in [3]. It is obtained by choosing a prime base $b \geq s$ and using generating matrices C_j given by the $(j-1)$th power of the (upper triangular) Pascal matrix P_b in $\mathbb{Z}_b$.

As shown by Tezuka in [25], if we take

$$C_j = A_j P_b^{j-1}, \qquad j = 1,\ldots,s \tag{4}$$

where each A_j is a nonsingular lower triangular (NLT) matrix, then we also get a $(0,s)$-sequence. This family of constructions is called *generalized Faure sequences* in [25]. Note that the truncation is required for such generalized Faure sequences.

The Sobol' sequences [22] are digital sequences in base 2 for which t is not necessarily 0. In fact, a necessary condition for having $t = 0$ is that we must have $b \geq s$ (see [17, Cor. 5.17]). This construction is not as simple as the one from Faure and requires arithmetic operations on polynomials. The generating matrices are obtained by choosing a primitive polynomial $p_j(z)$ in the ring $\mathbb{F}_2[z]$ of polynomials over the finite field $\mathbb{F}_2$, given by $p_j(z) = z^{d_j} + a_{j,1}z^{d_j-1} + \cdots + a_{j,d_j-1}z + 1$, where each $a_{j,l} \in \mathbb{F}_2$ and d_j is $p_j(z)$'s degree. We then need d_j *direction numbers* of the form $v_{j,r} = m_{j,r}/2^r$, where $m_{j,r}$ is an odd integer between 1 and 2^r for $r = 1,\ldots,d_j$. Once these d_j direction numbers are chosen, the following ones are obtained through the recurrence

$$v_{j,r} = a_{j,1}v_{j,r-1} \oplus \cdots \oplus a_{j,d_j-1}v_{j,r-d_j+1} \oplus v_{j,r-d_j} \oplus (v_{j,r-d_j}/2^{d_j}), \tag{5}$$

where $\oplus$ represents the addition of vectors with components in $\mathbb{F}_2$.

The rth column of C_j is then formed by the base 2 expansion of $v_{j,r}$. That is, each direction number is assigned to a column of C_j and fills it with its binary representa-

tion. By the definition of the initial vectors $v_{j,1},\ldots,v_{j,d_j}$ and the recurrence (5) used to obtain the next ones, one can see that each C_j is an NUT matrix. In turn, based on Proposition 1, this implies that each one-dimensional projection of the Sobol' sequence is a $(0,1)$-sequence (as shown also in [22, Remark 3.5]). This also implies that Sobol' sequences do not need truncation in their definition.

In what follows, we want to find ways of selecting simple matrices A_j in (4) so that the resulting constructions are competitive with the Sobol' sequence based on carefully chosen direction numbers. To do so, it is useful to use the framework of generalized Niederreiter sequences, which include these two families of constructions and can therefore help understanding the analogies between them.

3 Framework of generalized Niederreiter sequences

This construction from [24] builds digital sequences in a given base b by first choosing s polynomials $p_1(z),\ldots,p_s(z)$ in $\mathbb{F}_b[z]$, where $e_j=\deg(p_j(z))$. For each $j=1,\ldots,s$, we also need a sequence $y_{j,1}(z),y_{j,2}(z),\ldots$ of polynomials. These polynomials, when reduced modulo $p_j(z)$, must be independent within each group of size e_j. That is the polynomials $y_{j,1}(z) \bmod p_j(z), y_{j,2}(z) \bmod p_j(z),\ldots,y_{j,e_j}(z) \bmod p_j(z)$ must be linearly independent.

We then consider the coefficients $a^{(j)}(k,l,r)$ in the development of

$$\frac{y_{j,l}(z)}{p_j(z)^k}=\sum_{r=w}^{\infty} a^{(j)}(k,l,r)z^{-r}$$

and use them to construct the generating matrices. More precisely, for $l\le e_j$, the lth row of C_j is given by $a^{(j)}(1,l,1),a^{(j)}(1,l,2),\ldots$. That is, the lth row contains the coefficients from development of $y_{j,l}(z)/p_j(z)$ for $l\le e_j$. Then the next block of e_j rows is based on the coefficients in the development of $y_{j,l}(z)/(p_j(z))^2$ and so on.

As shown in [24], a digital sequence constructed in this way has a quality parameter $t=(e_1-1)+\ldots+(e_s-1)$.

The name *generalized Niederreiter sequences* comes from the fact that this construction builds on the principles proposed by Niederreiter for his construction of digital (t,s)-sequences [17], also known as *Niederreiter sequences*, but with a subtle generalization in the way the polynomials $y_{j,l}(z)$ are chosen.

$(0,s)$-sequences revisited

It is easy to see that the generating matrix resulting from $p_j(z)=z-j+1$ and $y_{j,l}=1$ for all j,l is the $(j-1)$th power of the Pascal matrix, P_b^{j-1}. Hence, Faure sequences are generalized Niederreiter sequences with $e_j=1$ for all j so that we find again that they are $(0,s)$-sequences. However, powers of the Pascal matrix have ones on their diagonal, which in turn implies that points from these sequences will

tend to clump along the main diagonal in $[0,1)^s$ initially, and the larger b is, the more time it takes before this behavior goes away.

An important new step was achieved in [5], where it is shown that the L_2-discrepancy T of NUT digital $(0,1)$-sequences only depend on the diagonal entries of the generating matrices. Hence the discrepancy T of one-dimensional projections of Faure sequences can be improved by multiplying on the left the generating matrices P_b^{j-1} by diagonal matrices. For more details, see [6, Corollary 2 and Section 5]. This theoretical result is at the origin of new scramblings for Halton sequences (see [8]) and Faure sequences (see Sections 4 and 5 below).

The generalization proposed by Tezuka [24] amounts to replacing the $y_{j,l}(z)$ by arbitrary constants in $\mathbb{Z}_b$, i.e., multiply P_b^{j-1} on the left by an NLT matrix, as in (4). It has been widely used under the name GFaure, in the software FinDer.

Sobol' sequences revisited

As mentioned in Section 2, here we take $p_j(z)$ to be the jth primitive polynomial in base 2 (sorted in increasing degree). We also need polynomials $y_{j,1}(z),\ldots,y_{j,e_j}(z)$ to initialize each block of e_j rows. Note that the requirement that the "direction numbers" be odd means that $y_{j,1}(z)$ must be of degree e_j-1, $y_{j,2}(z)$ must be of degree e_j-2, and so on, up to $y_{j,e_j}(z)=1$, which must be of degree 0. Hence, each generating matrix C_j is NUT, as mentioned before.

Various proposals for finding good direction numbers have been proposed in the literature [13, 15]. It should be noted that the naive choice of selecting direction numbers so that the first submatrix is the $e_j\times e_j$ identity I_{e_j} is not good. It causes the same kind of sub-optimal behavior as the one resulting from the fact that we have ones on the diagonal of the Pascal matrices for the original Faure $(0,s)$-sequences. Going further, we can think of the direction numbers as a way of performing a deterministic scrambling of the naive Sobol' sequence, using an NUT block-diagonal matrix as the scrambling matrix. More formally, we have the following result.

Proposition 2. *Let C be a generating matrix for the Sobol' sequence based on a primitive polynomial $p(z)$ of degree e, and initial direction numbers given by an NUT $e\times e$ matrix V. Let G be another generating matrix for the Sobol' sequence based on $p(z)$ but now with $V=I_e$. Then we can write $C=DG$, where*

$$D=\begin{pmatrix} V & 0 & \cdots \\ 0 & V & \ddots \\ \vdots & \ddots & \ddots \end{pmatrix}.$$

Proof. Let $p(z)=z^e+a_1z^{e-1}+\cdots+a_{e-1}z+1$. The key point of the proof is to understand the structure of the matrix C. To do so, it is useful to partition it into blocks of $e\times e$ matrices. Let $B_{l,c}$ be the cth such matrix on the lth block of e rows that form C. We will prove that $B_{l,c}$ has the form $VH_{l,c}$ for some matrix $H_{l,c}$ independent

of V, for all $l,c \geq 1$. To do so, we introduce two $e \times e$ matrices Q and F, defined as

$$Q = \begin{pmatrix} 1 & 0 & \dots & 0 \\ a_{e-1} & 1 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ a_1 & a_2 & \dots & 1 \end{pmatrix}, \qquad F = (I_e + R_2)(I_e + R_3)\dots(I_e + R_e),$$

where R_k is an $e \times e$ matrix with the first $k-1$ terms of its kth column given by $a_{k-1}, a_{k-2}, \dots, a_1$ and filled with zeros otherwise.

First, by definition of the Sobol' sequence we have $B_{1,1} = V$. Then, we show by induction that $B_{1,c} = V(QF)^{c-1}$ for any $c > 1$. Hence we need to show that $B_{1,c} = B_{1,c-1}(QF)$. To do so, we apply the recurrence (5) in two steps: first, we take the appropriate combinations of columns from $B_{1,c-1}$ explicitly described by (5). This step is performed through the multiplication by Q. Then, the kth column in $B_{1,c}$, for $k = 2, \dots, e$, is recursively updated by adding the appropriate combination of the $k-1$ preceding columns. One can easily verify that this is achieved by multiplying the matrix $B_{1,c-1}Q(I_e + R_2)\dots(I_e + R_{k-1})$ by $(I_e + R_k)$.

For the next rows of C, one can easily verify that the lth block of e rows of C starts with $l-1$ zero matrices. Then, we have that $B_{l,l} = VF^{l-1}$, for any $l > 1$. This comes from the last term $(v_{j,r-d_j}/2^{d_j})$ in (5), which has the effect of pasting $B_{l-1,l-1}$ into $B_{l,l}$. But at the same time, we need to take into account the effect of the recurrence (5) on $B_{l,l}$. This is achieved by successively multiplying by terms of the form $(I_e + R_k)$, for $k = 2, \dots, e$.

For $c > l$, we have that $B_{l,c} = (B_{l,c-1}Q + B_{l-1,c-1})F$, where the first term $B_{l,c-1}QF$ corresponds to the application of (5) but without the $(v_{j,r-d_j}/2^{d_j})$ term, which is handled separately by the second term $B_{l-1,c-1}F$).

Hence for all $l,c \geq 1$, $B_{l,c} = VH_{l,c}$ for some matrix $H_{l,c}$ independent of V. Therefore, the generating matrix G has submatrices of the form $G_{l,c} = H_{l,c}$, and hence C is obtained by multiplying (from the left) each submatrix $G_{l,c}$ in G by V, which amounts to setting $C = DG$. □

We can use this result to draw an interesting analogy with $(0,s)$-sequences. Indeed, for $(0,s)$-sequences, $e = 1$ and so from the point of view of Proposition 2, finding "good" direction numbers amounts to choosing one factor f_j for each dimension j and then use $A_j = f_j I$, where I is the $\mathbb{N} \times \mathbb{N}$ identity matrix. This is precisely the type of simple scrambling that we choose to focus on in this work. Note that the resulting matrix C_j (4) has diagonal entries given by f_j.

4 New efficient scramblings of $(0,s)$-sequences

Here we propose two ways of constructing a $(0,s)$-sequence. Both suggestions amount to carefully choose the matrices A_j in (4). In both cases, as announced above in our interpretation of Proposition 2, we take A_j to be of the form $A_j = f_j I$, where

f_j is an appropriately chosen factor. The choice of this factor is based on the same ideas as those used in [8] to build generalized Halton sequences. For this reason, we first briefly recall our method from [8].

Generalized Halton sequences from [8]

The approach here is to first build a "short list" of at most 32 multipliers based on the criterion $\theta_p^f(1)/\log p$ described in [6], which is related to a bound on the L_2-discrepancy of S_p^{fI}. More precisely, $\theta_p^f(1)$ is defined as [6, Prop. 2]

$$\theta_p^f(1) = \max_{1\le N\le p}\left(T^2(N,S_p^{fI}) - \frac{N^2}{12p^2}\right),$$

where $T(N,S)$ measures the L_2-discrepancy of the first N points of the sequence S. Multipliers f for which $\theta_p^f(1)$ is small thus give rise to good one-dimensional sequences. While the construction of this short list rests on a nice theoretical result, we use a more pragmatic method to select a multiplier from this list. Our idea is to make sure that two-dimensional projections over nearby indices are well distributed, hence avoiding the most well-known defect of the original Halton sequences [16].

That is, to select a multiplier f_j for the jth coordinate, we use the criterion

$$\tau_j^{W,N_0}(f_1,\ldots,f_{j-1},f) = \max_{1\le l\le W} T(N_0,(S_{p_{j-l}}^{f_{j-l}},S_{p_j}^f)), \tag{6}$$

where W and N_0 have to be chosen, and $f_1 = 1$ since we use $p_1 = 2$. That is, for each candidate f in the short list for p_j, we compute the value T of the L_2-discrepancy for the first N_0 points of the two-dimensional sequence based on the $(j-l)$th and jth coordinates (using the multipliers f_{j-l} chosen for the $(j-l)$th coordinate, and the candidate f under study, respectively), for $l = 1,\ldots,W$, where W is the "window" size of the criterion. Then we keep the worst (largest) of these W values of T as our quality measure for f. The multiplier for p_j is chosen as the one that minimizes $\tau_j^{W,N}(f_1,\ldots,f_{j-1},f)$ among all candidates in the short list.

Construction 1 for $(0,s)$-sequences: GF1

Our first idea is to choose b to be the smallest prime larger or equal to s—as typically done when building $(0,s)$-sequences—and then let $A_j = f_j \times I$ where the multipliers f_j are chosen similarly as in the approach described above for generalized Halton sequences. More precisely, we build a list $\mathscr{L}$ containing the m best multipliers f according to θ_b^f, where m is approximately equal to $b/2$. This is a valid approach since any NUT matrix C with diagonal entries given by f is such that $T(N,S_p^f) = T(N,S_p^C)$ [5]. We set f_1 equal to the best multiplier in $\mathscr{L}$ and then choose, for every $j = 2,\ldots,s$, the multiplier $f_j = f \in \mathscr{L}$ that minimizes

$$\tau_j(N_0,W) = \max_{1\le l\le W} T(N_0,(S_b^{C_{j-l}},S_b^C)),$$

where $C_{j-l} = f_{j-l}P_b^{j-l-1}$ and $C = fP_b^{j-1}$.

The results reported in Section 6 were done with $N_0 = 2500$ and $W = 7$, which are the values used in [8]. As discussed in [8], taking $N_0 = 2500$ is somewhat arbitrary but ensures $N_0 > b$ in our examples and appears to be an appropriate length to discriminate "good" sequences from "bad" sequences. Note that the construction of the "short list" of multipliers is done independently of N_0, as it is based on the criterion $\theta_p^f(1)$, which provides a bound on the discrepancy *for all N*.

Construction 2 for $(0,s)$-sequences: GF2

The second idea starts with a somewhat unusual choice, which is to take the base b to be about twice as large as the dimension s. More precisely, we take b the smallest prime larger than $2s$. A larger base implies a larger choice of multipliers, and so by taking $b \approx 2s$, we can simply set $A_j = f_j I$ as in the first proposal, but with f_j equal to the multiplier f with the jth smallest value for θ_b^f. This is a very simple method, as no search based on two-dimensional projections needs to be performed. One simply needs to order the multipliers according to θ_b^f, which can be done quickly.

5 A construction extensible in the dimension–GF3

Because $(0,s)$-sequences need to be defined in a base $b \ge s$, they do not have the property of being *extensible in the dimension*, which we now define.

Definition 1. A family of constructions is *extensible in the dimension* if, for every $s \ge 1$, an s-dimensional sequence $\{X_i, i \ge 1\}$ in that family can be transformed into an r-dimensional sequence $\{\tilde{X}_i, i \ge 1\}$ in that family, where $r > s$, in such a way that X_i equals the first s coordinates of $\tilde{X}_i$.

From their definition, it is clear that Sobol' and Halton sequences are extensible in the dimension. The problem with $(0,s)$-sequences is that if we choose a base $b \ge s$, then for any $r > b$ we need to choose a new base in order to define a $(0,r)$-sequence, and thus we cannot simply extend each s-dimensional point to an r-dimensional one.

Now, if we weaken the property of $(0,s)$-sequences by introducing another quality parameter to replace t, then we will be able to create a construction that is closely connected to $(0,s)$-sequences, but has the advantage of being extensible in the dimension. So first, we define this new quality parameter, which has similarities with other criteria discussed in, e.g., [14].

Definition 2. For an s-dimensional digital sequence $\{X_1, X_2, \ldots\}$ in base b and integer $k \le s$, its quality parameter t_k, is defined as the smallest value so that each

projection of the form

$$\{(X_{i,j_1}, X_{i,j_2}, \ldots, X_{i,j_r}), i \geq 1\}$$

with $1 \leq r \leq k$, $j_1 < \cdots < j_r$ and $j_r - j_1 < k$, is a (t_k, r)-sequence.

The quality parameter t_k thus focuses on projections over indices $j_1, \ldots, j_r$ that span a range no larger than k. Note that for a (t,s)-sequence, we have $t_s = t$.

The idea is then to fix the base b, and for any dimension $s \geq 1$, construct an s-dimensional digital sequence in base b using the generating matrices

$$C_j = A_j P_b^{j-1}, \qquad j = 1, \ldots, s, \tag{7}$$

where $A_j = f_{(j \bmod m)} I$ and $m \approx b/2$. As for the GF-2 construction, $f_{(j \bmod m)}$ is the multiplier in $\mathbb{Z}_b$ with the $(j \bmod m)$th smallest value of θ_b^f.

Note that the matrices A_j make the period of the sequence $C_1, C_2, \ldots$ longer than that of the sequence $P_b, P_b^1, P_b^2, \ldots$, which equals b since $P_b^j = P_b^{j+lb}$ for any $l \geq 1$. Furthermore, as done in our numerical experiments, we can add a random shift as in Section 2 to these sequences, which completely breaks the periodic behavior of the sequence's coordinates for a given point, much like the use of a shift modulo 1 breaks the periodic behavior of the coordinates of points from a Korobov lattice when the number of points N is smaller than s. We also have the following result, which can be easily proved using the well-known fact that any projection of a $(0,s)$-sequences has a quality parameter $t = 0$.

Proposition 3. *For any $s \geq 1$, the s-dimensional digital sequence in base b based on the generating matrices (7) has a quality parameter $t_b = 0$.*

Note that this construction can handle problems where the dimension is unbounded, because once b and m are chosen and the factors $f \in \mathbb{Z}_b$ are sorted according to θ_b^f, *no extra parameters need to be chosen*. Korobov lattices also have this property, since the only parameter that needs to be chosen is the generator a of the lattice. It should be noted, however, that when $s \geq b$ this construction is not uniformly distributed. Also, because of its periodic behavior and the fact that it has some bad projections, it may give erroneous results when used to integrate certain classes of functions which heavily depend on the subset of variables corresponding to these projections. Hence this construction should be used carefully. A possible class of functions for which it could be proved to do well are those having *finite-order weights*, as defined in [21].

6 Numerical results

In this section, we compare the performance of different low-discrepancy sequences on a few problems. Our main goal is to illustrate the importance of the parameter choice for each construction. Hence we compare the (i) original Halton sequences

(H); (ii) generalized Halton sequence proposed and recommended in [8] (GH); (iii) the naive Sobol' sequence where the direction numbers are set to the unit vectors (S); (iv) the Sobol' sequence with direction numbers from [15] (GS); (v) the original $(0,s)$-sequence from [3] (F), and (vi) the alternatives GF-1 and GF-2 described in Section 4. We refer to the group GH, GS, GF-1 and GF-2 as the "scrambled constructions", and the remaining ones as the "non-scrambled" ones.

For each integration problem f, we show the estimated variance of the average

$$Q_{N,M} = \frac{1}{NM}\sum_{l=1}^{M}\sum_{i=1}^{N} f(X_i^l),$$

where $X_i^l = (X_{i,1}^l,\ldots,X_{i,s}^l)$ is obtained by shifting each of the s individual (generalized/linearly scrambled) van der Corput sequences forming the construction under study by a random shift as described in Section 2, for $l = 1,\ldots,M$. The number of points N considered are of the form $N = 2000k$ for $k = 1,\ldots,50$, and we use $M = 25$ i.i.d. random shifts to estimate the variance. For comparison, we show the estimated variance $\hat{\sigma}^2/NM$ of the Monte Carlo estimator based on NM evaluations, where $\hat{\sigma}^2$ is estimated using a grand total of $100000M$ simulations (or computed exactly when possible). Due to lack of space, we do not show comparative results based on the deterministic error or the average absolute error, as in [8].

We first consider a mortgage-backed security problem from [1]. Here the function f represents the discounted cash-flows paid by this security, which in turns come from a pool of mortgages where the mortgagors have the option of prepaying their entire balance at any time, with no penalty. The cash flows and discount factors both depend on a random interest rate process. The integral $I(f)$ corresponds to the theoretical price for this security and cannot be determined exactly.

As can be seen in Figure 1, the scrambled constructions do significantly better than the non-scrambled ones for this problem. In fact, some of the non-scrambled sequences do *worse* than Monte Carlo.

We then consider a *digital option* problem. This example has been used by Papageorgiou in [20] to show that the Brownian bridge technique—popular in financial problems—does not always help quasi-Monte Carlo methods. Hence for problems like this, it is important to have access to low-discrepancy sequences that work well, even in large dimensions. Here also, as seen in Figure 2, the scrambled constructions clearly do better than the non-scrambled ones.

The next three problems look at test-functions defined over I^s that have been used in the literature [8, 23], and are given by

$$g_1(X) = \prod_{j=1}^{96}(1+0.25(X_j-0.5)); \; g_2(X) = \prod_{j=1}^{75}\frac{|4X_j-2|+(75-j)^2}{1+(75-j)^2};$$

$$g_3(X) = \alpha_{120}\pi^{-60}\cos\left(\sqrt{\frac{1}{2}\sum_{j=1}^{120}[\Phi^{-1}(X_j)]^2}\right),$$

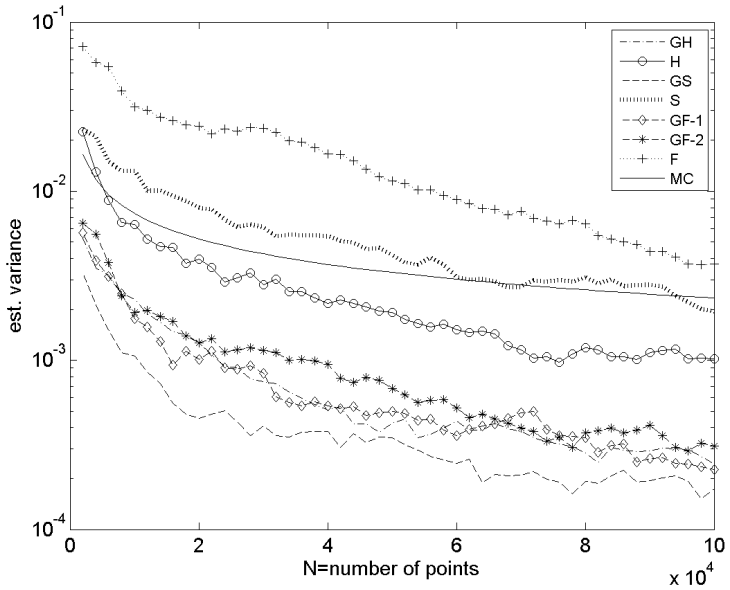

Fig. 1 MBS problem ($s = 360$): the scrambled constructions perform much better than the non-scrambled ones

where α_{120} is such that g_3 integrates to 1 and Φ is the CDF of a standard normal.

The first function is such that constructions with good low-dimensional projections over all indices should do relatively well on this problem, while constructions whose low-dimensional projections deteriorate as they are defined over larger indices should not perform well. Results are shown on Figure 3.

Indeed, as before the scrambled versions (especially Sobol' and the two GF's) do much better than Monte Carlo and the non-scrambled constructions. Note that the naive Sobol' "S" is clearly worse than Monte Carlo.

The second function is such that the coordinates are in increasing order of importance, which means sequences whose coordinates with higher indices are not so well distributed will do badly on this problem. As we see on Figure 4, the non-scrambled sequences have a variance much larger than for the scrambled sequences. It is worth noting that the GF-2 construction, which does not pay attention to two-dimensional projections as GF-1 does, has a variance significantly larger than the other scrambled sequences, although not as bad as the non-scrambled sequences.

Finally, the results shown on Figure 5 for g_3 confirm once again the superiority of the scrambled sequences, this time on a problem where f is not a product. Table 1 summarizes which methods perform best for each function for $N = 100000$ (variance within a factor of two of the smallest variance). We see GS and GF1 are always among the best. The construction GF2 is less competitive, but we are confident that its performance would improve significantly if multipliers were selected with the

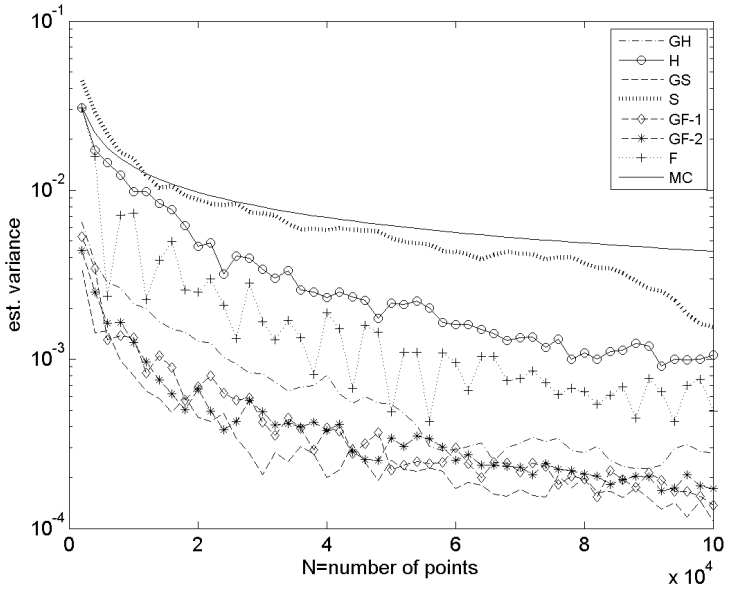

Fig. 2 Digital option with $s = 75$: here also, the scrambled constructions perform much better

help of two-dimensional projections as in GF1 (but still with the larger base), which is in our plans for the near future.

Table 1 Best methods for each function

MBS	Dig	g_1	g_2	g_3
GS,GF1,GH	GS,GF1	GS,GF2,GF1	GH,GS,GF1	GS,GF1,GH

We end with some results showing the performance of the construction described in Section 5, which is extensible in the dimension and can therefore handle a problem in any dimension. The problem considered here is a simple queueing system where clients arrive according to a Poisson process with frequency 1/minute, and the service times are exponentially distributed with mean 55 seconds. We wish to determine the expected number of clients, among the first 1000 ones, who wait more than 5 minutes in the queue before being served. Since we need to generate one interarrival time and one service time per client, the dimension s of this problem is 2000. We compare the performance of the GF-3 construction in base $b = 727$ with an *extensible Korobov lattice* taken from [9] (based on the generator $a = 14471$) and the Monte Carlo method. Results are shown on Figure 6. Both low-discrepancy sequences do much better than Monte Carlo on this problem, and the simple GF-3 sequence is competitive with the extensible Korobov one.

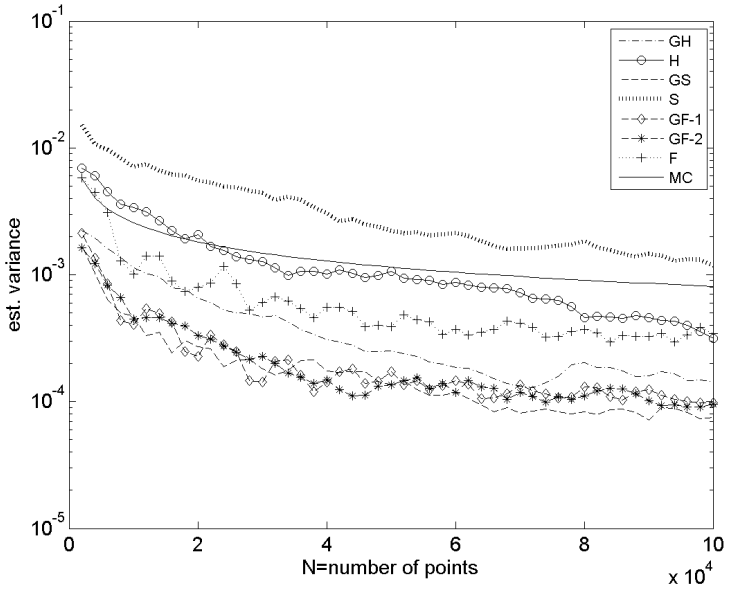

Fig. 3 Function g_1 where $s = 96$: the scrambled sequences GS, GF1 and GF2 are best and some non-scrambled constructions do worse than MC

7 Conclusion

In this paper, we proposed two new ways of choosing parameters for $(0,s)$-sequences that attempt to correct some of the problems encountered in large dimensions by the original construction of Faure. Our approach shares similarities with other ideas used for the Halton and Sobol' sequence. Numerical results on various problems suggest that for all three families, the constructions based on carefully chosen parameters can perform significantly better than the more naive choices. Furthermore, we saw that the new constructions are competitive with the Sobol' sequence, which is popular among practitioners.

We proposed a new construction based on $(0,s)$-sequences where we fix the base b and allow $s > b$. Hence this construction is extensible in the dimension and can handle problems of unbounded dimension. Our numerical results done with $s = 2000$ suggest it is a promising approach, at least for some classes of problems.

For future research, we plan to extend our search for efficient scramblings, for example by building NLT scrambling matrices A_j—not just diagonal—and also by fine-tuning the multipliers for GF-2 and GF-3 so that the resulting sequences have good low-dimensional projections, as done for the GF-1 construction. We would also like to study further the properties of the GF-3 construction, so as to get a better understanding of the classes of functions for which it can perform well.

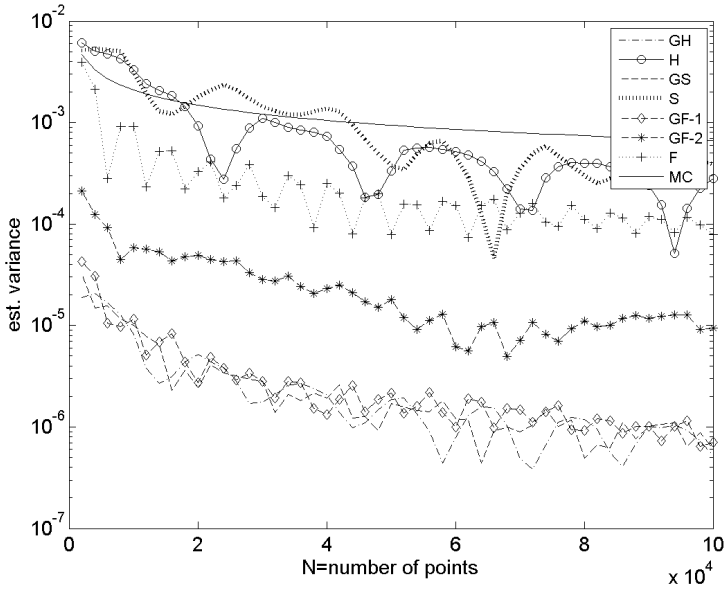

Fig. 4 Function g_2 where $s = 75$: here GF2 does not do as well as GH, GS and GF1, but is still better than the non-scrambled sequences

8 Acknowledgments

We thank the program committee of MCQMC'08 for giving us the opportunity to present this work in Montréal, and the editor and referees for their useful suggestions. The second author also acknowledges the financial support of NSERC.

References

1. Caflisch, R.E., Morokoff, W., Owen, A.B.: Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. J. Comput. Finance **1**(1), 27–46 (1997)
2. Faure, H.: Discrépance des suites associées à un système de numération (en dimension un). Bull. Soc. Math. France **109**, 143–182 (1981)
3. Faure, H.: Discrépance des suites associées à un système de numération (en dimension s). Acta Arit. **41**, 337–351 (1982)
4. Faure, H.: Multidimensional quasi-Monte Carlo methods. Theoret. Comput. Sci. **123**(1), 131–137 (1994)
5. Faure, H.: Discrepancy and diaphony of digital (0,1)-sequences in prime base. Acta Arit. **117**, 125–148 (2005)
6. Faure, H.: Selection criteria for (random) generation of digital $(0,s)$-sequences. In: H. Niederreiter, D. Talay (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2004, pp. 113–126. Springer, New York (2006)

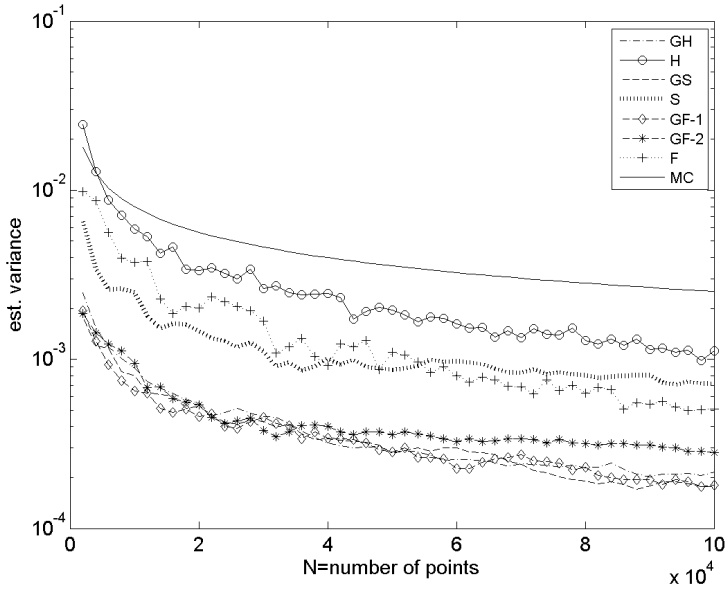

Fig. 5 Function g_3 where $s = 120$: all scrambled sequences do better than the non-scrambled ones, with GF2 slightly less competitive than the others

7. Faure, H.: Van der corput sequences towards general $(0,1)-$sequences in base b. J. Théor. Nombres Bordeaux **19**, 125–140 (2007)
8. Faure, H., Lemieux, C.: Generalized Halton sequences in 2008: A comparative study (2009). To appear in ACM Trans. Model. Comp. Sim.
9. Gill, H.S., Lemieux, C.: Searching for extensible Korobov rules. J. Comp. **23**, 603–613 (2007)
10. Halton, J.H.: On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer. Mathem. **2**, 84–90 (1960)
11. Hofer, R.: On the distribution properties of Niederreiter-Halton sequences. J. Number Th. **129**, 451–463 (2009)
12. Hofer, R., Kritzer, P., Larcher, G., Pillichshammer, F.: Distribution properties of generalized van der Corput-Halton sequences and their subsequences. Int. J. Number Th. (2009). In press.
13. Joe, S., Kuo, F.Y.: Constructing Sobol' sequences with better two-dimensional projections. SIAM J. Scient. Comput. **30**, 2635–2654 (2008)
14. Larcher, G.: Digital point sets: Analysis and applications. In: P. Hellekalek, G. Larcher (eds.) Random and Quasi-Random Point Sets, *LNS*, vol. 138, pp. 167–222. Springer, New York (1998)
15. Lemieux, C., Cieslak, M., Luttmer, K.: RandQMC user's guide: A package for randomized quasi-Monte Carlo methods in C. Tech. Rep. 2002-712-15, Department of Computer Science, University of Calgary (2002)
16. Morokoff, W.J., Caflisch, R.E.: Quasi-random sequences and their discrepancies. SIAM J. Scient. Comput. **15**, 1251–1279 (1994)
17. Niederreiter, H.: Point sets and sequences with small discrepancy. Monats. Math. **104**, 273–337 (1987)
18. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods, *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*, vol. 63. SIAM, Philadelphia (1992)

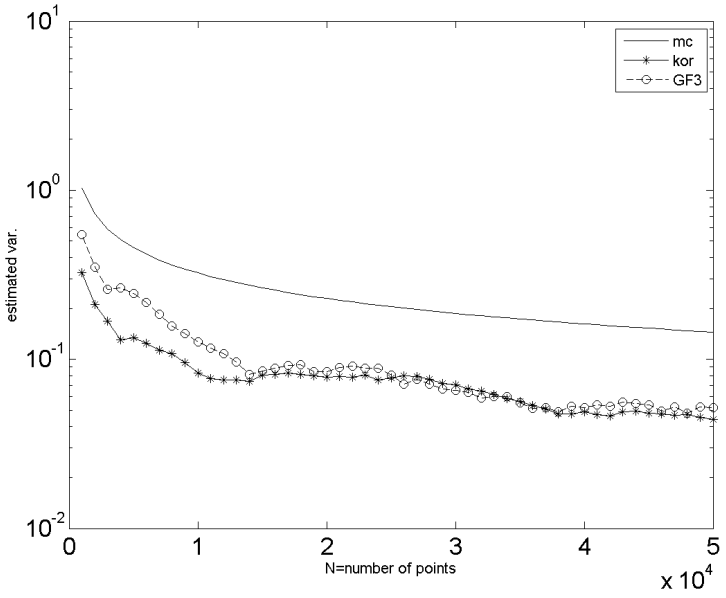

Fig. 6 Queueing problem: $s = 2000$.

19. Niederreiter, H., Xing, C.: Quasirandom points and global function fields. In: Finite fields and applications (Glasgow, 1995), *London Math. Soc. Lecture Note Ser.*, vol. 233, pp. 269–296. Cambridge Univ. Press, Cambridge (1996)
20. Papageorgiou, A.: The Brownian bridge does not offer a consistent advantage in quasi-Monte Carlo integration. J. Comp. **18**(1), 171–186 (2002)
21. Sloan, I.H., Wang, X., Woźniakowski, H.: Finite-order weights imply tractability of multivariate integration. J. Comp. **20**, 46–74 (2004)
22. Sobol', I.M.: On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comput. Math. and Math. Physics **7**, 86–112 (1967)
23. Sobol', I.M., Asotsky, D.I.: One more experiment on estimating high-dimensional integrals by quasi-Monte Carlo methods. Math. Comput. Simul. **62**, 255–263 (2003)
24. Tezuka, S.: Polynomial arithmetic analogue of Halton sequences. ACM Trans. Model. Comp. Sim. **3**, 99–107 (1993)
25. Tezuka, S.: A generalization of Faure sequences and its efficient implementation. Tech. Rep. RT0105, IBM Research, Tokyo Research Laboratory (1994)
26. Warnock, T.: Computational investigations of low discrepancy point sets. In: S.K. Zaremba (ed.) Application de la théorie des nombres à l'analyse numérique, pp. 319–343. Academic Press, New York (1972)
27. www.math.uwaterloo.ca/˜clemieux/0s.html