Generalized Halton Sequences in 2008: A Comparative Study

HENRI FAURE
Institut Mathématiques de Luminy and Université Paul Cézanne
and
CHRISTIANE LEMIEUX
University of Waterloo

Halton sequences have always been quite popular with practitioners, in part because of their intuitive definition and ease of implementation. However, in their original form, these sequences have also been known for their inadequacy to integrate functions in moderate to large dimensions, in which case (t, s)-sequences such as the Sobol' sequence are usually preferred. To overcome this problem, one possible approach is to include permutations in the definition of Halton sequences—thereby obtaining *generalized Halton sequences*—an idea that goes back to almost thirty years ago, and that has been studied by many researchers in the last few years. In parallel to these efforts, an important improvement in the upper bounds for the discrepancy of Halton sequences has been made by Atanassov in 2004. Together, these two lines of research have revived the interest in Halton sequences. In this paper, we review different generalized Halton sequences that have been proposed recently, and compare them by means of numerical experiments. We also propose a new generalized Halton sequence which, we believe, offers a practical advantage over the surveyed constructions, and that should be of interest to practitioners.

Categories and Subject Descriptors: F.2.1 [**Numerical Algorithms and Problems**]: Number-theoretic computations; G.1.4 [**Quadrature and Numerical Differentiation**]: Multidimensional (multiple) quadrature

General Terms: Algorithms,Theory, Performance

Additional Key Words and Phrases: Halton sequences, permutations, scrambling, discrepancy.

1. INTRODUCTION

Halton sequences [Halton 1960] are the oldest multidimensional quasi-random sequences. They have always been popular with practitioners, mainly because their definition is simple, and they are easy to implement. However, it is also well-known that their behavior starts to deteriorate quickly in higher dimensions, which makes them useless for typical real-life problems.

For a long time, the theoretical behavior of these sequences, as measured by

Authors' address: C. Lemieux, Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada. H. Faure, Institut de Mathématiques de Luminy, UMR 6206 CNRS, 163 Av. de Luminy, case 907, 13288 Marseille cedex 9, France.

the concept of *discrepancy*, was also thought to be quite bad in large dimensions. More precisely, for a while the best known upper bounds on the component c_s of the discrepancy bound that is independent of the number of points was increasing exponentially fast with the dimension s for Halton sequences. By contrast, other types of quasi-random sequences, such as the constructions proposed in Faure [1982], Niederreiter [1988], and Niederreiter and Xing [1996], are such that this quantity c_s goes to 0 exponentially fast with the dimension s.

In an effort to improve the behavior of Halton sequences, several researchers have studied various ways of generalizing their definition by including permutations, chosen either deterministically or randomly. This idea goes back to about thirty years ago [Braaten and Weller 1979; Faure 1978], and has been studied by several authors since then. In parallel to these efforts, Atanassov [2004] has provided a significant improvement on the discrepancy upper bounds of the original Halton sequence, making c_s go to 0 with the dimension s rather than growing exponentially fast with s. Moreover, in the same paper, Atanassov produced generalized Halton sequences (by means of the so-called "admissible integers") for which the component c_s of the discrepancy bounds has an even better asymptotic behavior than the one for the original Halton sequences.

Together, these two lines of research have revived the interest in Halton sequences and their generalizations. The goal of this paper is two-fold. First, given all the recent work done in this area by various authors, we believe an up-to-date survey of the most recent generalized Halton sequences available to practitioners is in order. We provide this along with an extensive numerical study comparing these different constructions on several problems. Second, while studying the available constructions, it appeared to us that there was room for improvement. Namely, we think it would be useful to practitioners to have access to a generalized Halton sequence that is (i) based on sound theoretical justifications; (ii) based on an explicit method to choose the required parameters; and (iii) whose performance is as good as—or even better than—the other available generalized Halton sequences. In an effort to fulfill this need, we also provide a new construction that satisfies these requirements, and whose usefulness in practice thus appears to be promising.

The rest of the paper is organized as follows: in Section 2, we provide the required background on Halton sequences, their generalizations, define the different concepts of discrepancy used in our work, make some comparisons with the so-called (t, s)-sequences, and review the special class of permutations used in our new construction. In Section 3, we describe several generalized Halton sequences that have been proposed over the last few years. Our new construction is described in Section 4, along with its parameters for the first 50 dimensions. Section 5 describes the performance measures that we chose to use to compare the different constructions. In particular, we discuss the use of randomizations to estimate the error and variance of an approximation, in the context of quasi-Monte Carlo integration. The description of our numerical experiments is given in Section 6, along with an excerpt of representative graphs illustrating the performance of the different constructions. The online appendix contains the entire set of graphs for all the experiments that we conducted, and also the list of parameters required to implement our new sequence up to 360 dimensions. We conclude in Section 7 with a summary of our

findings and some recommendations.

2. BACKGROUND INFORMATION

We start with a review of Halton sequences and their generalizations, which constitute one type of *quasi-random sequences*. Generally speaking, quasi-random sequences are formed by points in the s-dimensional hybercube $I^s = [0,1)^s$, and are designed to achieve a high level of uniformity, as measured by the concept of discrepancy, which we discuss next. Quasi-random (or low-discrepancy) sequences can then be used for numerical tasks such as integration, and can thus provide an alternative to the random sampling used by Monte Carlo methods. In that context, using such sequences is referred to as *quasi-Monte Carlo* (QMC). As we will see later in our numerical experiments, carefully constructed generalized Halton sequences can provide more accurate approximations than Monte Carlo for several types of integration problems.

2.1 Halton sequences and their generalizations

The definition of the Halton sequence makes use of the *van der Corput sequence in base b*, denoted S_b, which has its nth term ($n \geq 1$) defined as

$$S_b(n) = \sum_{r=0}^{\infty} \frac{a_r(n)}{b^{r+1}}, \tag{1}$$

where $a_r(n)$ is the r-th digit of the b-adic expansion of

$$n - 1 = \sum_{r=0}^{\infty} a_r(n)\ b^r.$$

The *Halton sequence* is an s-dimensional sequence $X_1, X_2, \ldots$ in I^s defined as

$$X_n = (S_{b_1}(n), \ldots, S_{b_s}(n)), \tag{2}$$

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. So, for instance, with $s = 2$ the Halton sequence with $b_1 = 2$ and $b_2 = 3$ starts as $(0,0), (1/2, 1/3), (1/4, 2/3), (3/4, 1/9), (1/8, 4/9)$ and so on. Closely connected to Halton sequences are the *Hammersley point sets*, which are sets of size N defined as

$$P_N = \left\{ \left(\frac{n-1}{N}, S_{b_1}(n), \ldots, S_{b_{s-1}}(n) \right), n = 1, \ldots, N \right\}.$$

A *generalized van der Corput sequence* 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

$$S_b^{\Sigma}(n) = \sum_{r=0}^{\infty} \frac{\sigma_r\big(a_r(n)\big)}{b^{r+1}}. \tag{3}$$

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 defined in (1) is obtained by taking $\sigma_r = I$ for all $r \geq 0$, where I stands for the identity permutation over Z_b.

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

$$X_n = (S_{b_1}^{\Sigma_1}(n), \ldots, S_{b_s}^{\Sigma_s}(n)), \ n \geq 1, \tag{4}$$

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.

2.2 Discrepancy

To measure the quality of the above constructions, the concept of *discrepancy* is often used. The discrepancy of a point set $P_N = \{X_1, \ldots, X_N\} \subseteq I^s$ measures by how much the empirical distribution induced by P_N deviates from the uniform distribution over $[0,1)^s$. As a building block for this measure, we consider for a subinterval of I^s of the form $J = \prod_{j=1}^s [y_j, z_j)$, where $0 \leq y_j < z_j \leq 1$, the difference

$$E(J; N) = A(J; N) - NV(J),$$

where $A(J; N) = \#\{n; 1 \leq n \leq N, X_n \in J\}$ is the number of points in P_N that fall in the subinterval J, and $V(J) = \prod_{j=1}^s (z_j - y_j)$ is the volume of J. We use the notation J^* to denote subintervals J that have a corner at the origin. That is, J^* is of the form $J^* = \prod_{j=1}^s [0, z_j)$, where $0 \leq z_j \leq 1$.

Then, the *star (extreme) discrepancy* D^* and the *(extreme) discrepancy* D of P_N are defined by

$$D^*(P_N) = \sup_{J^*} |E(J^*; N)| \quad \text{and} \quad D(P_N) = \sup_J |E(J; N)|.$$

Note that several authors have a $1/N$ factor when defining the above quantities, for instance Niederreiter [1992].

Since the extreme discrepancy $D(P_N)$ measures the same thing as the star discrepancy $D^*(P_N)$ but over more subintervals J, it is clear that we always have $D(P_N) \geq D^*(P_N)$. It is well known that $D(P_N) \leq 2^s D^*(P_N)$ (see for instance Niederreiter [1992]). For an infinite sequence X, we denote by $D(N, X)$ and $D^*(N, X)$ the discrepancies of its first N points. A sequence satisfying $D^*(N, X) \in O((\log N)^s)$ is typically considered to be of low-discrepancy. Relations between sequences and point sets can be found in Niederreiter [1992, Lemma 3.7] and Faure [1986, Section III], but will not be discussed in this paper.

Computing $D(P_N)$ and $D^*(P_N)$ for dimensions $s > 2$ is usually not feasible in practice. Instead, one can look at the asymptotic behavior—both as N and s increases—of upper bounds on these quantities to assess the quality of a given construction. This will be done in the next two subsections. In addition, the star-discrepancy shows up in upper bounds for the error of the approximation

$$Q_N = \frac{1}{N} \sum_{i=1}^N g(X_i)$$

for the integral $I(g) = \int_{I^s} g(x)dx$ of a function $g : I^s \to R$ whose *variation in the sense of Hardy and Krause*, denoted $V(g)$, is bounded [Niederreiter 1992]. Namely, the Koksma-Hlawka inequality gives the following upper bound on the error of Q_N:

$$E_N = |Q_N - I(g)| \leq \frac{1}{N} D^*(P_N) V(g),$$

when $V(g) < \infty$. Hence a low-discrepancy sequence provides an approximation Q_N for $I(g)$ whose error E_N is in $O(N^{-1}(\log N)^s)$ for functions of bounded variation.

Going back to the definition of discrepancy, it turns out that if we replace the sup norm by the L_2-norm in the star and extreme discrepancy, we obtain discrepancy measures that can be computed in practice. More precisely, let the *star L_2–discrepancy* $T^*(P_N)$ be defined as

$$T^*(P_N) = \Big(\int_{I^s} \Big(E(\textstyle\prod_{j=1}^s [0, y_j); N) \Big)^2 \, dy_1 \dots dy_s \Big)^{\frac{1}{2}},$$

and the *L_2–discrepancy* $T(P_N)$ be defined as

$$T(P_N) = \Big(\int_{\{(y,z) \in I^{2s};\ y_j < z_j\}} \Big(E(\textstyle\prod_{j=1}^s [y_j, z_j); N) \Big)^2 \, dy_1 dz_1 \dots dy_s dz_s \Big)^{\frac{1}{2}}.$$

A formula to compute $T^*(P_N)$ is given in Warnock [1972], while $T(P_N)$ was introduced in Morokoff and Caflisch [1994], where a formula is also given. Since the L_2-discrepancy measure is used in the construction of our new generalized Halton sequence, we give their formula below (for the square $T^2(P_N)$ of $T(P_N)$), where $X_{i,k}$ denote the kth coordinate of the ith point $X_i \in I^s$ of P_N:

$$\begin{aligned} T^2(P_N) \ = \ & \sum_{i=1}^N \sum_{j=1}^N \prod_{k=1}^s (1 - \max(X_{i,k}, X_{j,k})) \min(X_{i,k}, X_{j,k}) \\ & - N 2^{-s+1} \sum_{i=1}^N \prod_{k=1}^s X_{i,k}(1 - X_{i,k}) + N^2 12^{-s}. \end{aligned} \tag{5}$$

Hence $T(P_N)$ can be computed in $O(N^2 s)$.

As we mentioned in the introduction, the discrepancy of Halton sequences was thought for a long time to behave quite badly in large dimensions. In our discussion of this topic, we will look at bounds of the form

$$D(N, X) \le c_s (\log N)^s + O((\log N)^{s-1}) \tag{6}$$

for low-discrepancy sequences. In what follows, we discuss the progress made on understanding the behavior of c_s for Halton sequences.

First, in Faure [1982], it was proved that for the original Halton sequence $X = (X_n)_{n \ge 1}$ as defined in (2), the bound (6) holds with

$$c_s = \prod_{j=1}^s \frac{b_j - 1}{\log b_j}. \tag{7}$$

This improved the original (loose) bound found by Halton [1960], and then Meijer in 1968. For a generalized Halton sequence X, the same bound was shown to hold in Faure [1986, Section 5.4]. Then, Atanassov [2004] improved this result to

$$c_s = \frac{1}{s!} \prod_{j=1}^s \frac{b_j - 1}{\log b_j} =: e^{\lambda_s}, \tag{8}$$

for the original Halton sequence, where we introduced the quantity λ_s so that later, we can use $\log c_s$ when comparing different constructions. With the additional

term $s!$ in the denominator, c_s now goes to 0 as $s \to \infty$. Also, it is quite obvious from the proof given in Atanassov [2004, Lemma 3.1, p.19] that this result also holds for generalized Halton sequences built with arbitrary sequences of permutations. Furthermore, in the same paper, Atanassov showed that for a generalized Halton sequence defined by taking the b_j's to be the first s prime numbers, and by defining the sequence of permutations $\Sigma_j = (\sigma_{j,r})_{r\geq 0}$ using the so-called *admissible integers*—which we will define in more detail in Section 3—then an improved value for c_s holds in (6), and is given by

$$c_s = \frac{2^s}{s!}\left(\sum_{j=1}^{s} \log p_j\right) \prod_{j=1}^{s} \frac{p_j(1+\log p_j)}{(p_j-1)\log p_j} =: e^{\mu_s}, \tag{9}$$

where, as stated at the end of Section 2.1, we use the notation p_j to denote the jth base, which is given by the jth prime number in this case. In the next subsection, we refer to the construction that satisfies this as the *Halton-Atanassov sequence*.

2.3 Comparison with (t, s)-sequences

Digital (t, s)*-sequences* are another family of low-discrepancy sequences that include the well-known Sobol', Faure, and Niederreiter sequences. Niederreiter [1987] developed the general theory of (t, s)-sequences in arbitrary base b, including the original Sobol' and Faure sequences, and he achieved with Xing the best order of magnitude for the quality parameter t with respect to the dimension ($t = O(s)$, see Niederreiter and Xing [1996]). These sequences also use van der Corput sequences as a building block, but with the same base b for each coordinate. Transformations are then performed by means of well-chosen *generating matrices*.

Since we do not include these sequences in our numerical comparisons given in Section 6, we will not give more details about how they are constructed. However, we believe it is of interest to compare the behavior of the discrepancy bounds of these two families in light of the recent results of Atanassov [2004], just so that the reader can understand how these results have revived the interest for the family of Halton sequences, and allowed them to not be ruled out because of what was thought to be a suboptimal behavior for their discrepancy. Furthermore, since all these sequences can be categorized as "low-discrepancy sequences", i.e., they satisfy $D^*(N, X) \in O((\log N)^s)$, one way to compare them is to look at the behavior of the quantity c_s in the bound (6) as the dimension s increases.

First, for the Sobol' sequences [Sobol' 1967], which are (t, s)-sequences in base 2,

$$c_s = \frac{2^t}{s!}\Big(\frac{2}{\log 2}\Big)^s =: e^{\tau_s}. \tag{10}$$

Next, for the $(0, s)$-sequences in prime bases $b \geq s$ proposed by Faure, it is shown in Faure [1982] that

$$c_s = \frac{1}{s!}\Big(\frac{b-1}{\log b}\Big)^s =: e^{\rho_s}. \tag{11}$$

This result also applies to Faure sequences whose coordinates are "scrambled" by permutations of the digits in the $b-$adic expansion on N, in a similar fashion to those used to define the generalized van der Corput sequence (3).

Next, for (t, s)-sequences in arbitrary bases b (hence for Niederreiter sequences), it is proved that [Niederreiter 1988]

$$c_s = \frac{b^t}{s!}\frac{b-1}{2\lfloor \frac{b}{2} \rfloor}\left(\frac{2\lfloor \frac{b}{2} \rfloor}{\log b}\right)^s =: e^{\nu_s}. \tag{12}$$

When $b \geq s$ is an odd prime, in which case t can be minimal ($t = 0$), Equation (12) reduces to (11). But when $b = 2$, (12) improves (10) by a factor $\frac{1}{2}$. Slightly better constants hold for $2 \leq s \leq 4$ (see Niederreiter [1988]). As it was the case for (11), (12) also applies to variants of these constructions in which the coordinates have been scrambled by permutations of the digits (see Niederreiter [1992], Niederreiter and Xing [1996], Owen [1995], Tezuka [1993], Tezuka [1995]). Note also that in these constructions, the base b and the dimension s can be chosen independently from each other, but the parameter t depends on b and s. While $t \in O(s \log s)$ for the Sobol' and Niederreiter sequences in a given base b, $t \in O(s)$ for the Niederreiter-Xing sequences studied in Niederreiter and Xing [1996]. For the Faure sequence, $t = 0$, but we must have $b \geq s$, and thus the base b depends on the dimension s. In what follows, when looking at the behavior of ν_s we will assume that we are working with Niederreiter-Xing sequences.

Finally, Kritzer [2006] recently improved the constants c_s in (11) and (12) by a factor $1/2$ for odd $b \geq 3$ and $s \geq 2$, and by a factor $1/3$ for $b = 2$ and $s \geq 5$ (a similar result holds for even b). Table I takes into account these new values when computing the constants c_s.

There are two possibilities for comparing the quantities c_s given in Sections 2.2 and 2.3 and denoted $e^{\lambda_s}, e^{\mu_s}, e^{\tau_s}, e^{\rho_s}, e^{\nu_s}$ for the different low-discrepancy sequences of interest: computing c_s for different values of the dimension s, or search for asymptotic estimates as s grows to infinity.

For this second approach, using number-theoretic arguments that we shall not develop here due to space constraints (see the online appendix), one can show that

$$-s\log(1.5) < \lambda_s < -\log s + O(1), \tag{13}$$
$$-s\log s < \mu_s < -s\log s + O(s), \tag{14}$$
$$-s\log\log s < \rho_s < -s\log\log s + O(s), \tag{15}$$
$$\tau_s < s\log\log s + O(s), \tag{16}$$
$$-s\log s < \nu_s < -s\log s + O(s). \tag{17}$$

Therefore, from the point of view of asymptotic upper bounds for $\log c_s$, the best sequences are Halton-Atanassov and Niederreiter-Xing sequences, with order $-s\log s$, followed by Faure $(0, s)$-sequences, with order $-s\log\log s$, Halton sequences, with order $-\log s$, and finally the Sobol' sequences, with order $s\log\log s$. Hence, this shows that generalized Halton sequences can have bounds on their discrepancy that are asymptotically as good as the Niederreiter-Xing sequences, which as mentioned before, are designed so that the quality parameter t behaves optimally as a function of s. By constrast, the value of c_s given in (7)—so what was the best bound prior to the work of Atanassov—is such that $\log c_s \in O(s \log s)$.

Next, we give in Table I the values of c_s for Halton, Faure, Niederreiter-Xing and Halton-Atanassov, for values of s ranging between 10 and 50. The values t_s for the

Table I. Values of the constants in the discrepancy bounds for each sequence and different dimensions s

s	Halton ($e^{\lambda s}$)	Faure ($e^{\rho s}$)	Nied-Xing ($e^{\nu s}$)	Halton-Atan ($e^{\mu s}$)
10	2.18e-01	2.19e-01	4.46e-03	2.82e+00
20	2.85e-02	1.72e-02	1.75e-08	1.44e-07
30	3.53e-03	3.27e-05	1.40e-16	1.91e-17
40	5.78e-04	1.20e-07	1.07e-24	6.17e-29
50	9.50e-05	1.19e-09	7.56e-35	1.34e-41

Niederreiter-Xing sequences were obtained using the database `MinT` [Schürer and Schmid 2006]. That is, for each s we searched for the pair (b, t) that was yielding the smallest value for c_s in (12). These results show that for smaller values of s, the Niederreiter-Xing sequences have the smallest constant, but around $s = 30$, the Halton-Atanassov sequences start to have a better constant, hence suggesting an advantage in favor of the Halton-Atanassov sequences inside the "$-s \log s$" order. Recall however that these are only part of upper bounds on the discrepancy. Obviously, the sequences might be ranked differently based on the actual discrepancy (if we were able to compute it). Also, such comparisons do not take into account the $O((\log N)^{s-1})$ term in the bound (6), which for reasonable values of N might completely dominate the first term in that bound. Moreover, in the first component of (6) there is the prominent term $(\log N)^s$ which becomes very large as s and N increase, and for such large values the ranking of the constants c_s may become irrelevant. Finally, this classification might change with new possible improvements on the discrepancy bounds of some sequences, especially the Sobol' ones, because although they appear to be the "worst" according to the above ranking, they have been found by several authors to perform very well in practice.

2.4 Linear Scramblings

We end this section by a discussion of a class of permutations that is often used to define generalized Halton sequences. In particular, our new construction described in Section 4 makes use of this type of permutations called *linear scramblings*.

These scramblings are a special class of permutations introduced in Matoušek [1998] in his attempt to classify more general scramblings of (t, s)-sequences proposed by Owen [1995]. A special sub-class was first considered by Tezuka [1995] with very successful implementations in finance (see also the *GFaure sequences* in the software Finder discussed in Paskov and Traub [1995]). Of course such scramblings can also apply to Halton sequences. That is, they can be used to define the permutations $\sigma_{j,r}$ used in (4) to alter the digits $a_0 = a_0(n), a_1 = a_1(n), \ldots$ defining the coordinates of the sequence.

A *linear digit scrambling* corresponds to using permutations of the form

$$\sigma_{j,r}(a_r) = f_{j,r} a_r + g_{j,r} \pmod{b_j} \quad \text{for } 0 \le a_r \le b_j - 1, \tag{18}$$

in (4), for $j = 1, \ldots, s$, $r \ge 0$, and where $f_{j,r} \neq 0$ and $g_{j,r}$ are in Z_{b_j}, with $f_{j,r}$ and b_j coprime.

If $g_{j,r} = 0$, we obtain the so-called *multipliers* f studied in Atanassov [2004],Chi et al. [2005] and Faure [2006]. The additive factor $g_{j,r}$ is a translation also called *digital shift*. Also, for a fixed j, the multipliers $f_{j,r}$ are often chosen to be the same for all $r \ge 0$. That is, we have $f_{j,r} = f_j$ for all $r \ge 0$.

More generally, a *linear scrambling* is a permutation of the digits a_r in the set Z_b depending also on the preceding digits a_c for $0 \le c \le r-1$. It is of the form

$$\sigma_{j;a_0,\ldots,a_{r-1}}(a_r) = \sum_{c=0}^{r} f_{j,r,c} a_c + g_{j,r} \pmod{b_j}, \tag{19}$$

for $j = 1, \ldots, s$, $r \ge 0$, and where $f_{j,r,c}, g_{j,r} \in Z_{b_j}$, with $f_{j,r,r} \neq 0$ and $f_{j,r,c}$ and b_j coprime. The transformations proposed in Tezuka [1995] correspond to the case $g_{j,r} = 0$. Because of the dependence on the previous digits, this type of permutation is more general than the one used in our definition of generalized Halton sequence.

For our new construction, we use deterministic linear digit scramblings that are homogeneous (i.e., $g_{j,r} = 0$), and with $f_{j,r} = f_j$ for $r \ge 0$. We also use random digital shifts for the purpose of variance estimation, as explained in Section 5. This amounts to setting $f_{j,r} = 1$ and taking random independent $g_{j,r}$'s in (18).

3. OVERVIEW OF PROPOSED GENERALIZED HALTON SEQUENCES

The idea of using permutations—or scramblings—to improve the quality of Halton sequences goes back to Faure [1978], and Braaten and Weller [1979]. Since then, several other authors have proposed different ways of choosing these permutations [Atanassov 2004; Atanassov and Durchova 2003; Chi et al. 2005; Faure 1992; 2006; Kocis and Whiten 1997; Mascagni and Chi 2004; Tuffin 1998; Warnock 1995; 2002]. Among those proposals, we have selected four to be tested in our numerical experiments, which are the most recent ones. In addition, we also consider the "randomized Halton sequence" proposed by Wang and Hickernell [2000]. It is not a generalized Halton sequence based on deterministic permutations, but it improves the Halton sequence via a certain type of randomization, and can therefore fit in another category with the randomization proposed in Morokoff and Caflisch [1994]. Hence, with our new scrambling and the original Halton sequence—labeled **H** in what follows—we have tested seven sequences. We now describe the four chosen generalized Halton sequences, and provide the key ideas and a detailed analysis for each. Following that, we briefly discuss the randomized Halton sequence of Wang and Hickernell [2000], labeled **WH** in what follows. Note that although the permutations found in Tuffin [1998] for $s \le 16$ appear to do well on the problems tested in that paper, they are only given up to $s \le 16$ (in his Ph.D. thesis), and were found by a random search in a space whose size grows exponentially fast with s. Getting good permutations in this way would require too much computation time for the dimensions considered here, which is why we have not selected this approach.

(1) **KW [Kocis and Whiten 1997]:** The first few papers describing scrambling schemes for Halton sequences were based on the idea of using one permutation σ_j for each base b_j ($1 \le j \le s$), and then apply it to each digit in order to break the correlations observed on many two-dimensional projections of the sequence. In these early papers, no restriction was imposed on the choice of permutations. The choices presented in Braaten and Weller [1979] and Tuffin [1998] are based on computer searches that can be rather intensive for a large dimension s. As an alternative to computer searches, Kocis and Whiten [1997] suggest a choice of permutations that can be easily determined for any choice of s. Note that

Faure [1992] also proposes permutations that can be determined easily for any base. Preliminary experiments suggested that these permutations were not as good for high-dimensional problems as those from Kocis and Whiten [1997], which is why they were not considered further.

The following presentation differs from the one given in Kocis and Whiten [1997], who simply use small examples to describe their idea. Instead, we use a general description based on the van der Corput sequence, which goes as follows. Let $n_s = \lceil \log b_s / \log 2 \rceil$. Define the permutation π of $Z_{2^{n_s}}$ by $\pi(k) = 2^{n_s} S_2(k+1)$ for $0 \le k < 2^{n_s}$, where $S_2(k+1)$ is the $(k+1)$th term of the van der Corput sequence in base 2. Then, the s permutations $\sigma_1, \ldots, \sigma_s$ of the Kocis–Whiten's scrambling are derived from π through the formula

$$\sigma_j(h) = \pi(k_h) \text{ for } 0 \le h \le b_j - 1, \qquad j = 1, \ldots, s,$$

in which $k_0 < k_1 < \ldots < k_{b_j - 1}$ is the set $\{k;\ 0 \le \sigma(k) \le b_j - 1\}$ numbered from 0 to $b_j - 1$. In other words, the permutation σ_j is obtained by skipping the values greater than or equal to b_j in π. For instance, take $s = 3$. Then $n_3 = \lceil \log 5 / \log 2 \rceil = 3$ and so σ corresponds to the permutation $[0, 4, 2, 6, 1, 3, 5, 7]$. Hence σ_1 corresponds to $[0, 1]$, σ_2 to $[0, 2, 1]$ and σ_3 to $[0, 4, 2, 1, 3]$.

There is no theoretical motivation for this choice: the authors speak of "a question of art", arguing that "there is no recipe which leads to optimal permutations of coefficients". Our description based on the van der Corput sequence suggests that one could as well choose another sequence S_b^τ to define π (e.g., S_3^I, which has a smaller discrepancy than S_2^I). We refer to this sequence as **KW** in our experiments.

In the same paper, another method is introduced to break the correlations, and consists in leaps of length equal to a prime number L, different of the p_j's, for each of the one-dimensional sequences $\{S_{p_j}(n), n \ge 1\}$, $j = 1, \ldots, s$. These leaped Halton sequences often give better numerical results than the permuted ones, but the leaps they use are tailor-made for 5 of the 9 test-functions of the paper. More generally, such leaps can be applied to any scrambling and to any sequence.

Finally, Kocis and Whiten also perform computations with Sobol' and Faure leaped sequences, with improvements on the original ones. Their conclusion is that detailed investigations should be completed for these sequences to assess the method (see Kocis and Whiten [1997, Section 3.2]).

(2) **AD [Atanassov and Durchova 2003; Atanassov 2004]:** The next sequence is the Halton-Atanassov sequence, which improves the behavior of c_s in the discrepancy bound of Halton sequences, as discussed in Section 2. Its generation and testing can be found in Atanassov and Durchova [2003], which is why we label this sequence as **AD** in what follows. Actually, the result of Atanassov has been known for some years before being published, therefore the apparent shift between the two papers. The proofs in Atanassov [2004] involve a deep knowledge of Halton sequences and cannot be even surveyed here.

So for this sequence, the permutations $\sigma_{j,r}$ in (4) make use of the "admissible integers" m_j introduced in Atanassov [2004]. More precisely, they are linear digit scramblings of the form

$$\sigma_{jr}(a_r) = m_j^r \, a_r \pmod{p_j}, \qquad r \ge 0 \tag{20}$$

for the rth digit in dimension $j = 1, \ldots, s$. The integers $m_1, \ldots, m_s$ are said to be *admissible integers* for $p_1, \ldots, p_s$ if (i) for each $j = 1, \ldots, s$, p_j does not divide m_j; (ii) for any set of integers $d_1, \ldots, d_s$ such that p_j does not divide d_j, there exist integers $\alpha_1, \ldots, \alpha_s$ satisfying the congruences

$$m_j^{\alpha_j} \prod_{1 \le l \le s, l \ne j} p_l^{\alpha_l} \equiv d_j \pmod{p_j}, \qquad j = 1, \ldots, s. \tag{21}$$

For instance, consider the case where $s = 3$. Then $p_1 = 2, p_2 = 3, p_3 = 5$, and $\{m_1 = 1, m_2 = 2, m_3 = 3\}$ are admissible integers, since (i) m_j is not a multiple of p_j; (ii) the congruence (21) is solved with $\alpha_2 = \alpha_3 = 1$ for any d_1 not a multiple of $p_1 = 2$ (i.e., for any odd d_1); it is solved with $\alpha_1 = 0, \alpha_2 = \alpha_3 = 1$ for $d_2 \equiv 1 \bmod 3$ and with $\alpha_1 = \alpha_2 = \alpha_3 = 1$ for $d_2 \equiv 2 \bmod 3$; solutions for the four cases $d_3 \equiv i \bmod 5$, $i = 1, \ldots, 4$ can also be easily found.

The admissible integers used in the experiments of Section 6 are given in the file `haltondat.h` available at `parallel.bas.bg/~emanouil/sequences.html`. Clearly, other admissible integers could be chosen. For example, Vandewoestyne and Cools [2006] list their own choice for s up to 30.

In Atanassov and Durchova [2003], the authors claim that if the power r in (20) is increased to $r + 1$ and an additive (random) digital shift is added in this type of permutation, then one can carry through the proof for the improved behavior of the constant c_s given in Atanassov [2004] for a deterministic generalized Halton sequence based on (20). As they point out, using powers $r + 1$ should be intuitively better than using r, the latter implying that the first digit of each coordinate in this type of generalized Halton sequence is the same as that in the original Halton sequence. The online appendix shows a few numerical results suggesting that the obtained sequence—using powers r—gives results that can hardly be distinguished from the original Halton sequence. This has also been observed in Vandewoestyne and Cools [2006], who established comparisons with the Halton-Atanassov method, but using the powers r instead of $r + 1$. In our experiments, we have used the powers $r + 1$ both for the deterministic version of the AD sequence, and the one randomized with an additive digital shift.

(3) **CMW [Chi et al. 2005]:** More recently, new linear digit scrambling schemes (without digital shifts and with the same multiplier $f_{j,r} = f_j$ for each digit in dimension j) have been proposed to improve Halton sequences, first by Mascagni and Chi [2004], Chi et al. [2005] and then Vandewoestyne and Cools [2006] (construction (4), to be discussed next). In Mascagni and Chi [2004], an interesting analysis of the correlations that occur in the original Halton sequence is done. The multiplier f_j applied to each digit in dimension j (for $1 \le j \le s$) is then obtained by means of a criterion which involves the serial test for two-dimensional sequences produced by *Linear Congruential Pseudorandom Number Generator* (LCPRN). In fact, the authors use an upper bound given in Niederreiter [1978, p. 1025]) for the two-dimensional extreme discrepancy D—not L_2, as they claim —without mentioning what is the sequence involved in the formula, and refer to Niederreiter [1978] without any more precision. In Chi et al. [2005], this criterion is recalled and blended with an added feature derived from earlier work in Warnock [1995; 2002] involving

the Weyl sequences, which are another type of quasi-random sequences that make use of irrationals $\sqrt{p_j}$. Such a mixture obtained from two sequences—LCPRN and Weyl—having no connection with van der Corput–Halton sequences nevertheless gives good results in our tests until dimension 50 (the multipliers are not available beyond that). The corresponding sequence is denoted **CMW** in what follows.

(4) **VC [Vandewoestyne and Cools 2006]:** Vandewoestyne and Cools [2006] give a good survey of generalized Halton sequences, including Chi et al. [2005], and then define their own scrambling which, in fact, reduces to the linear digit scrambling (18) with $f_{j,r} = f_j = b_j - 1$ for all $r \geq 0$. Their tests based on computations of T^* and T show a clear advantage for this choice of multipliers, but only for a small number of points (1 to 10^3 or 10^4). However, computations of D and T done in Faure [2006] for one-dimensional van der Corput sequences scrambled with $\sigma(k) = fk$, $1 \leq f \leq b-1$, show that using $f = b-1$ or $f = 1$ give the same result. From this point of view, the one-dimensional projections of this construction are not better than for the Halton sequence.

As we can see in Figure 2, this construction fails to break the correlations present in the original Halton sequence. Our tests also confirm that such a scrambling is close to the identity permutation. Maybe the only difference is that the first point after the origin is close to $(1, \ldots 1)$ (see Vandewoestyne and Cools [2006], end of Section 4.8). In that context, we must recall the very pertinent observation of Matoušek [1998] on T^*: "nearly the lowest possible L_2-discrepancy is attained by a pathological point set if the number of points is not large enough in terms of the dimension, for instance a dimension around 30 and a number of points less than 10^4". This sequence is labeled **VC** in our experiments.

Finally, and as mentioned at the beginning of this section, we also tested the **WH** sequence proposed in Wang and Hickernell [2000] in our experiments, which makes use of ideas discussed in Pagès [1992] and Struckmeier [1995] to define what is called by Struckmeier a "generalized Halton sequence". As explained in Wang and Hickernell [2000], it can be described as follows: the first point X_1 of the sequence is randomly and uniformly generated in $[0,1)^s$, and written as

$$X_{1,j} = \sum_{r=0}^{k_j} x_{1,j,r} b_j^{-r-1}.$$

Then define the jth coordinate of the nth term of this sequence as

$$X_{n,j} = S_{b_j}(n + m_0^{(j)}), \qquad \text{where} \qquad m_0^{(j)} = \sum_{r=0}^{k_j} x_{1,j,r} b_j^r,$$

for $j = 1, \ldots, s$ and $n \geq 2$. In other words, the obtained sequence is made up of van der Corput sequences that each start at a different index, determined by X_1. It is not hard to see that this is in fact very similar to performing a random digital shift—described in detail in Section 5—on the original Halton sequence, with slight differences only arising from the fact that in WH, additions are not done modulo b_j. That is, if, say $b_j = 23$, $n = 7$ and $X_{1,j} = 16/23 + 4/23^2 + 17/23^3 + 12/23^4$, then in WH, we would have $X_{7,j} = 5/23^2 + 17/23^3 + 12/23^4$, while if we use X_1 as a

random digital shift for the Halton sequence, then $X_{7,j} = 4/23^2 + 17/23^3 + 12/23^4$.

As it turns out, the WH sequence performs very similarly to the randomly digitally shifted Halton sequence in our experiments. This is also consistent with remarks in Chi et al. [2005] and Vandewoestyne and Cools [2006] to the effect that this method does not break the two-dimensional correlations that are known to cause problems to the original Halton sequence in large dimensions.

4. A NEW CONSTRUCTION

While reviewing constructions that have been proposed to improve Halton sequences, it appeared to us that it would be useful for practitioners to have access to a generalized Halton sequence satisfying the following requirements. First, we think it is preferable to try to achieve an improvement in a deterministic way, rather than relying on random choices, as in the WH sequence. Second, we believe it is important to come up with a systematic way of choosing the parameters required to define such sequences. Note that this is not the case for Halton-Atanassov sequences, where one must still choose admissible integers among the (possibly large) set of possible choices. Third, it seems imperative that the criterion used to drive this systematic search should have good theoretical foundations. The AD sequence clearly satisfies this requirement, at least asymptotically, but the same cannot be said of the KW and CMW sequences. Finally, it is obvious that the newly proposed sequence should be competitive with others that have been proposed. Our sequence has been designed to fulfill these requirements, and is described next.

The theoretical foundation of our sequence lies in the study described in Faure [2005] and its application to QMC methods in Faure [2006]. In this work, the selection of multipliers $f_{j,r}$ for linear digit scramblings of the form (18) was made for $(0, s)$-sequences generated by nonsingular upper triangular matrices (see Faure [2006, corollary 2, p.118]), using the extreme discrepancy D or the diaphony $F := 2\pi T$ in one dimension. But since these multipliers give very good discrepancy or diaphony for the resulting scrambled van der Corput sequences (in one dimension), it is natural to use them also for Halton sequences, whose one-dimensional projections are precisely van der Corput sequences.

Also, these "good" multipliers are about the same whether we use the criterion based on D or T. In what follows, we chose to work with T, because in dimension $s > 1$, the extreme discrepancy D is difficult to compute, while T can be computed in $O(N^2 s)$, as discussed in Section 2. Since our search method attempts to not only look at one-dimensional projections but also two-dimensional ones, working with T instead of D is a reasonable choice.

We now describe how we chose the multipliers $f_1, \ldots, f_s$ used to define our new generalized Halton sequence, based on linear digit scramblings of the form (18) with $f_{j,r} = f_j$ for $r \geq 0$. In what follows, we work with prime bases, and let S_p^f be the generalized van der Corput sequence in base p in which the permutations σ_r are given by $\sigma_r(k) = fk \pmod p$. First, for each prime base p among the first s prime numbers, we sort the set $\{1, \ldots, p-1\}$ of potential multipliers f according to the criterion $\theta_p^f(1)/\log p$ described in Faure [2006], which is related to a bound on the

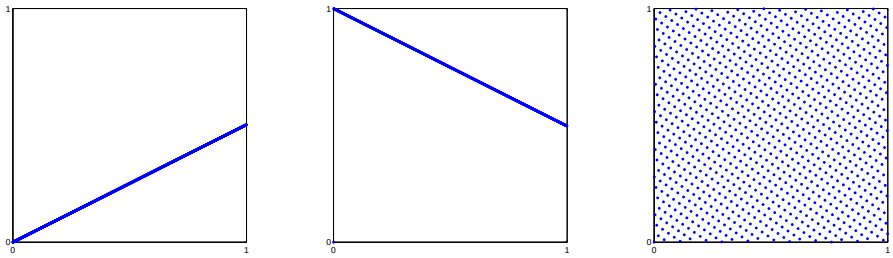

Fig. 1. Hammersley point sets with $N = 1000$ based on generalized van der Corput sequence in base 1987, using multiplier $f = 1$ (left), corresponding to H; $f = 1986$ (middle), corresponding to VC; and $f = 555$ (right), corresponding to FL with a single element in the short list

L_2-discrepancy of S_p^f. More precisely, $\theta_p^f(1)$ is defined as [Faure 2006, Prop. 2]

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

and is such that if $\Sigma = (\sigma_r)_{r \ge 0}$ satisfies $\sigma_r(k) = f_r k \bmod p$ with f_r such that $\theta_p^{f_r}(1) \le \theta_p^f(1)$, then for all $N \ge 1$ we have

$$T^2(N, S_p^\Sigma) \le \frac{\theta_p^f(1)}{\log p} \log N + \theta_p^f(1) + \frac{1}{12}.$$

Hence the multiplier f that minimizes $\theta_p^f(1)$ is such that the corresponding sequence S_p^f has the smallest bound on its L_2-discrepancy, among all van der Corput sequences in base p permuted by an homogeneous linear digit scrambling. The second multiplier f on the list is such that S_p^f has the second smallest bound, and so on. Our short list thus retains the multipliers with the smallest upper bound on the L_2-discrepancy of the corresponding sequence S_p^f.

The motivation for this criterion is as follows. First, $\theta_p^f(1)$ can be computed relatively fast. For instance, for the 300th prime $p_{300} = 1987$, the list of all 1986 multipliers between 1 and 1986 can be sorted according to $\theta_{1987}^f(1)/\log 1987$ in just a few seconds. Second, this bound on the discrepancy can be thought of as a way of measuring how good is the multiplier f at improving the "space-filling" property of the van der Corput sequence in base p, which is quite bad for large bases p. Figure 1 shows the first 1000 points of the two-dimensional Hammersley point set based on S_p^1, S_p^{p-1} and $S_p^{f^*}$, where $f^* = 555$ is the factor with the smallest value of $\theta_{1987}^f(1)$ for the base $p = 1987$. That is, we plot $\{((n-1)/1000, x_n), n = 1, \ldots, 1000\}$, with x_n the nth point of the sequence under study. These plots offer a two-dimensional depiction of the space-filling properties of the sequence $\{x_n\}_{n \ge 1}$. For instance, if $x_1 < \ldots < x_{1000}$—which is not a good way of filling the space—then the corresponding two-dimensional Hammersley point set lies on a straight line, as we can see for S_p^1 on Figure 1. Note that S_p^1 and S_p^{p-1} correspond to the 300th coordinate of the Halton and VC sequence, respectively. It is clear from these plots that the factor f^* chosen according to $\theta_p^f(1)/\log p$ manages to greatly improve the space-filling property of the van der Corput sequence.

At this point, one possible approach would be to simply choose, for each dimension and corresponding prime p_j, the best factor f according to $\theta^f_{p_j}(1)$. While this would surely provide an s-dimensional generalized Halton sequence with good one-dimensional projections, it is not clear that the higher-dimensional projections would also be good. For instance, if we are unlucky, we could end up choosing for two successive dimensions, j and $j+1$, factors f_j and f_{j+1} that give rise to strong correlations between the one-dimensional sequence $S^{f_j}_{p_j}$ and $S^{f_{j+1}}_{p_{j+1}}$, which is one of the main problems that the original Halton sequence has. (This behavior can also be observed on the plot of the VC sequence in Figure 2 (top middle).) Our approach to avoid this is to also make sure that two-dimensional projections for coordinates j and k that are nearby are also of a good quality. Search criteria that consider projections of nearby indices are also discussed in Cheng and Druzdzel [2000] and L'Ecuyer and Lemieux [2002], for other types of quasi-random sequences.

More precisely, once the multipliers are sorted according to $\theta^f_p(1)/\log p$ for a given base p, we keep a "short list" of at most 32 potential candidates. Also, we only keep the multipliers for which $\theta^f_p(1)/\log p \leq 0.1$, and for $p > 3$ we never keep 1 or $p-1$. The purpose of the short list is to make sure we have good one-dimensional projections. Once we have a short list for each prime base, we select multipliers using a step-by-step approach based on the criterion

$$\tau^{W,M}_j(f_1,\ldots,f_{j-1},f) = \max_{1\leq l\leq W} T(M,(S^{f_{j-l}}_{p_{j-l}},S^{f}_{p_j})), \tag{22}$$

where $T(M,X)$ represents the L_2-discrepancy of the first M points of a sequence X and W is a positive integer to be chosen, which determines how many bidimensional projections will be assessed. More precisely, to select a multiplier f_j for the jth coordinate, for each candidate f in the short list for p_j, we compute the value T of the L_2-discrepancy for the M first 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 $\tau^{W,M}_j(f_1,\ldots,f_{j-1},f)$ among all candidates in the short list.

Summing up, our construction combines an approach for building the short list that rests on a solid theoretical foundation with a more pragmatic method to select a multiplier from this list, using the criterion (22). This ensures that special attention is paid to bidimensional projections, hence avoiding the most well-known defect of the original Halton sequences.

Table II lists the multipliers used to define our new generalized Halton sequence—labeled **FL** from now on—up to dimension 50. The complete list up to dimension 360 is available in the online appendix. We used $M = 2500$ and a window size given by $W = \min(7, j-1)$ to obtain these parameters. Before we explain this choice of $M = 2500$, first note that the construction of the short list of multipliers is done independently of the value of M, because it is based on the criterion $\theta^f_p(1)$, which provides a bound on the discrepancy *for all* N. To assess the two-dimensional projections, we chose $M = 2500$ so that M was at least as large as the largest base $p_{360} = 2423$ considered. In our experience, a relatively small initial sample often gives a good idea of the behavior of the sequence. This is in line with a

Table II. List of the first 50 multipliers for the FL sequence

j	f_j	j	f_j	j	f_j	j	f_j	j	f_j
1	1	11	18	21	43	31	92	41	104
2	1	12	8	22	61	32	47	42	126
3	3	13	13	23	60	33	29	43	50
4	3	14	31	24	56	34	61	44	80
5	4	15	9	25	26	35	57	45	55
6	9	16	19	26	71	36	69	46	152
7	7	17	36	27	32	37	115	47	114
8	5	18	33	28	77	38	63	48	80
9	9	19	21	29	26	39	92	49	83
10	18	20	44	30	95	40	31	50	97

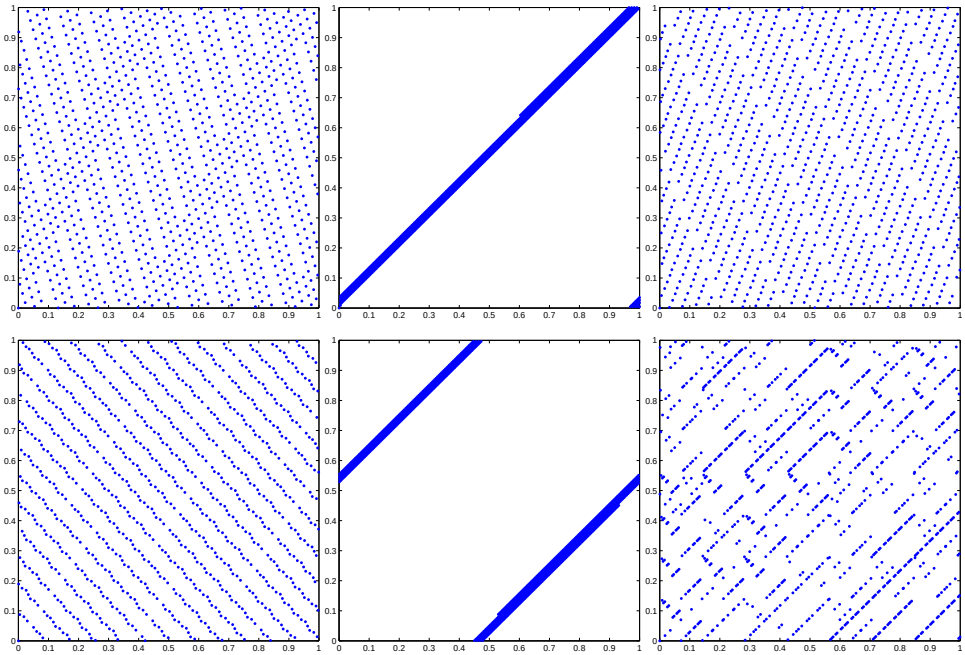

Fig. 2. Projection of the first $N = 1000$ points of different sequences over the 49th and 50th coordinates: CMW (top left), VC (top middle), AD (top right) and FL (bottom left), WH (bottom middle) and KW (bottom right)

common practice used by several authors, which is to use two-dimensional figures showing the first few thousand points of a sequence to illustrate their good or bad quality (see, e.g., Morokoff and Caflisch [1994]). On the other hand, theoretically speaking, we should use $M = p_i p_j$ points when measuring the discrepancy of the projection over the ith and jth coordinates. But this would require $M \approx 6 \times 10^6$ when i and j are close to 360. Instead, taking $M = 2500$ appears to be a good compromise that allows the search to be performed in a reasonable time. As for W, we experimented with a few other window sizes without noticing significant changes in the performance of the sequence.

5. PERFORMANCE MEASURES

To compare the different generalized Halton sequences studied in this paper, we perform empirical comparisons based on different types of integrands and assessment methods. By doing so, we wish to recognize the two main approaches that can be found in the literature for assessing the performance of low-discrepancy sequences. First, one can use test-functions g where the integral

$$I(g) = \int_{[0,1)^s} g(x)dx$$

can be computed exactly, and therefore (deterministic) low-discrepancy sequences $X_1, X_2, \ldots$ can be compared by looking at the behavior of the absolute error $E_N = |I(g) - Q_N|$ where $Q_N = \sum_{i=1}^{N} g(X_i)/N$ (the relative error $E_N/I(g)$ can also be used) as the number of function evaluations N increases. We can then adjust different features of the integrand g (e.g., related to the concept of *effective dimension*, to be defined in Section 6.1) by choosing appropriate parameters. Examples will be given in Section 6.

Test-functions may be considered somewhat artificial, and for this reason, integrands arising from more practical problems—such as the evaluation of financial products of various types—are often used to assess the performance of low-discrepancy sequences. For such problems, the exact value of $I(g)$ is typically unknown. Hence in this case, if one still wants to use the absolute error E_N to assess the accuracy of Q_N, first an approximation for $I(g)$ based on a very large value of N (e.g., in the order of 2^{19}, as in Caflisch et al. [1997], or one million in Ninomiya and Tezuka [1996]) must be performed. Although this approach is acceptable from an academic point of view, practitioners faced with the problem of estimating $I(g)$ cannot realistically be expected to perform such a large *a priori* estimation for $I(g)$ in order to obtain some kind of accuracy measure for the approximation Q_N. A more pragmatic approach is then to *randomize* the low-discrepancy sequence so that a variance estimator and/or a confidence interval can be computed.

Several randomization techniques are available for generalized Halton sequences. Most of them amount to apply randomly chosen permutations to the digits defining the coordinates of the underlying deterministic sequence. Restrictions on the kind of permutations used can be imposed. For instance, a *random linear scrambling* means randomly chosen permutations of the form (19) are applied, independently for each digit $r \geq 0$ and dimension $j = 1, \ldots, s$. A *random digital shift* is obtained by setting $f_{j,r} = 1$ and randomly choosing $g_{j,r} \in Z_{b_j}$ in (18), again for each digit $r \geq 0$ and dimension $j = 1, \ldots, s$. Compared to more general randomizations, the random digital shift has the advantage of making a smaller perturbation to the deterministic structure of the underlying quasi-random sequence. In particular, the unanchored discrepancy D and T are invariant under a random digital shift for one-dimensional generalized van der Corput sequences [Faure 2005].

Now, recall that our goal is to compare deterministic scramblings for the Halton sequence. Hence, it makes sense for us to use the randomization that has as little randomness as possible. In this way, we can make sure that comparisons based on variance estimates will be as consistent as possible with the ones that could have been made based on the absolute error of the underlying deterministic sequences,

if $I(g)$ had been known. To summarize, the three following assessment methods will be used to compare sequences in Section 6: (i) the absolute error E_N of the deterministic sequences; (ii) the average absolute error over m randomizations of the sequences, and (iii) the estimated variance of the randomized sequences. More precisely, let

$$\hat{\mu}_{N,l} = \frac{1}{N}\sum_{i=1}^{N} g(X_i \oplus V_l), \tag{23}$$

where $\oplus$ is a component-by-component digital addition performed in base b_j for the jth component, and $V_1, \ldots, V_m$ are independent randomly chosen shifts. Then the average absolute error is

$$\frac{1}{m}\sum_{l=1}^{m} |\hat{\mu}_{N,l} - I(g)| \qquad \text{and} \qquad \hat{\sigma}_N^2 = \frac{1}{(m-1)}\sum_{l=1}^{m}(\hat{\mu}_{N,l} - \tilde{\mu}_N)^2 \tag{24}$$

is an unbiased estimator for the variance of $\hat{\mu}_{N,l}$, where $\tilde{\mu}_N = (\sum_{l=1}^{m} \hat{\mu}_{N,l})/m$.

6. NUMERICAL EXPERIMENTS

As mentioned before, two main types of integrands will be tested. First, we consider different test-functions that have been used in previous studies to assess the performance of various constructions. Second, we look at integrands arising from practical problems in financial mathematics. This section presents the integrands used within these two categories.

6.1 Test-functions

There is a large number of test-functions that have been used by different authors in comparative studies involving QMC methods. A very useful description of these different functions can be found in Owen [2003]. We have chosen three such functions, which allow us to test different aspects of the constructions under study. These three functions are:

$$g_1(x) = \prod_{j=1}^{s} \frac{|4x_j - 2| + \alpha_j}{1 + \alpha_j} \qquad g_2(x) = \prod_{j=1}^{s} 1 + c(x_j - 0.5),$$

$$g_3(x) = \alpha_s \pi^{-s/2} \cos\left(\sqrt{\frac{1}{2}\sum_{j=1}^{s}[\Phi^{-1}(x_j)]^2}\right),$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal random variable, and $x = (x_1, \ldots, x_s)$. Note that g_1 and g_2 integrate to 1. For g_3, the constant α_s is determined numerically so that g_3 also integrates to 1.

The first function was proposed in Radovic et al. [1996] and is also used, e.g., in Wang and Hickernell [2000]. Here we consider the same choices that have been studied in these two papers, that is (i) $\alpha_j = 0.01$, (ii) $\alpha_j = 1$; (iii) $\alpha_j = j$, and (iv) $\alpha_j = j^2$, for $1 \le j \le s$.

As we go from (i) to (iv), the *effective dimension (in the truncation sense)* of the function g_1 decreases. The concept of effective dimension is defined precisely

in Caflisch et al. [1997]. Without going into the details, we say that g has an effective dimension d_T in the truncation sense if it can be well approximated by a function that depends only on the first d_T variables $x_1, \ldots, x_{d_T}$. If g can be well approximated by a sum of functions depending each on no more than d_S of the variables $x_1, \ldots, x_s$, then we say g has an effective dimension of d_S in the *superposition sense*.

Functions with a small effective dimension d_T are easy to integrate as long as the point set used has good properties for its projections over the first d_T coordinates. The original Halton sequence should be quite sensitive to the effective dimension d_T, since we know its projections deteriorate quickly as the dimension increases. A well-chosen generalized Halton sequence should be less sensitive, because the permutations should attenuate the bad behavior of the van der Corput sequence in large bases. Hence this function allows us to test how sensitive the different constructions are to the effective dimension (in the truncation sense) of the integrand.

In order to better assess the overall quality of the tested sequences, we also studied the choice $\alpha_j = (s - j + 1)^2$ for $1 \leq j \leq s$, which can be seen as the same choice as (iv), but where we reverse the order of the coordinates of each point. We denote this choice as (v). This means the effective dimension d_T becomes large (probably equal to the nominal dimension s), since now the most important variables are the last ones x_s, x_{s-1}, and so on. We expect all sequences to have more difficulty with this function, especially those who are not designed to break the bad behavior of the van der Corput sequence in large bases. The idea of testing the sensitivity of a sequence in this way has been used in Ninomiya and Tezuka [1996].

In our experiments, we tested the three dimensions $s = 20, 50$ and 150 for all five cases (i) to (v). We excluded the origin when evaluating g_1 as $g_1(0, \ldots, 0) = (2.01/1.01)^s$ is very large for the case (i). Everywhere else, we include the origin.

The function g_2 has been used in Sobol' and Asotsky [2003]. The effective dimension of this function can also be adjusted, this time by choosing c appropriately. Note that g_2 is defined in a symmetric way, i.e., all variables x_j contribute to g_2 in the same way. Hence for this function, it makes more sense to use the effective dimension in the superposition sense in order to measure its difficulty. In our experiments, we selected the same combinations (c, s) as in Sobol' and Asotsky [2003]: (0.1, 120), (0.25, 96), and (1, 150). For these three choices, we computed the effective dimension in the superposition sense and found it was equal to 4, 6, and 20, respectively (we used a threshold of 0.99 in the precise definition of d_S given in Caflisch et al. [1997]). This indicates that, for instance when $(c, s) = (0.1, 120)$, a point set with good projections over four-dimensional subspaces or less should perform well to integrate this function, but the quality of the projections should not deteriorate as the bases increase. Hence we expect the original Halton sequence to have problems with g_2 for all three choices of parameters. We also expect all sequences to have problems integrating g_2 when $(c, s) = (1, 150)$, because of the large peak at $(1, \ldots, 1)$.

By constrast with g_1 and g_2, the function g_3 is not defined as a product. It is instead part of a family of isotropic integrals, and was used in Papageorgiou and Traub [1997] to demonstrate the superiority of QMC over Monte Carlo for this

type of problems. This function is known to have a small effective dimension in the superposition sense. For instance, in Owen [2003] it is shown that for $s = 25$, d_S is about 3. In our experiments, we tried $s = 20$, 50, and 120 for this function.

Summing up, in total we have tested 21 different test-functions (5 types of g_1 with three dimensions each; three types of g_2 and three dimensions for g_3).

6.2 Integrands from finance

The first problem considered here is an option pricing problem. More precisely, we wish to estimate the value at time 0 of an *Asian call option* on an underlying asset that follows a lognormal distribution. Formally, the value at time 0 of this option is given by the expectation

$$C_0 = \mathrm{E}\left[e^{-rU}\left(\frac{1}{s}\sum_{j=1}^{s} S(u_j) - K\right)^+\right] \tag{25}$$

where U is the expiration time of the contract, K is the *strike price* (also defined in the option contract), $S(u)$ is the price of the underlying asset at time u, and $0 < u_1 < \ldots < u_s = U$ are s times at which the asset price is observed in order to compute the average used in (25), and r is the risk-free rate. The notation x^+ means $\max(0, x)$. The lognormal model means we assume $S(u)$ is of the form $S(u) = S(0)e^{(\mu-\sigma^2/2)u+\sigma\sqrt{u}Z}$, where $Z \sim N(0,1)$ is a standard Gaussian random variable, and μ is the return rate on the asset. In the expectation (25) though, μ is taken to be equal to r, based on risk-neutral pricing arguments. Hence the quantity C_0 can be written as an s-dimensional integral of the form

$$C_0 = e^{-rU}\int_{[0,1)^s}\left(\frac{1}{s}\sum_{j=1}^{s} S(0)e^{(r-\sigma^2/2)u_j+\sigma\sum_{l=1}^{j}\sqrt{\Delta_l}\Phi^{-1}(x_l)} - K\right)^+ dx_1\ldots dx_s,$$

where Φ^{-1} denote the inverse of the CDF of a $N(0,1)$, and $\Delta_l = u_l - u_{l-1}$. In our experiments, we used $S(0) = 50$, $U = 1$ year, $r = 0.05$, $\sigma = 0.3$, $u_j = j/s$, and two choices of dimension: $s = 40$ and $s = 75$. We consider three choices for the strike price: $K \in \{45, 50, 55\}$.

Several authors have studied the effective dimension for this type of problem [Wang and Sloan 2005; Lemieux and Owen 2001]. A common conclusion is that the effective dimension in the superposition sense is usually much smaller than s, especially when K is small.

The second problem has been widely used in the quasi-Monte Carlo literature [Paskov and Traub 1995; Ninomiya and Tezuka 1996; Caflisch et al. 1997]. The following description follows Caflisch et al. [1997]. The goal here is to estimate the value at time 0 of a financial instrument called a *mortgage-backed security* (MBS), which is given by $M_0 = \mathrm{E}\left(\sum_{j=1}^{s} v_j c_j\right)$, where v_j is the discount rate for month j, and c_j is the cash-flow for month j. Both v_j and c_j are random quantities that can be written as functions of the stochastic interest rate process $\{i_0, i_1, \ldots, i_s\}$, and that are determined by five parameters denoted K_1, K_2, K_3, K_4 and σ, in addition to the initial rate i_0. (Details are given in the online appendix.) The value for s is chosen to be 360, since the contracts under study are 30-year long, and the

Table III. Labels used for the tested sequences

MC	**H**	**KW**	**WH**
Monte Carlo	Halton	Koc.-Whit.	Wang-Hick.
1949	1960	1997	2000
		det. perm.	randomization
AD	**CMW**	**VC**	**FL**
Atan.-Durch.	Chi-Masc.-Warn.	Vand.-Cools	Fau.-Lem.
2003	2005	2006	2008
admissible int.	select "good" f_j	$f_j = p_j - 1$	select "good" f_j

quantities mentioned above are monitored monthly. In Caflisch et al. [1997], two choices of parameters are used. The first one—which we denote *linear* below—yields a 360-dimensional integrand that is nearly linear. By contrast, the second one yields a function referred to as *non-linear* by Caflisch et al., although, as they explain, it still has a large linear component, but not as extreme as in the "nearly linear" case. In our experiments, we also report results based on the parameters used in Ninomiya and Tezuka [1996], and refer to those as "Nin-Tez".

These two problems have no analytical solutions. Hence, when we look at the deterministic error, we use "benchmark prices" as given in Caflisch et al. [1997] and Ninomiya and Tezuka [1996] for the MBS problem. For the Asian option problem, we used a Sobol' sequence with 2.5×10^6 points to determine those prices.

Summing up, in total we have tested nine different finance problems. Before presenting the results, we recall in Table III the labels used for each construction.

6.3 Results

We have produced three figures for each of the 30 functions described in the previous sub-sections: one showing the deterministic error E_N for each construction, for values of N going from 2000 to 100,000, increasing by 2000 each time; the second shows the average error over 25 independent copies of the randomized estimators (all using a digital shift) at the same values of N as for the deterministic error; the third one shows the estimated variance $\hat{\sigma}_N^2$ as given in (24), again for values of N ranging between 2000 and 100,000. For all functions, as in Sobol' and Asotsky [2003], we give as a reference on the error plots the median error for the Monte Carlo (MC) estimator based on N points, which is given by $0.6745\sigma/\sqrt{N}$, where σ^2 is the variance of the function g. For the variance plots, we show the MC variance σ^2/N. For the financial problems, we replace (the unknown) σ^2 by an estimate based on $N = 100,000$ simulations.

For problems with dimension $s \leq 50$, we do not show VC among the randomized constructions; when $s > 50$, since parameters for CMW are not available, then we show VC so that with the randomized constructions, we always compare a total of six methods in addition to MC. We chose to drop VC because it was very similar to Halton in most cases.

The 90 figures for these 30 functions are all given in the online appendix, with one page for each function. (In the online appendix, we have also added the case $K = 60$ and $s = 40$ for the Asian option, and results for N up to 10^6 for g_1, case (i) with $s = 50$, and the Asian option with $K = 55$ and $s = 75$.) Here, we summarize the results in three tables: Table IV is for g_1, Table V is for g_2 and g_3, and Table VI is for the financial problems. In these tables, the notation "A>B" means the

Table IV. Summary of the results for g_1 for (i) $\alpha_j = 0.01$; (ii) $\alpha_j = 1$; (iii) $\alpha_j = j$; (iv) $\alpha_j = j^2$; and (v) $\alpha_j = (s-j+1)^2$; the notation $A > B$ means A did worse than B

g_1	(i)	$s = 20$	$s = 50$	$s = 150$
	Det.	H, VC > MC	H, VC >> MC KW > MC	H, VC, KW > MC
	Rand.:	Best are FL, KW	All have jumps	All have jumps
	This is a difficult function, and already with $s = 50$ we observe jumps in the error due to function evaluation in the problematic region near $(1, \ldots, 1)$. H, VC and KW get huge deterministic errors with $s \geq 50$. FL behaves well.			
	(ii)	$s = 20$	$s = 50$	$s = 150$
	Det.:	All good	H, VC >> MC KW > MC	H, VC >> MC KW > MC
	Rand.:	FL best	H, WH > MC FL and CMW best	All have jumps FL, WH and AD best
	KW does not do well on this function.			
	(iii)	$s = 20$	$s = 50$	$s = 150$
	Det.:	All good	All good	VC a bit worse
	Rand.:	All good	Clear separ. H, WH > AD, FL, KW, CMW	Clear separ. H, WH, VC > AD, FL, KW
	Here we see the separation H,WH,VC vs AD,FL,CMW,KW that appears elsewhere.			
	(iv)	$s = 20$	$s = 50$	$s = 150$
	Det.:	All good	All good	All good
	Rand.:	All good	All good	All good, FL seems slightly better
	This function is easy and does not allow to distinguish the constructions.			
	(v)	$s = 20$	$s = 50$	$s = 150$
	Det.:	All good	H, VC > AD,FL,KW,CMW	H, VC > MC
	Rand.:	Clear separ. H, WH > AD, KW > FL, CMW	Clear separ. H, WH > AD, KW > FL, CMW	Clear separ. H, VC, WH > KW > AD, FL
	Inversing the order is especially bad for H, VC and WH. All constructions show cyclical behavior in error. FL always among the best (not true for AD and KW).			

sequence "A" is worse than "B".

We also give an excerpt of the most representative results obtained.

For the function g_1, in some cases we get an easy integrand for which all constructions work well: the most obvious example is the case (iv), which has a very small effective dimension in the truncation sense. The cases (i) and (ii) turn out to be quite challenging, causing H, WH and VC to sometimes be much worse than MC—especially without a randomization—as shown on Figures 3 and 4. As we see there, the randomization significantly improves H.

Looking at what happens when the order of the variables is reversed—corresponding to (v)— is quite interesting: while with the original order—corresponding to (iv)—all sequences are very good, with the order reversed, the sequences do not behave as well, especially H, WH and VC, which are worse than MC for $s = 150$. The better constructions, although they still behave better than MC, surely have less uniformity in those high dimensions than in the earlier ones, thereby their associated error for this type of function is larger than it was for the case (iv), and can exhibit a cyclical behavior, as seen on Figure 5.

Both g_2 and g_3 have an effective dimension in the superposition sense smaller than

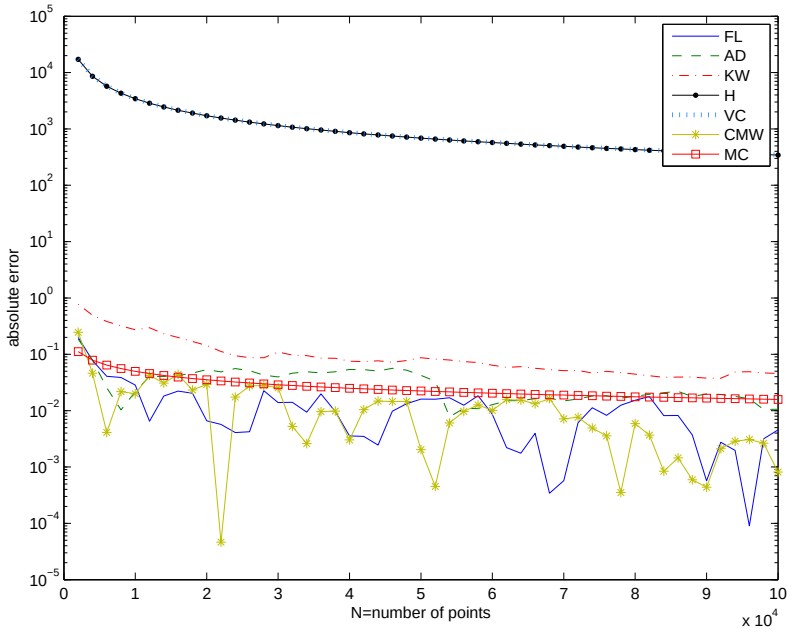

Fig. 3. Deterministic error for g_1, case (ii) with $s = 50$: H and VC are much worse than MC

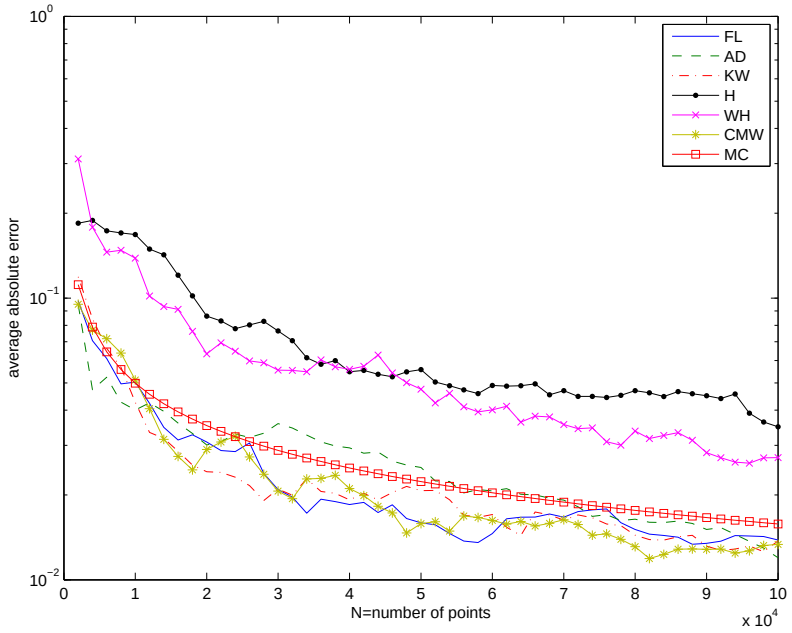

Fig. 4. Average absolute error for g_1, case (ii) with $s = 50$: although the random digital shift helps H, it is still worse than MC

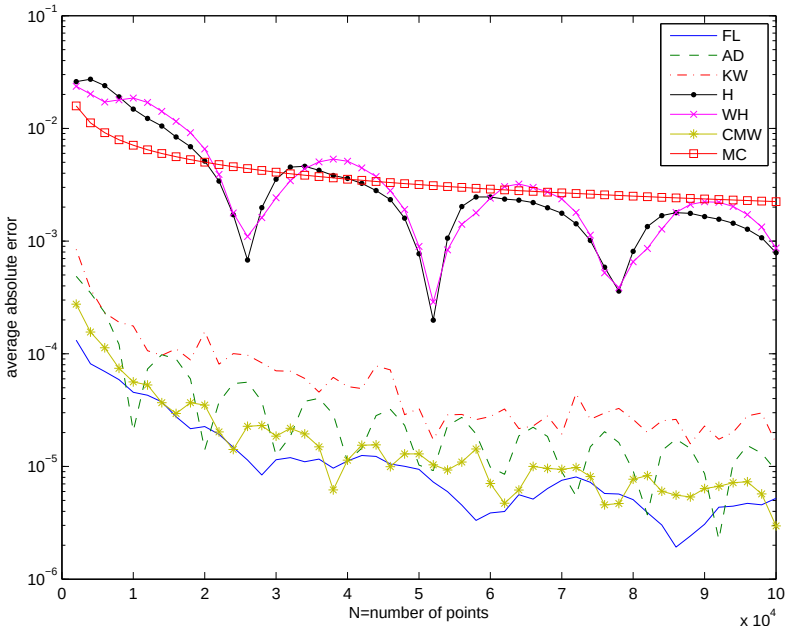

Fig. 5. Estimated variance for g_1, case (v) with $s = 50$: FL seems the best; H and WH barely improve upon MC

Table V. Summary of the results for g_2 and g_3; $A > B$ means A did worse than B

g_2	(c, s)	(0.1,120)	(0.25, 96)	(1,150)
	Det.:	H,VC > MC	H,VC > MC	H,VC,KW > MC
	Rand.:	Clear separ. H,WH,VC > AD,FL,KW	Clear separ. H,WH,VC > AD,FL,KW	Jumps due to problematic region
g_3	s	20	50	120
	Det.:	H close to MC	H, VC> MC	H, VC > MC
	Rand.:	All good	Clear separ.	Clear separ.
		H, WH slightly worse than FL, CMW, AD and KW	H,WH,VC > AD,FL,KW	H,WH,VC > AD,FL,KW

s. However, since the truncation effective dimension is not small, these functions can still be challenging for sequences whose quality deteriorate in the higher-valued indices. In Figures 6 and 7, we show the average error for g_2 with $(c, s) = (0.25, 96)$, and the estimated variance for g_3 when $s = 120$, respectively. In both cases, we see a clear separation between H, WH and VC versus AD, FL and KW. With g_2, the group H–WH–VC is worse than MC for small values of N.

For the Asian option problem, in dimension 40 there was not such a clear difference between the different constructions, but when we increased to $s = 75$, then it became clear that AD, FL, and KW were superior to H,WH and VC. As an illustration, Figure 8 shows the estimated variance when $s = 75$ and $K = 50$.

For the mortgage-backed security, the "Nin-Tez" set of parameters was such that all functions performed well. According to some preliminary tests we did, it seems like this particular choice of parameters causes the effective dimension in the

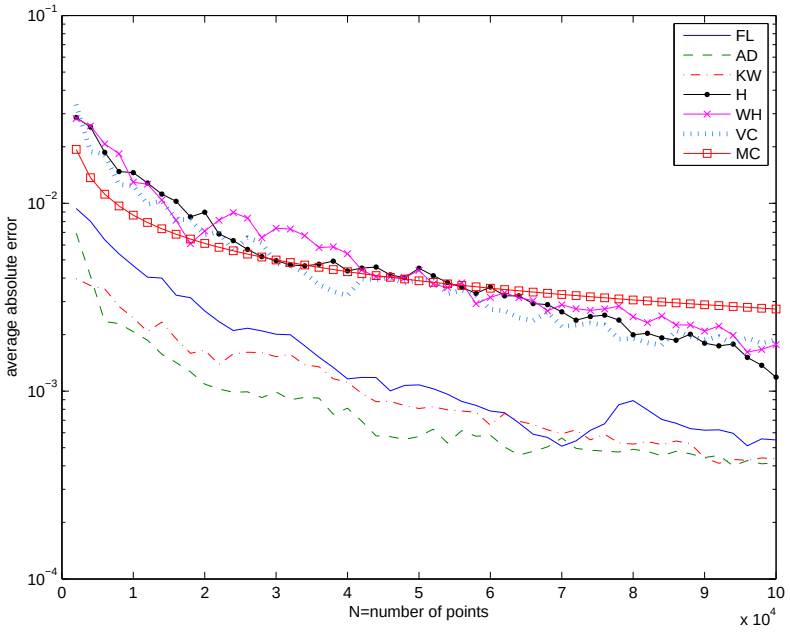

Fig. 6. Average error for g_2 with $(c, s) = (0.25,96)$: H, WH and VC are barely better than MC

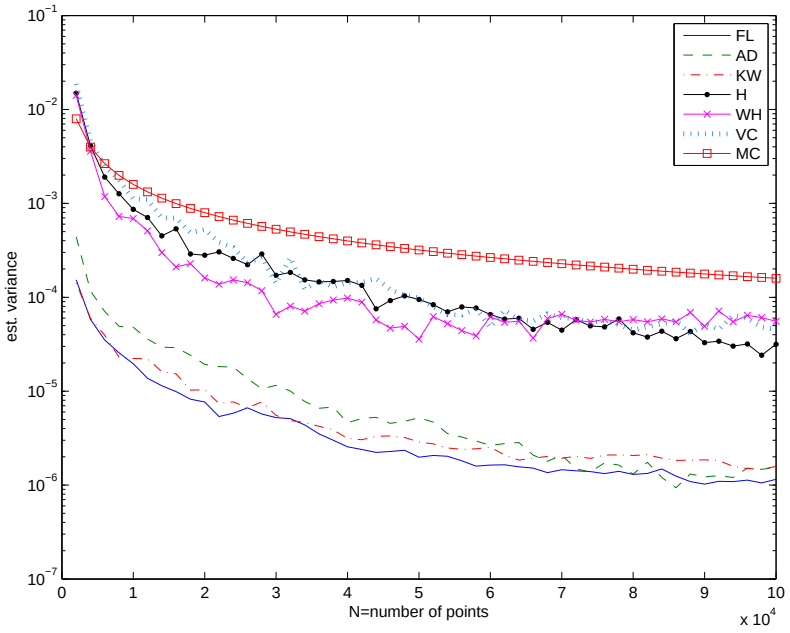

Fig. 7. Estimated variance for g_3 with $s = 120$: the group FL-AD-KW does clearly better than H-WH-VC

Table VI. Summary of the results for financial problems; $A > B$ means A did worse than B

As. opt.	$s = 40$	$K = 45$	$K = 50$	$K = 55$
	Det.:	H,VC> MC CMW > AD,FL,KW	H,VC> MC	H,VC > MC
	Rand.:	H,WH,AD > FL,KW,CMW	H worse	All good, FL slightly better
As. opt.	$s = 75$	$K = 45$	$K = 50$	$K = 55$
	Det.:	H,VC > MC	H,VC > MC	H,VC > MC
	Rand.:	Clear separ. H,WH,VC > AD,FL,KW	Clear separ. H,WH,VC > AD,FL,KW	Clear separ. H,WH,VC > AD,FL,KW
MBS	type	nearly lin.	non-lin.	Nin-Tez
	Det.:	H,VC > MC AD > FL, KW	H > MC VC erratic	H worse but better than MC
	Rand.:	Clear separ. H,WH,VC > AD > FL, KW	Clear separ. H,WH,VC > AD,FL,KW	All good

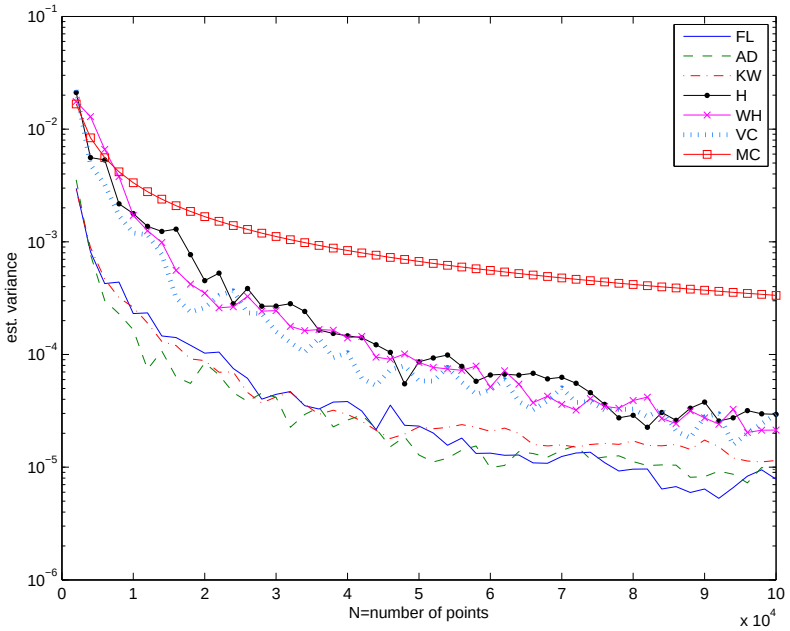

Fig. 8. Estimated variance for Asian option with $s = 75$ and $K = 50$: here also, the group FL-AD-KW is clearly better than H-WH-VC

truncation sense to be quite small (around 20 or so). We think this might be the explanation for the behavior observed. With the two other sets of parameters, there was once again a clear distinction between H, WH, VC and AD, FL, KW. However, with the "nearly linear" set of parameters, AD was not performing as well as FL and KW. Figure 9 illustrates this.

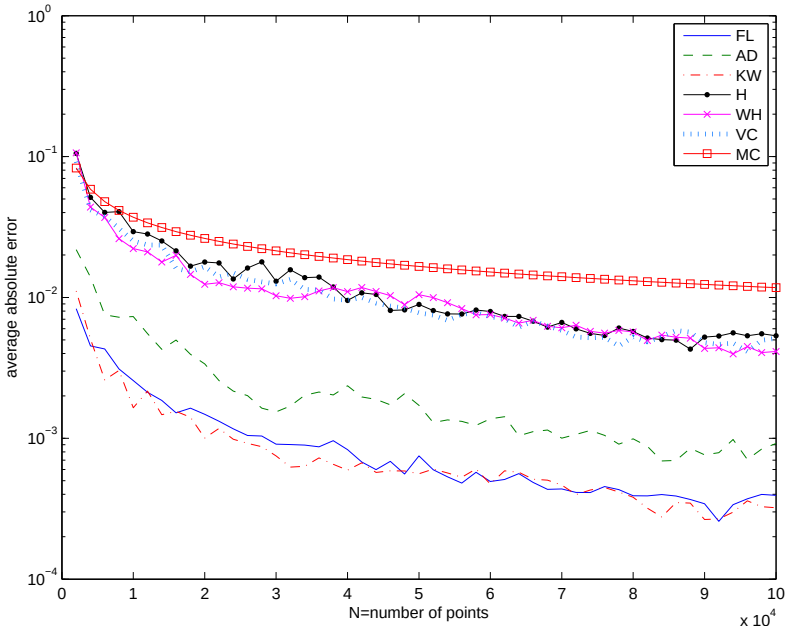

Fig. 9. Average error for mortgage-backed security, nearly linear case: FL and KW are best, followed by AD, then H-WH-VC

7. CONCLUSION

In this paper, we reviewed recent results on generalized Halton sequences, and proposed a new sequence (FL) whose parameters have been found by a computer search where the quality of one and two-dimensional projections is assessed via the L_2-discrepancy. Our numerical results indicate that deterministic generalized Halton sequences based on non-trivial permutations such as ours outperform the original Halton sequence and modifications (WH, VC) that retain some of the inherent problems of the Halton sequence, in particular with respect to bi-dimensional correlations. We think these constructions should be avoided, especially the deterministic versions (H,VC).

Among the remaining tested sequences, CMW has the disadvantage that factors only up to dimension 50 have been published. While this could presumably be extended, it does not seem obvious that the somewhat loose arguments used to find these factors would provide a good construction in higher dimensions. AD rests on a strong theoretical background, but the issue concerning the powers of the admissible integers—starting at 0 or 1—shows a gap between the underlying theoretical results of this construction and its quality in practice. Furthermore, it is quite possible that another choice of admissible integers would have given significantly different results, which would imply that one needs a complementary method for choosing "good" admissible integers. KW appears to perform relatively well, although there were some cases in our numerical experiments where it was doing worse than Monte Carlo. We thus think it cannot be safely recommended. It also requires much more memory to store the parameters and time to generate

the sequence.

Based on this, we think the FL sequence is a generalized Halton sequence with good properties that can be recommended to practitioners. Our numerical experiments indicate that it is always among the group of best constructions, and never does worse than Monte Carlo. This suggests that it can be used safely for a variety of problems. Furthermore, our approach for finding multipliers for the FL sequence could easily be adjusted to take into account specific information one might have on the functions to be integrated, by simply adjusting the search criterion (22) to another one that measures properties that are relevant for the problem at hand (for instance, higher-dimensional projections). We should also point out that our kind of search for "good" multipliers can be done relatively quickly on a computer, since we only search one integer per dimension, and do so among a short list of candidates, which themselves can be obtained very quickly with an effective theoretical support in one dimension. This is much faster than having to search among all possible permutations, in which case one has to rely on a random search in order to keep the computation time reasonably low, as done in Tuffin [1998].

Another point to discuss in light of the results presented in the previous section is the question of whether or not randomizations should be used with generalized Halton sequences. Given the fact that in no case we observed a deterioration of the quality of an approximation when performing a random digital shift—and in some cases, significant improvements were observed by using such shifts—we think it is generally recommendable to use such randomizations. In addition, as discussed in Section 5, it has the advantage of providing an error estimate for the approximation.

Finally, in the numerical experiments presented in this paper, we have not compared generalized Halton sequences with other constructions, such as those falling in the category of (t, s)-sequences. Such comparisons would clearly have to be made if one was trying to determine which construction works best for a given problem. Our goal here was instead to focus on generalized Halton sequences and study which constructions in this group could be recommended to practitioners for a large class of problems. Based on the knowledge acquired in this paper, we plan to pursue our study and establish comparisons with several (t, s)-sequences in the near future.

ELECTRONIC APPENDIX

The electronic appendix for this article can be accessed in the ACM Digital Library by visiting the following URL: `http://www.acm.org/pubs/citations/journals/jn/2008-V-N/p1-URLend`.

ACKNOWLEDGMENTS

We wish to thank Harald Niederreiter for his useful comments on this work. We also thank Emanouil Atanassov for useful discussions on his modified sequence. Xiaoheng Wang—while he was an undergraduate student in the Faculty of Mathematics at the University of Waterloo—helped improve the running time of some computations, and we also thank him for useful insight and discussion on this work. This revised version has greatly benefited from comments of three reviewers. The second author gratefully acknowledges the financial support of NSERC.

REFERENCES

Atanassov, E. 2004. On the discrepancy of the Halton sequences. *Mathematica Balkanica 18*, 15–32.

Atanassov, E. and Durchova, M. 2003. Generating and testing the modified Halton sequences. In *Fifth International Conference on Numerical Methods and Applications, Borovets 2002*, Springer-Verlag, Ed. Lecture Notes in Computer Science, vol. 2542. Berlin, 91–98.

Bach, E. and Shallit, J. 1996. *Algorithmic Number Theory*. Vol. 1. MIT Press, Cambridge, MA.

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

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

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

Chi, H., Mascagni, M., and Warnock, T. 2005. On the optimal Halton sequence. *Mathematics and Computers in Simulation 70*, 9–21.

Faure, H. 1978. Discrépance de suites associées à un système de numération. *C. R. Acad. Sc.,Paris 286–A*, 293–296.

Faure, H. 1982. Discrépance des suites associées à un système de numération (en dimension s). *Acta Arithmetica 41*, 337–351.

Faure, H. 1986. On the star-discrepancy of generalized Hammersley sequences in two dimensions. *Monatsh. Math. 101*, 291–300.

Faure, H. 1992. Good permutations for extreme discrepancy. *J. Number Theory 42,* 1, 47–56.

Faure, H. 2005. Discrepancy and diaphony of digital (0,1)-sequences in prime base. *Acta Arithmetica 117*, 125–148.

Faure, H. 2006. Selection criteria for (random) generation of digital $(0, s)$-sequences. In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, H. Niederreiter and D. Talay, Eds. Springer, 113–126.

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

Kocis, L. and Whiten, W. J. 1997. Computational investigations of low-discrepancy sequences. *ACM Transactions on Mathematical Software 23,* 2, 266–294.

Kritzer, P. 2006. Improved upper bounds on the star discrepancy of (t, m, s)-nets and (t, s)-sequences. *Journal of Complexity 22*, 336–347.

L'Ecuyer, P. and Lemieux, C. 2002. Recent advances in randomized quasi-Monte Carlo methods. In *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*, M. Dror, P. L'Ecuyer, and F. Szidarovszki, Eds. Kluwer Academic Publishers, Boston, 419–474.

Lemieux, C. and Owen, A. 2001. Quasi-regression and the relative importance of the ANOVA components of a function. In *Monte Carlo and Quasi-Monte Carlo Methods 2000*, K.-T. Fang, F. J. Hickernell, and H. Niederreiter, Eds. Springer, 331–344.

Mascagni, M. and Chi, H. 2004. On the scrambled Halton sequence. *Monte Carlo Methods and Applications 10*, 435–442.

Matoušek, J. 1998. On the L_2-discrepancy for anchored boxes. *J. Complexity 14*, 527–556.

Morokoff, W. J. and Caflisch, R. E. 1994. Quasi-random sequences and their discrepancies. *SIAM J. Sci. Comp. 15*, 1251–1279.

Niederreiter, H. 1978. Quasi-Monte Carlo methods and pseudorandom numbers. *Bulletin of the American Mathematical Society 84,* 6, 957–1041.

Niederreiter, H. 1987. Point sets and sequences with small discrepancy. *Monatsh. Math. 104*, 273–337.

Niederreiter, H. 1988. Low discrepancy and low dispersion sequences. *J. Number Theory 30*, 51–70.

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

Niederreiter, H. and Xing, C. 1996. Quasirandom points and global function fields. In *Finite fields and applications (Glasgow, 1995)*. London Math. Soc. Lecture Note Ser., vol. 233. Cambridge Univ. Press, Cambridge, 269–296.

Ninomiya, S. and Tezuka, S. 1996. Toward real-time pricing of complex financial derivatives. *Applied Mathematical Finance 3*, 1–20.

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

Owen, A. B. 2003. The dimension distribution and quadrature test functions. *Stat. Sinica 13*, 1–17.

Pagès, G. 1992. Van der Corput sequences, Kakutani transforms and one-dimensional numerical integration. *J. Comp. Appl. Math. 44*, 21–39.

Papageorgiou, A. and Traub, J. 1997. Faster evaluation of multidimensional integrals. *Computers in Physics 11*, 547–578.

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

Radovic, I., Sobol', I. M., and Tichy, R. F. 1996. Quasi-Monte Carlo methods for numerical integration: Comparison of different low-discrepancy sequences. *Monte Carlo Methods and Applications 2*, 1–14.

Schürer, R. and Schmid, W. 2006. MinT: a database for optimal net parameters. In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, H. Niederreiter and D. Talay, Eds. Springer, Berlin, 457–469.

Sobol', I. M. 1967. On 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.

Sobol', I. M. and Asotsky, D. I. 2003. One more experiment on estimating high-dimensional integrals by quasi-Monte Carlo methods. *Math. Comput. Simul. 62*, 255–263.

Struckmeier, J. 1995. Fast generation of low-discrepancy sequences. *J. Comp. Appl. Math. 91*, 29–41.

Tezuka, S. 1993. Polynomial arithmetic analogue of Halton sequences. *ACM Transactions on Modeling and Computer Simulation 3*, 99–107.

Tezuka, S. 1995. *Uniform Random Numbers: Theory and Practice*. Kluwer Academic Publishers, Norwell, MA.

Tuffin, B. 1998. A new permutation choice in Halton sequences. In *Monte Carlo and Quasi-Monte Carlo Methods 1996*, P. Hellekalek, G. Larcher, H. Niederreiter, and P. Zinterhof, Eds. Lecture Notes in Statistics, vol. 127. Springer-Verlag, New York, 427–435.

Vandewoestyne, B. and Cools, R. 2006. Good permutations for deterministic scrambled Halton sequences in terms of L_2-discrepancy. *J. Comp. Appl. Math. 189*, 341–361.

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

Wang, X. and Sloan, I. H. 2005. Why are high-dimensional finance problems of low effective dimension? *SIAM Journal on Scientific Computing 27*, 159–183.

Warnock, T. 1972. Computational investigations of low discrepancy point sets. In *Application de la théorie des nombres à l'analyse numérique*, S. Zaremba, Ed. Academic Press, New York, 319–343.

Warnock, T. 1995. Computational investigations of low discrepancy point sets ii. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, H. Niederreiter and P. J.-S. Shiue, Eds. Number 106 in Lecture Notes in Statistics. Springer-Verlag, 354–361.

Warnock, T. 2002. Effective error estimates for quasi-Monte Carlo computations. Manuscript available at public.lanl.gov/kmh/uncertainty/meetings/warnvgr.pdf.

...

THIS DOCUMENT IS THE ONLINE-ONLY APPENDIX TO:

Generalized Halton Sequences in 2008: A Comparative Study

HENRI FAURE
Institut Mathématiques de Luminy and Université Paul Cézanne
and
CHRISTIANE LEMIEUX
University of Waterloo

ACM Journal Name, Vol. V, No. N, October 2008, Pages 1–30.

A. LIST OF FACTORS

Table VII gives the first 360 factors for the FL sequence. These multipliers can also be downloaded from `www.math.uwaterloo.ca/~clemieux/FLFactors.html`.

B. GETTING ASYMPTOTIC ORDERS FOR C_S

Here, we use the following results from number theory [Bach and Shallit 1996]: (i) for the n-th prime p_n ($n \geq 3$), we have $n \log n + 1 < p_n < n^{\frac{3}{2}}$; (ii) from Stirling formula, we have that $s \log s - s < \log(s!) < s \log s$; (iii) below, $q_s \geq s$ represents the base used in dimension s for the Faure sequence, and we will make use of Bertrand's postulate, which says that we can find a prime q_s satisfying $s \leq q_s \leq 2s$.

Also, it is easier to compare the logarithms, so we will compare the quantities λ_s and μ_s for $\log c_s$ with c_s given by (8) and (9), which are respectively for the Halton and Halton-Atanassov sequences. Similarly, ρ_s, τ_s and ν_s are used to denote $\log c_s$ for the value of c_s given in (11), (10), and (12), which are respectively for $(0, s)$-sequences in varying base q_s, Sobol' sequences in base 2, and (t, s)-sequences in a fixed base b.

Hence we get:

$$\begin{aligned}
\lambda_s &= \sum_{j=1}^{s} \log\Big(\frac{p_j - 1}{\log p_j}\Big) - \log(s!) \\
\mu_s &= s \log 2 + \log\Big(\sum_{j=1}^{s} \log p_j \prod_{j=1}^{s} \frac{p_j(1 + \log p_j)}{(p_j - 1)\log p_j}\Big) - \log(s!) \\
\rho_s &= s \log\Big(\frac{q_s - 1}{\log q_s}\Big) - \log(s!) \\
\tau_s &= (t_s + s)\log 2 - s \log\log 2 - \log(s!),
\end{aligned}$$

Table VII. First 360 factors for the FL sequence

j	f_j	j	f_j	j	f_j	j	f_j	j	f_j	j	f_j				
1	1	46	152	91	269	136	431	181	305	226	1007	271	1075	316	856
2	1	47	114	92	292	137	295	182	460	227	622	272	682	317	1229
3	3	48	80	93	215	138	557	183	599	228	549	273	1245	318	1619
4	3	49	83	94	182	139	172	184	335	229	613	274	401	319	774
5	4	50	97	95	294	140	343	185	258	230	799	275	774	320	1229
6	9	51	95	96	152	141	472	186	649	231	408	276	1026	321	1300
7	7	52	150	97	148	142	604	187	771	232	856	277	499	322	1563
8	5	53	148	98	144	143	297	188	619	233	601	278	1314	323	1551
9	9	54	55	99	382	144	524	189	666	234	1072	279	743	324	1265
10	18	55	80	100	194	145	251	190	669	235	938	280	693	325	905
11	18	56	192	101	346	146	514	191	707	236	322	281	1282	326	1333
12	8	57	71	102	323	147	385	192	737	237	1142	282	1003	327	493
13	13	58	76	103	220	148	531	193	854	238	873	283	1181	328	913
14	31	59	82	104	174	149	663	194	925	239	629	284	1079	329	1397
15	9	60	109	105	133	150	674	195	818	240	1071	285	765	330	1250
16	19	61	105	106	324	151	255	196	424	241	1063	286	815	331	612
17	36	62	173	107	215	152	519	197	493	242	1205	287	1350	332	1251
18	33	63	58	108	246	153	324	198	463	243	596	288	1144	333	1765
19	21	64	143	109	159	154	391	199	535	244	973	289	1449	334	1303
20	44	65	56	110	337	155	394	200	782	245	984	290	718	335	595
21	43	66	177	111	254	156	533	201	476	246	875	291	805	336	981
22	61	67	203	112	423	157	253	202	451	247	918	292	1203	337	671
23	60	68	239	113	484	158	717	203	520	248	1133	293	1173	338	1403
24	56	69	196	114	239	159	651	204	886	249	1223	294	737	339	820
25	26	70	143	115	440	160	399	205	340	250	933	295	562	340	1404
26	71	71	278	116	362	161	596	206	793	251	1110	296	579	341	1661
27	32	72	227	117	464	162	676	207	390	252	1228	297	701	342	973
28	77	73	87	118	376	163	425	208	381	253	1017	298	1104	343	1340
29	26	74	274	119	398	164	261	209	274	254	701	299	1105	344	1015
30	95	75	264	120	174	165	404	210	500	255	480	300	1379	345	1649
31	92	76	84	121	149	166	691	211	581	256	678	301	827	346	855
32	47	77	226	122	418	167	604	212	345	257	1172	302	1256	347	1834
33	29	78	163	123	306	168	274	213	363	258	689	303	759	348	1621
34	61	79	231	124	282	169	627	214	1024	259	1138	304	540	349	1704
35	57	80	177	125	434	170	777	215	514	260	1022	305	1284	350	893
36	69	81	95	126	196	171	269	216	773	261	682	306	1188	351	1033
37	115	82	116	127	458	172	217	217	932	262	613	307	776	352	721
38	63	83	165	128	313	173	599	218	556	263	635	308	853	353	1737
39	92	84	131	129	512	174	447	219	954	264	984	309	1140	354	1507
40	31	85	156	130	450	175	581	220	793	265	526	310	445	355	1851
41	104	86	105	131	161	176	640	221	294	266	1311	311	1265	356	1006
42	126	87	188	132	315	177	666	222	863	267	459	312	802	357	994
43	50	88	142	133	441	178	595	223	393	268	1348	313	932	358	923
44	80	89	105	134	549	179	669	224	827	269	477	314	632	359	872
45	55	90	125	135	555	180	686	225	527	270	716	315	1504	360	1860

and for ν_s we have (using the bounds of Kritzer [2006] for $s \geq 5$)

$$\nu_s = \begin{cases} t_s \log b + s(\log(b-1) - \log\log b) - \log 2 - \log(s!) & \text{for } b \text{ odd} \\ t_s \log b + s(\log b - \log\log b) + \log \frac{b-1}{2(b+1)} - \log(s!) & \text{for } b \geq 4 \text{ even} \end{cases}$$

In the formula for τ_s, we will use the bound $t_s \leq (s \log s + s \log\log s)/\log 2 +$

$o(s \log\log s)$, while as mentioned before, $t_s = O(s)$ in the best constructions of Niederreiter and Xing.

With the help of our reminders on prime numbers and with $s \log s - s < \log(s!) < s \log s$ (from Stirling's formula), we get

$$\begin{aligned}
-s\log(1.5) < \lambda_s &< -\log s + O(1),\\
-s\log s < \mu_s &< -s\log s + s(2+\log 2) + \log\log(s!) + \log(1.5) = -s\log s + O(s),\\
-s\log\log s < \rho_s &< -s\log\log s + s(1+\log(2)) = -s\log\log s + O(s),\\
\tau_s &< -s\log s + t_s \log 2 + s(1+\log 2) = s\log\log s + O(s),\\
-s\log s < \nu_s &< -s\log s + t_s \log b + s(1+\log b) = -s\log s + O(s).
\end{aligned}$$

C. EXPERIMENTS ON G_1

For the function g_1, as mentioned in Section 6 of the paper, the effective dimension in the truncation sense decreases as we go from case (i) to (iv). Furthermore, case (v) is the same as (iv) but with the order of the coordinates reversed. Hence (v) has a truncated effective dimension that is presumably equal to the nominal dimension s: however the superposed effective dimension is no larger than the truncated effective dimension of (iv).

What we see for the case (i) is that when $s = 20$ (Figure 10), some constructions are already worse than MC: in fact, only FL and KW are always better than MC. When $s = 50$ (Figure 11), the deterministic versions of H and VC are quite bad. When randomized, all methods are usually better than MC, but the large jumps in the error or variance (due to an evaluation of f in the problematic region near $(1, \ldots, 1)$) might cause some methods to become worse than MC: it is observed with the variance of KW. We also show on Figure 12 results for N up to 10^6. There, we see that H is still showing jumps in its error around $N = 800,000$, while, for example, FL seems to settle around $N = 200,000$. When $s = 150$, Figure 13 shows that the deterministic H and VC methods are very bad; KW is worse than MC; FL and AD are "good", but what happens here is that the function has such a huge peak that what seems the best possible behavior is to return an estimate of 0 with error 1. In fact, all methods end up doing that when randomized.

For the case (ii), with $s = 20$ (Figure 14) all methods are better than MC, and FL seems to be the best in the randomized case, at least for larger values of N. Already when $s = 50$ (Figure 15), deterministic H and VC are much worse than MC, and KW is also slightly worse. AD struggles to be better than MC. So here FL and CMW seem best. With $s = 150$ (Figure 16), in the deterministic case the results are similar to case (i), but with the randomization, some methods (WH, AD, FL, KW) achieve an error of about 0.4 by the time $N = 100,000$.

The case (iii) and (iv) with $s = 20$ are similar (Figures 17 and 20), in that all methods perform quite well. However for (iii) when $s = 50$ or $s = 150$ (Figures 18 and 19), when the methods are randomized there is a clear difference between the performance of H, WH, and VC compared to FL, AD, KW and CMW (with $s = 50$). This does not happen with (iv), where all methods are quite comparable when $s = 50$ or $s = 150$ (Figures 21 and 22).

With (v), already at $s = 20$ we see on Figure 23 that H, VC, and WH are not doing as well as FL, AD, KW and CMW. When $s = 50$ (Figure 24), they are about

the same as MC, and then when $s = 150$ (Figure 25) they are much worse than MC. When $s = 50$ we see some cycles in H, VC and WH, while when $s = 150$, it is FL and AD who have cycles. We think perhaps KW avoid this cyclical behavior for its error because it uses permutations that are less regular than a simple multiplication by a factor, the latter being more subject to produce this type of behavior in large bases.

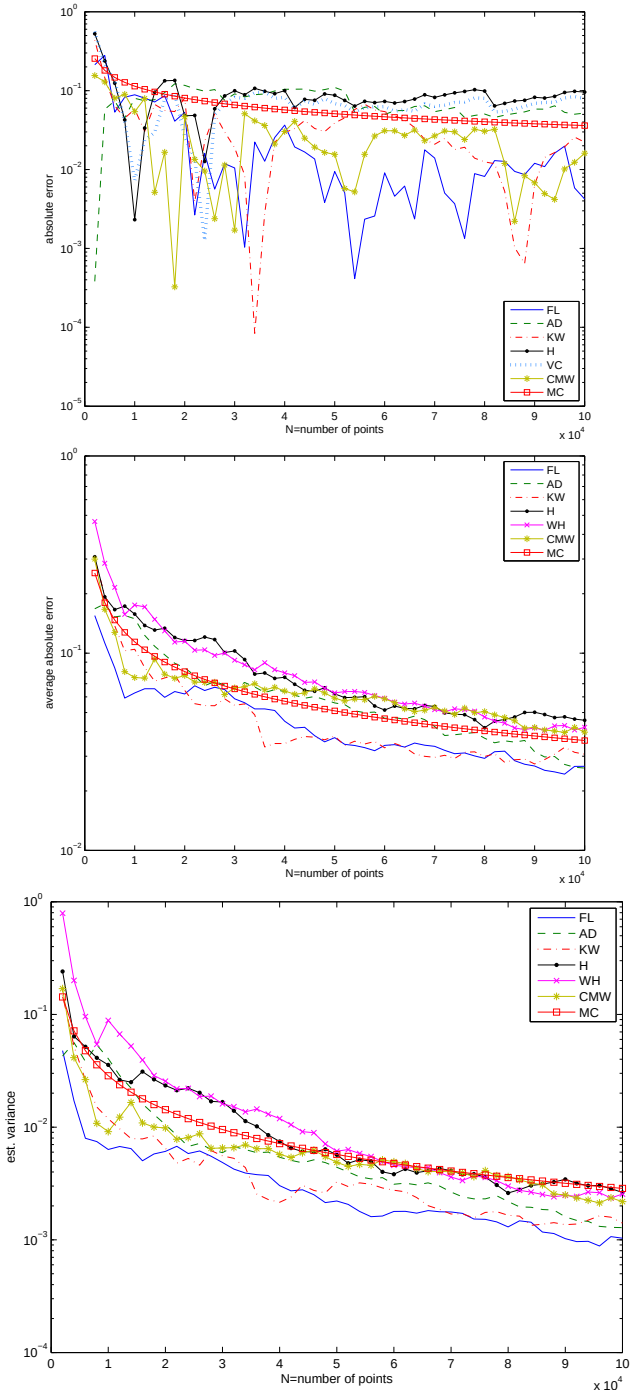

Fig. 10. Function g_1, case (i) with $s = 20$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

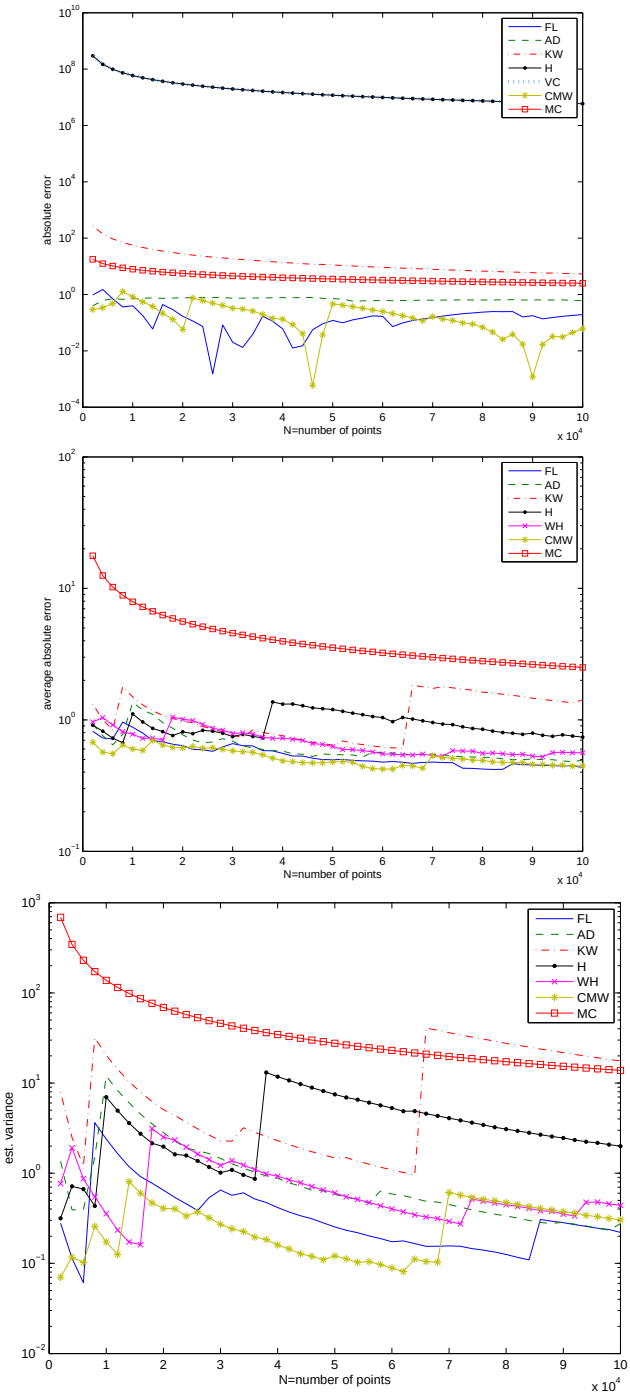

Fig. 11. Function g_1, case (i) with $s = 50$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

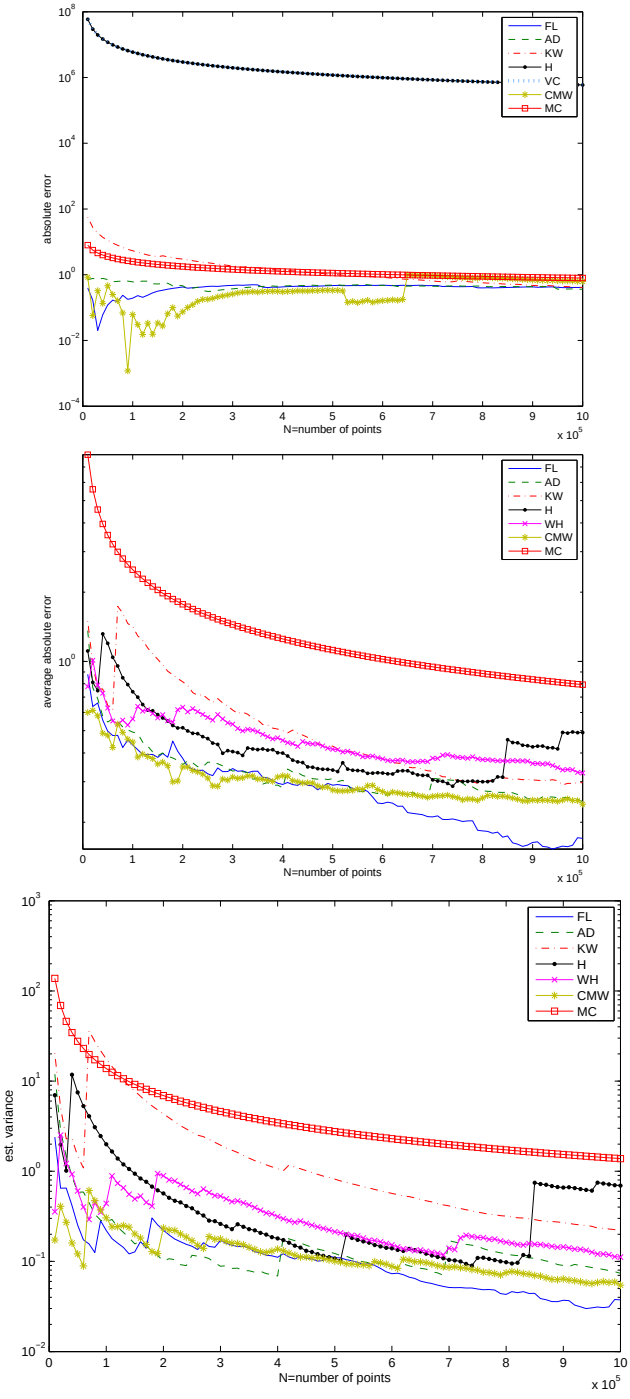

Fig. 12. Function g_1, case (i) with $s = 50$ and N up to 10^6; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

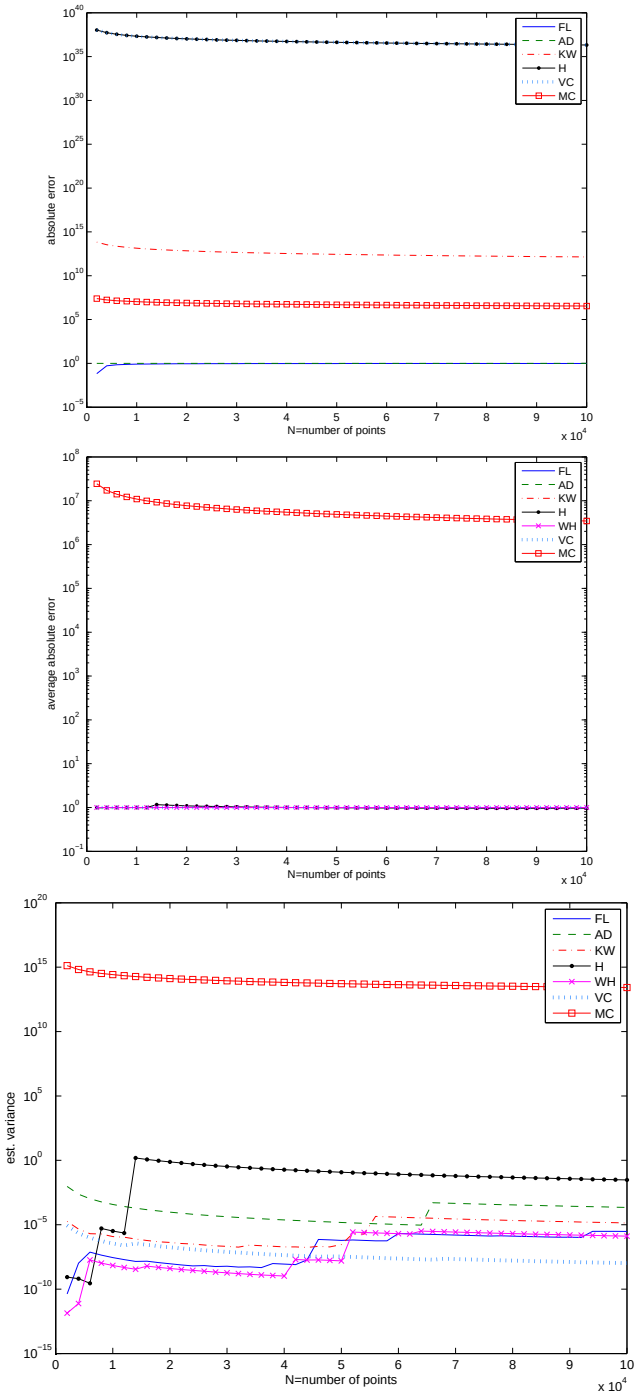

Fig. 13. Function g_1, case (i) with $s = 150$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

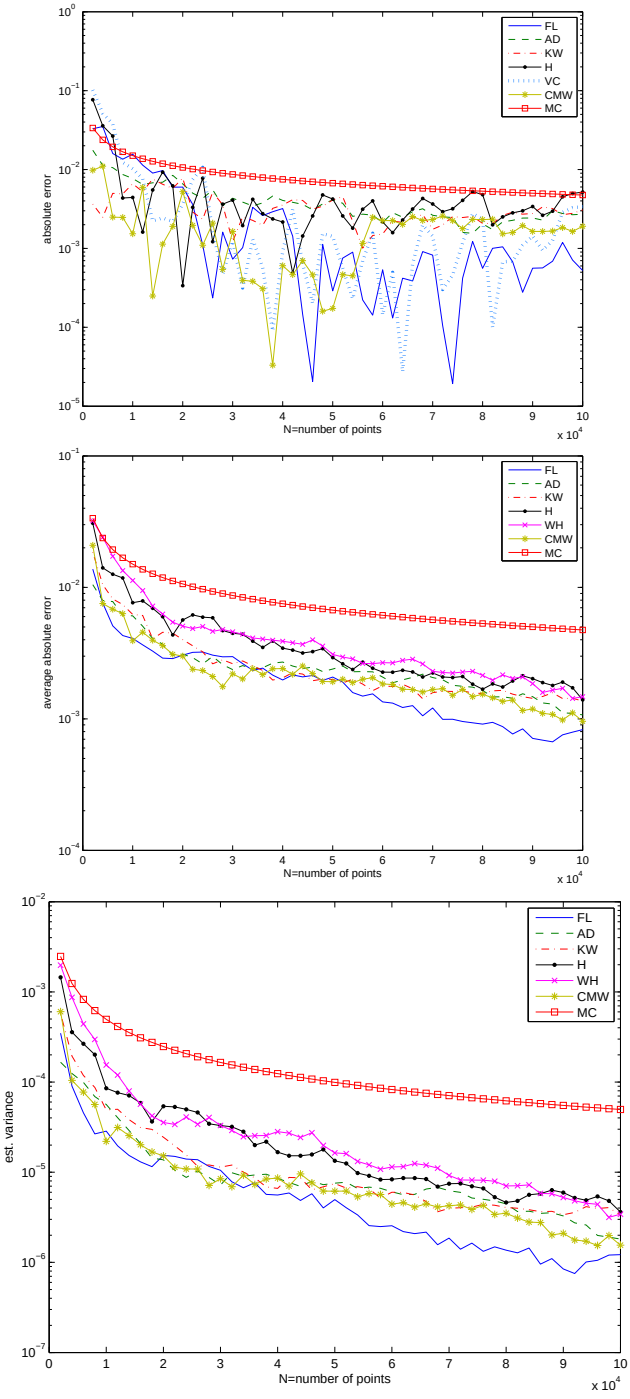

Fig. 14. Function g_1, case (ii) with $s = 20$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

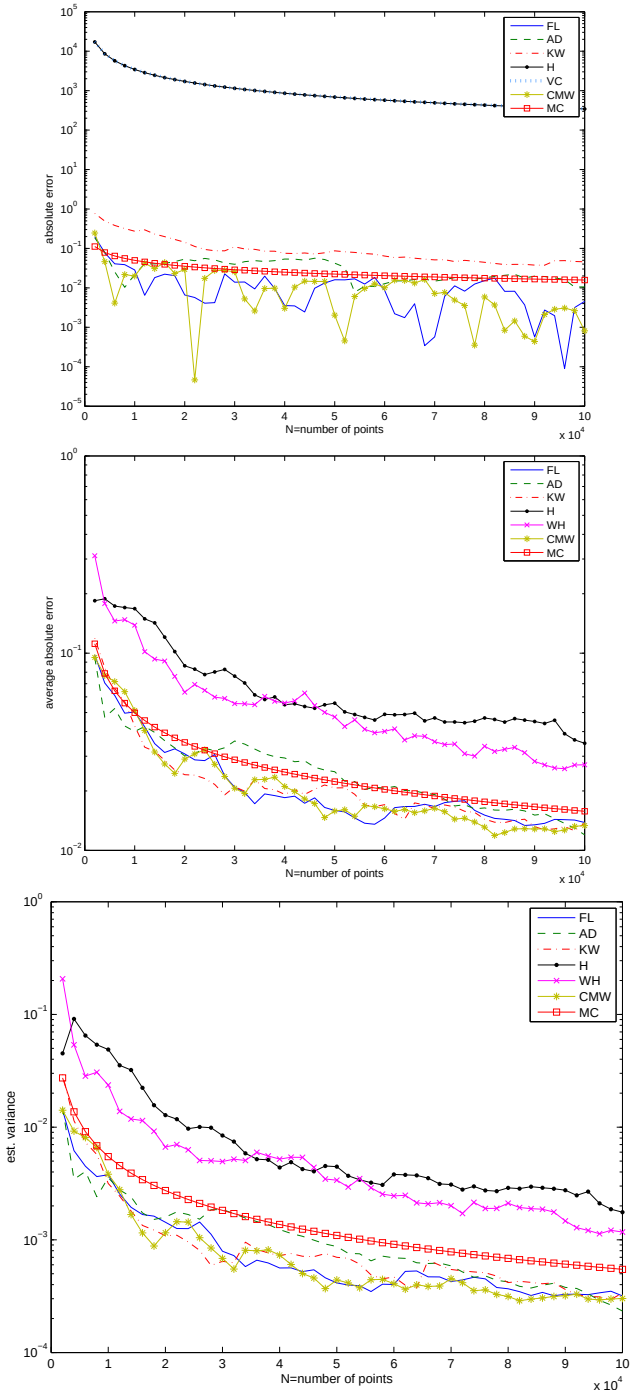

Fig. 15. Function g_1, case (ii) with $s = 50$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

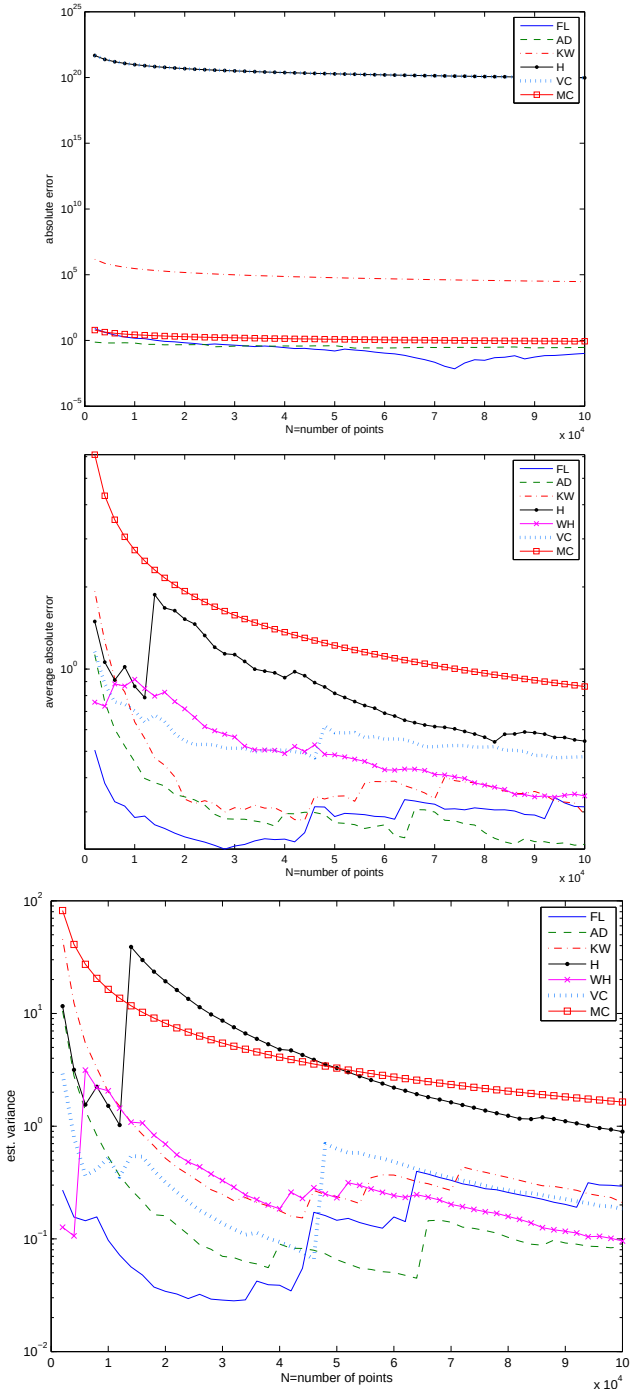

Fig. 16. Function g_1, case (ii) with $s = 150$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

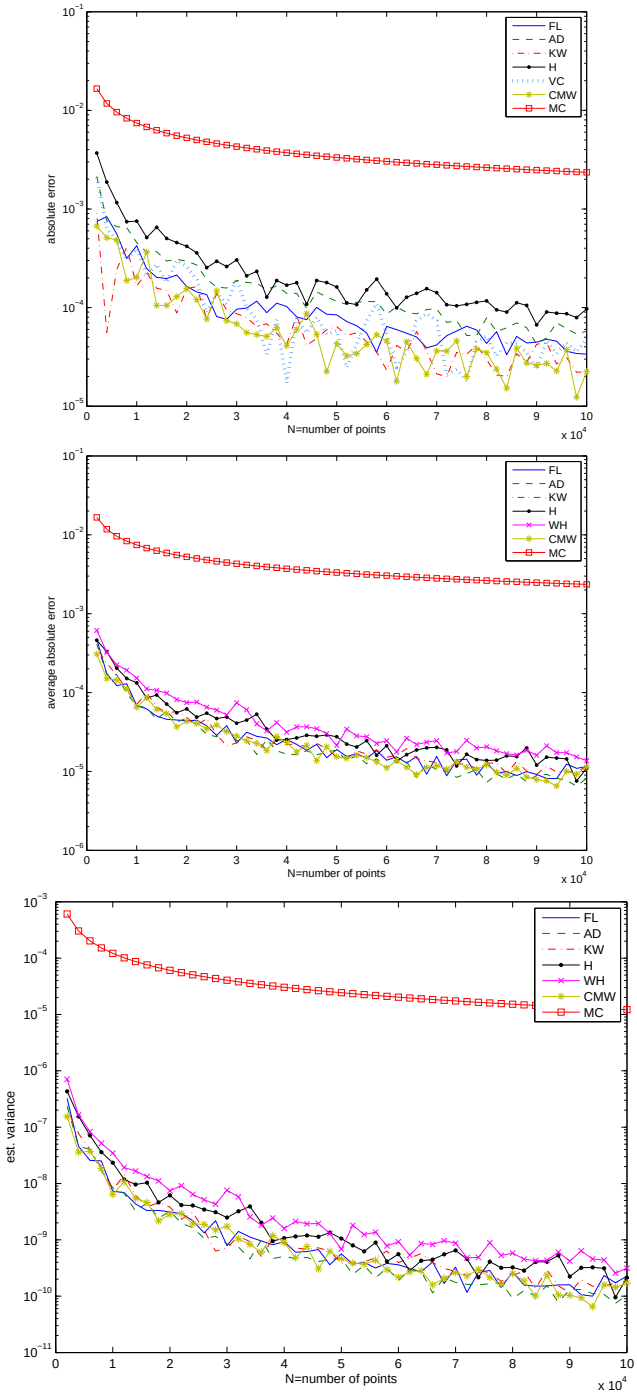

Fig. 17. Function g_1, case (iii) with $s = 20$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

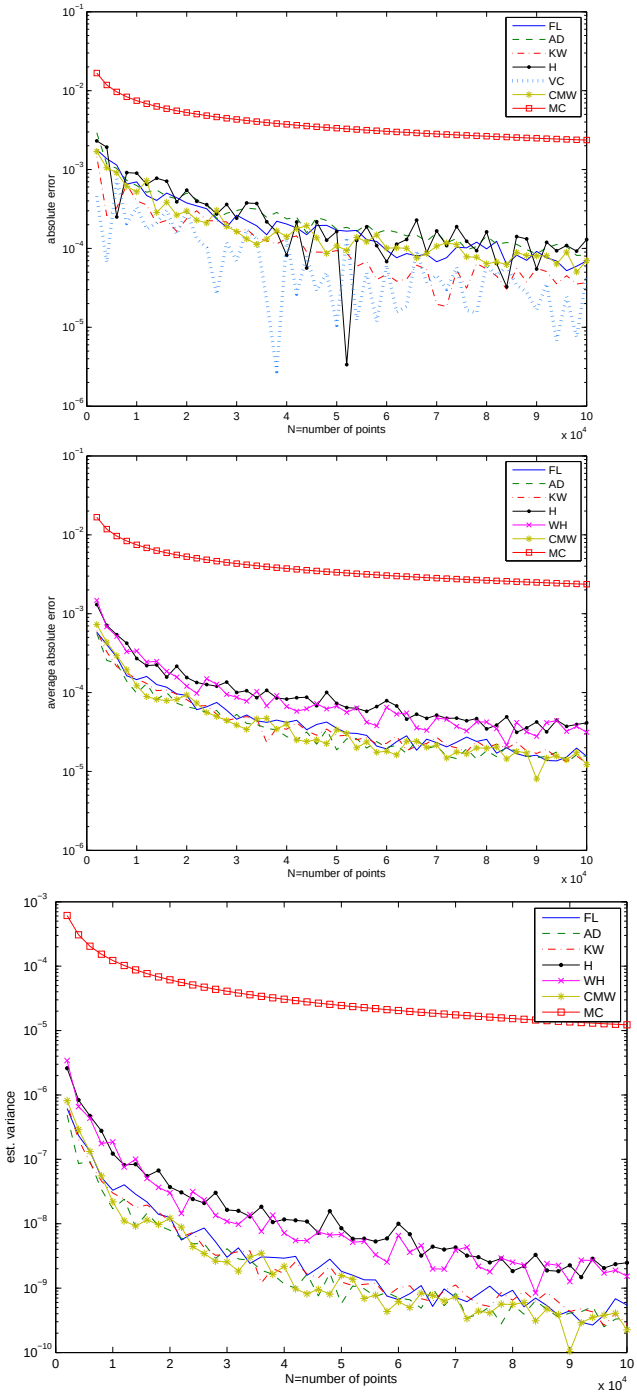

Fig. 18. Function g_1, case (iii) with $s = 50$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

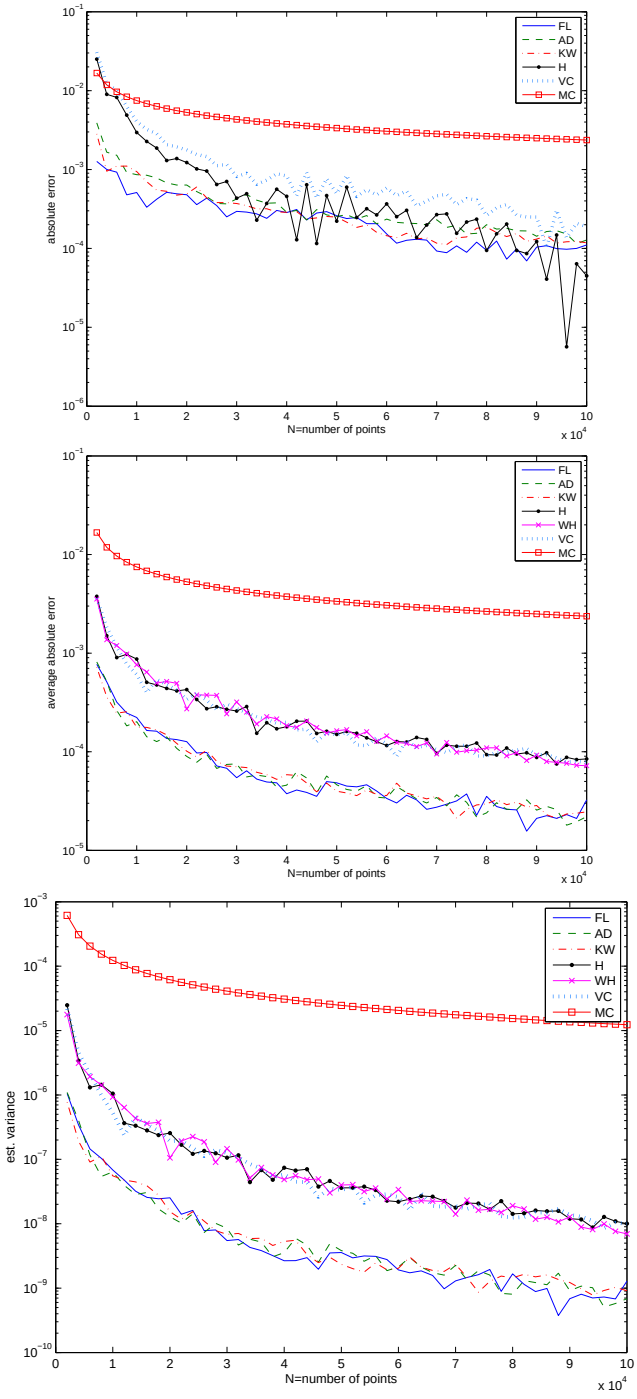

Fig. 19. Function g_1, case (iii) with $s = 150$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

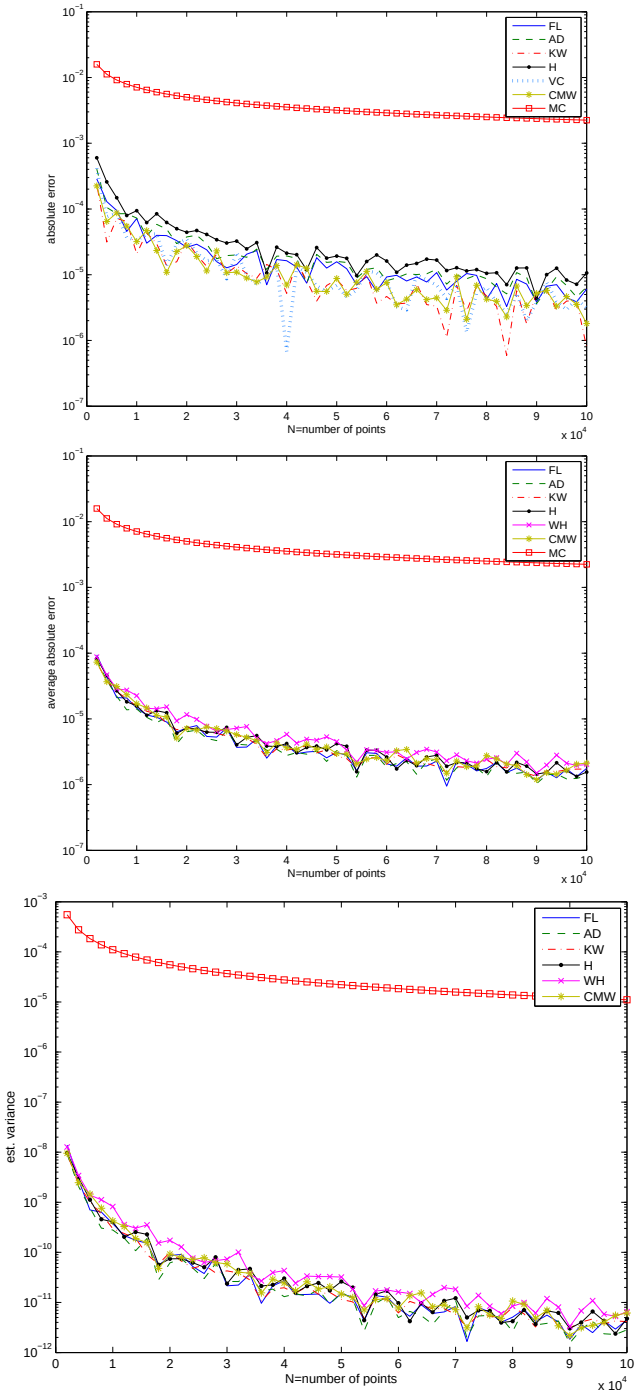

Fig. 20. Function g_1, case (iv) with $s = 20$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

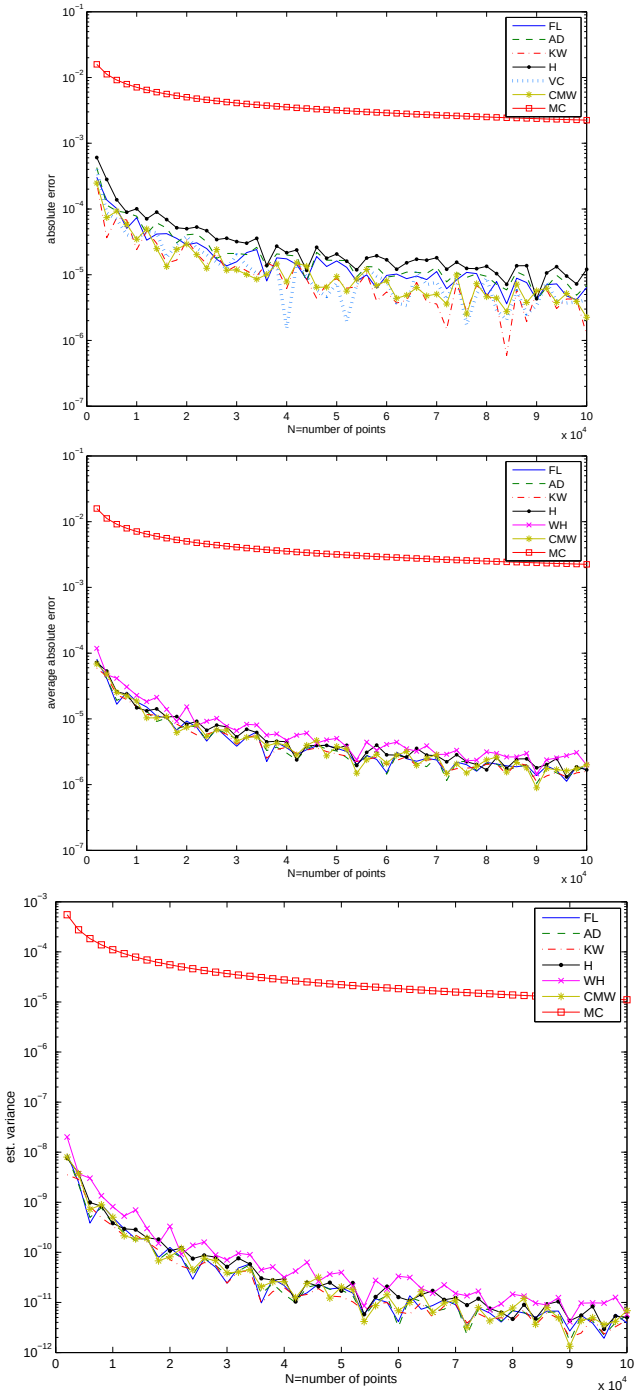

Fig. 21. Function g_1, case (iv) with $s = 50$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

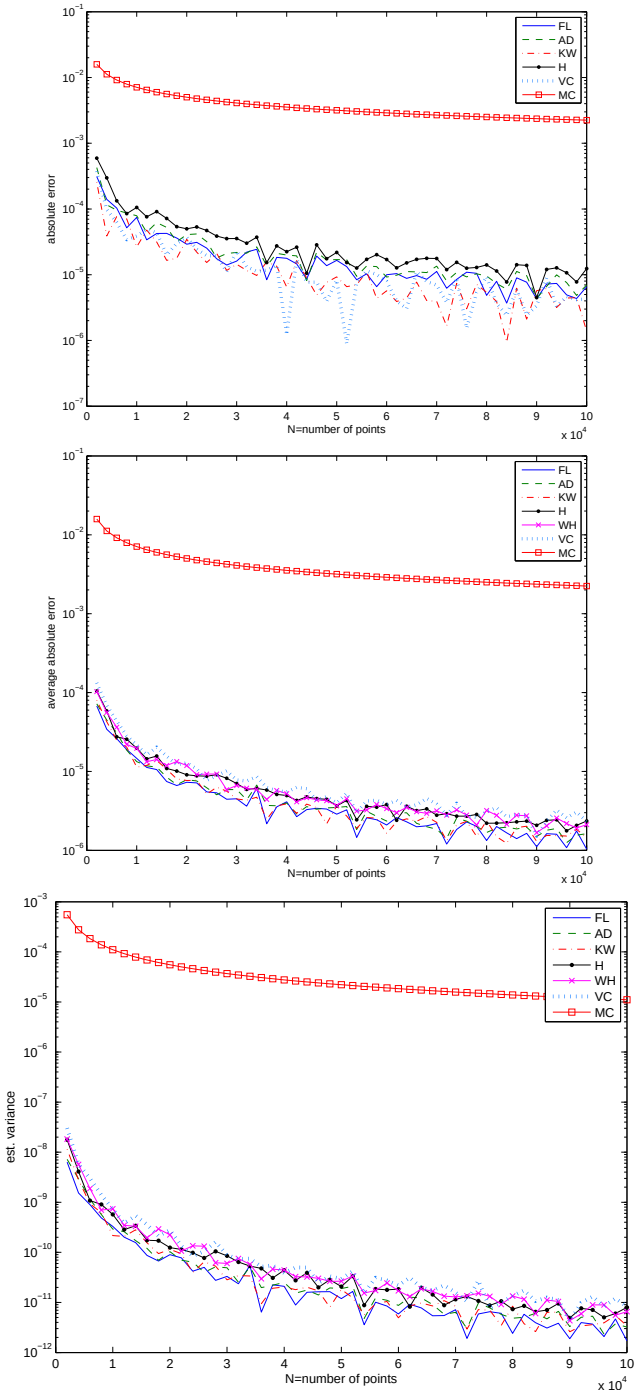

Fig. 22. Function g_1, case (iv) with $s = 150$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

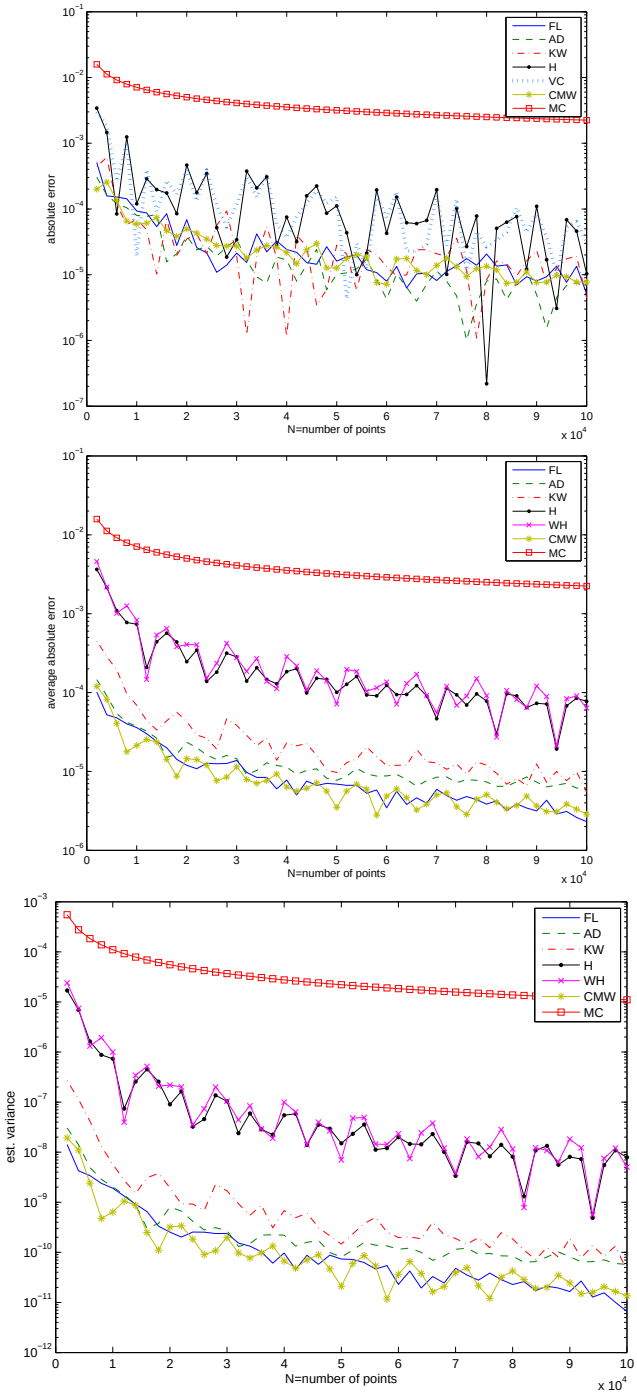

Fig. 23. Function g_1, case (v) with $s = 20$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

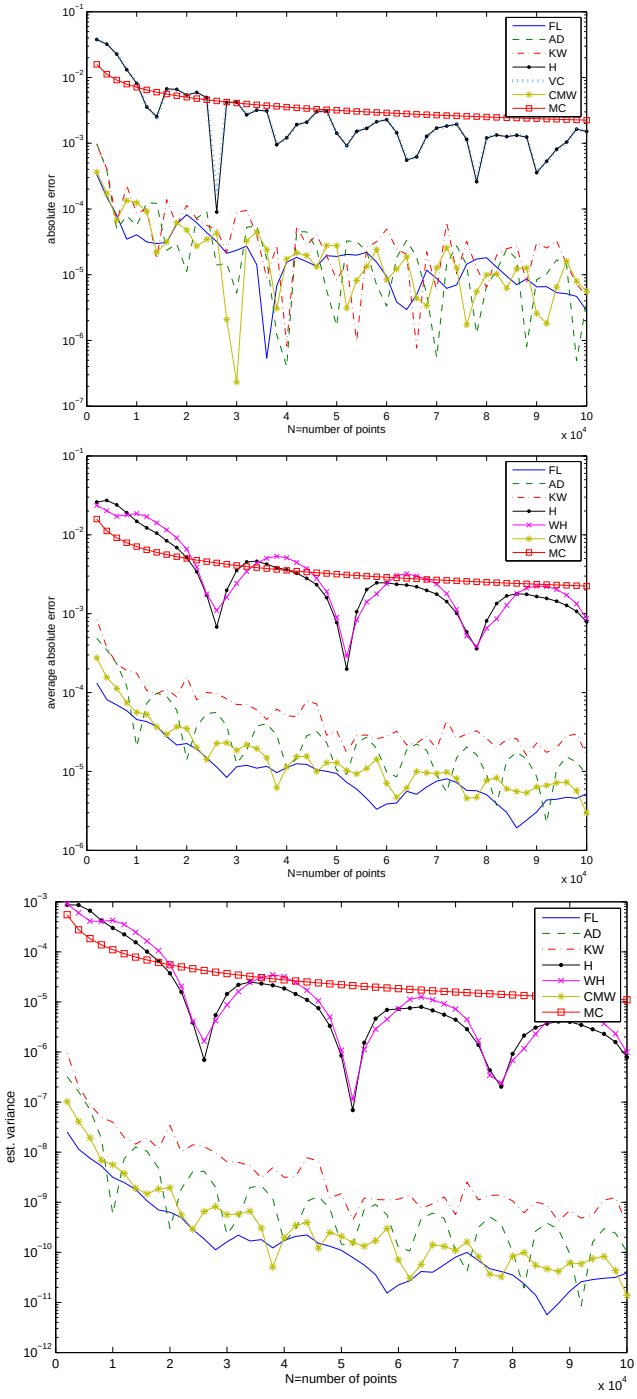

Fig. 24. Function g_1, case (v) with $s = 50$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

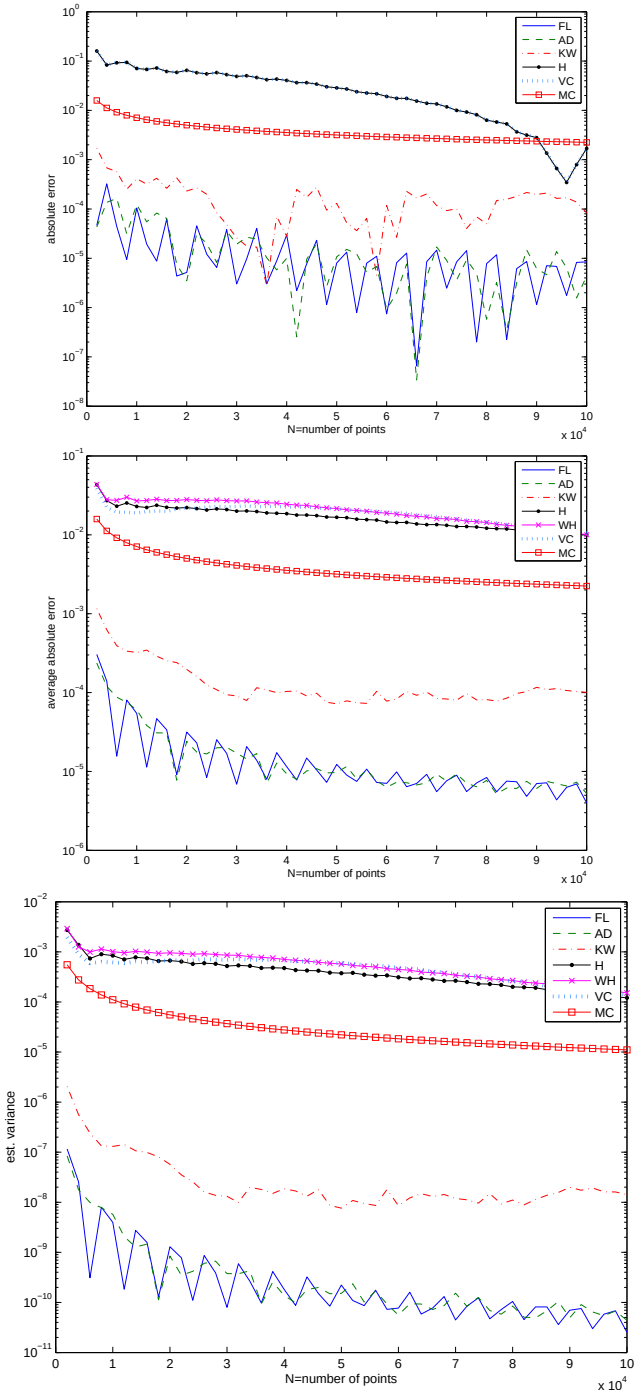

Fig. 25. Function g_1, case (v) with $s = 150$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

D. EXPERIMENTS ON G_2

For the function g_2, although the case $(c, s) = (0.1, 120)$ (Figure 26) has a small superposed effective dimension, we see that H, VC and WH are significantly less performing than AD, FL and KW, because the truncated effective dimension of this function is quite large, and thus poor low-dimensional projections for those higher indices are causing the approximation to be not so good. The same kind of behavior is observed in Figure 27, for the case $(c, s) = (0.25, 96)$. There, even with the randomization, H, VC and WH are worse than MC for smaller values of N. The (hard) case $(c, s) = (1, 150)$, shown in Figure 28, exhibits behavior similar to the one seen on the difficult cases (i) and (ii) for g_1. It is worth noting that deterministic KW does quite badly for this problem.

E. EXPERIMENTS ON G_3

For the function g_3, we see that with $s = 20$ (Figure 29), all sequences perform well, although H and WH seem slightly worse in the randomized versions. When $s = 50$ (Figure 30), H and VC do worse than MC in the deterministic case, and there is a clear separation between the performance of FL, AD, CMW and KW, versus H and WH in the randomized case. Similar observations—perhaps just amplified—can be made when $s = 120$, as can be seen in Figure 31.

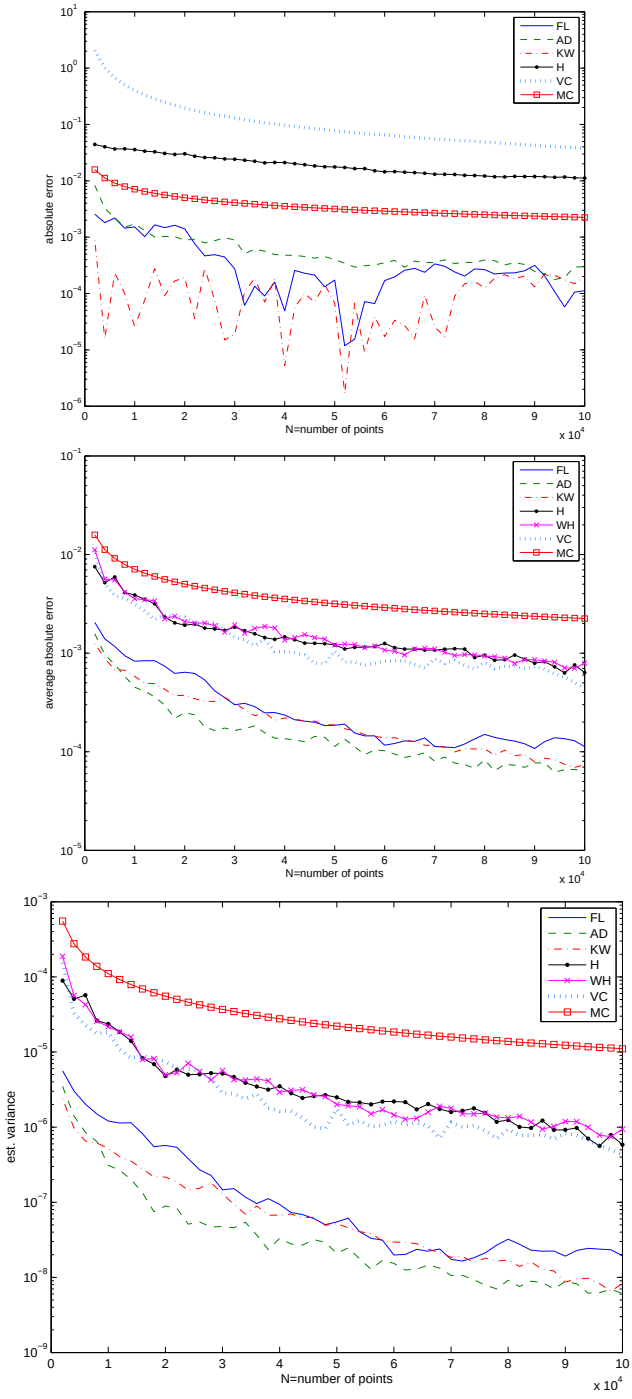

Fig. 26. Function g_2 with $c = 0.1$ and $s = 120$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

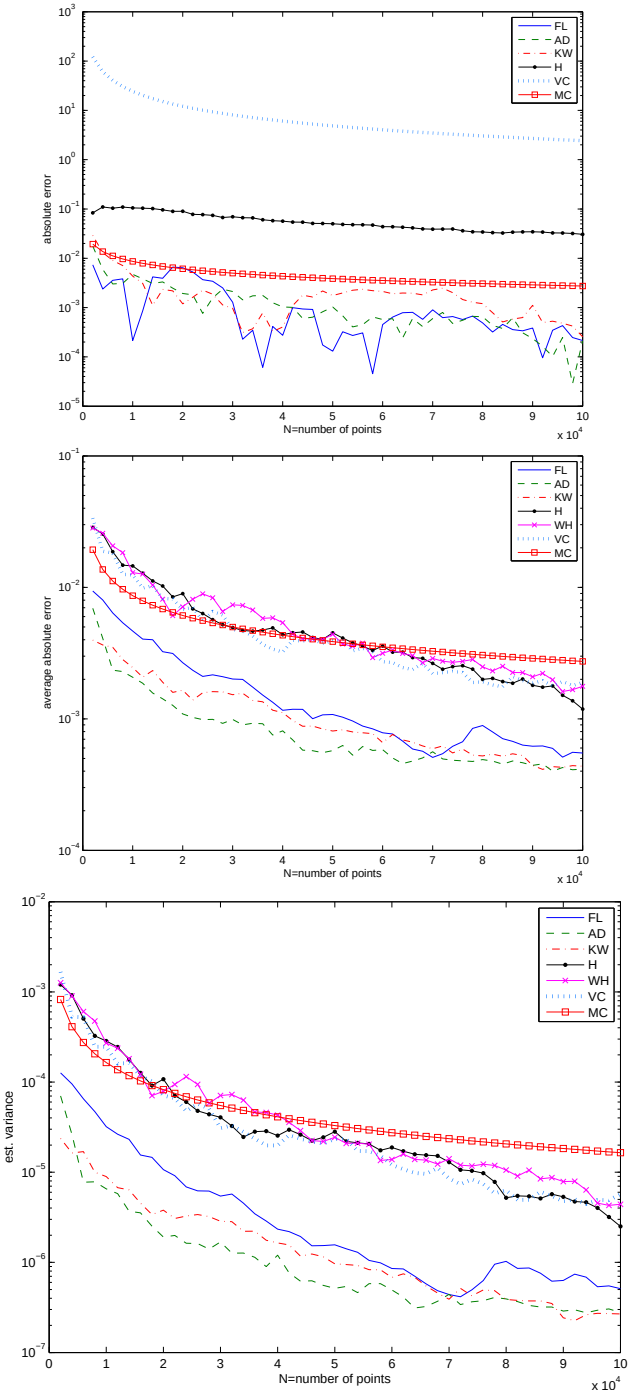

Fig. 27. Function g_2 with $c = 0.25$ and $s = 96$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

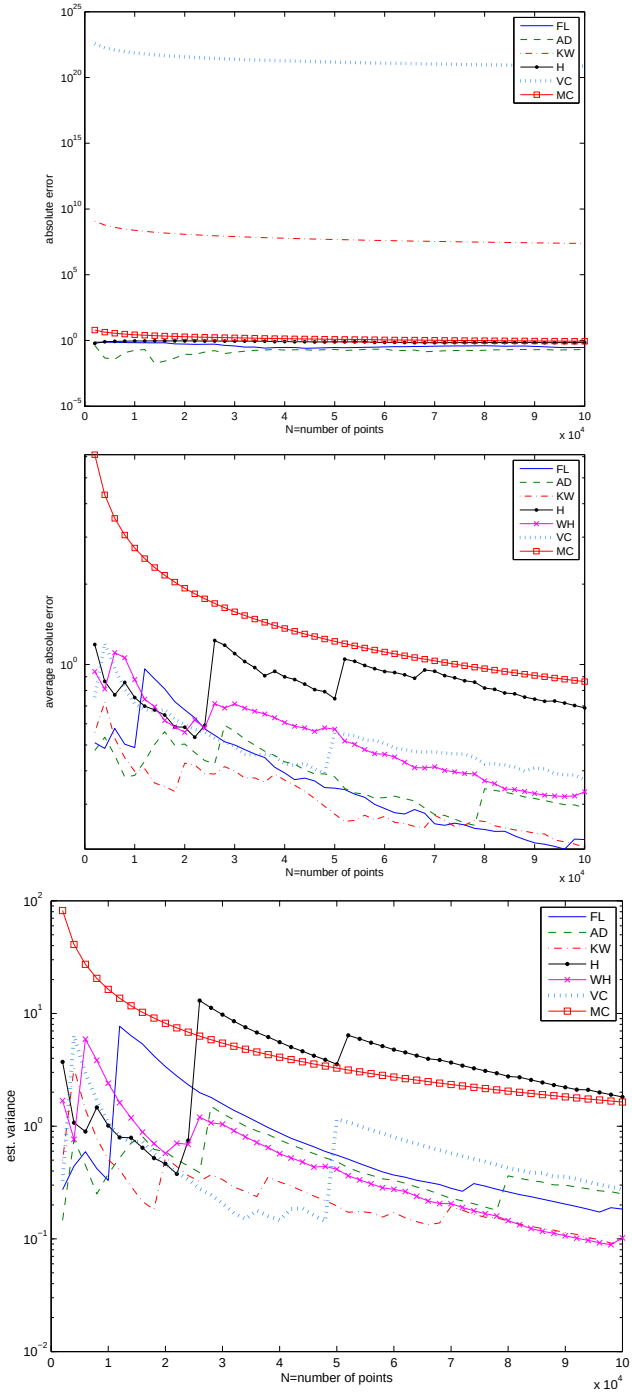

Fig. 28. Function g_2 with $c = 1$ and $s = 150$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

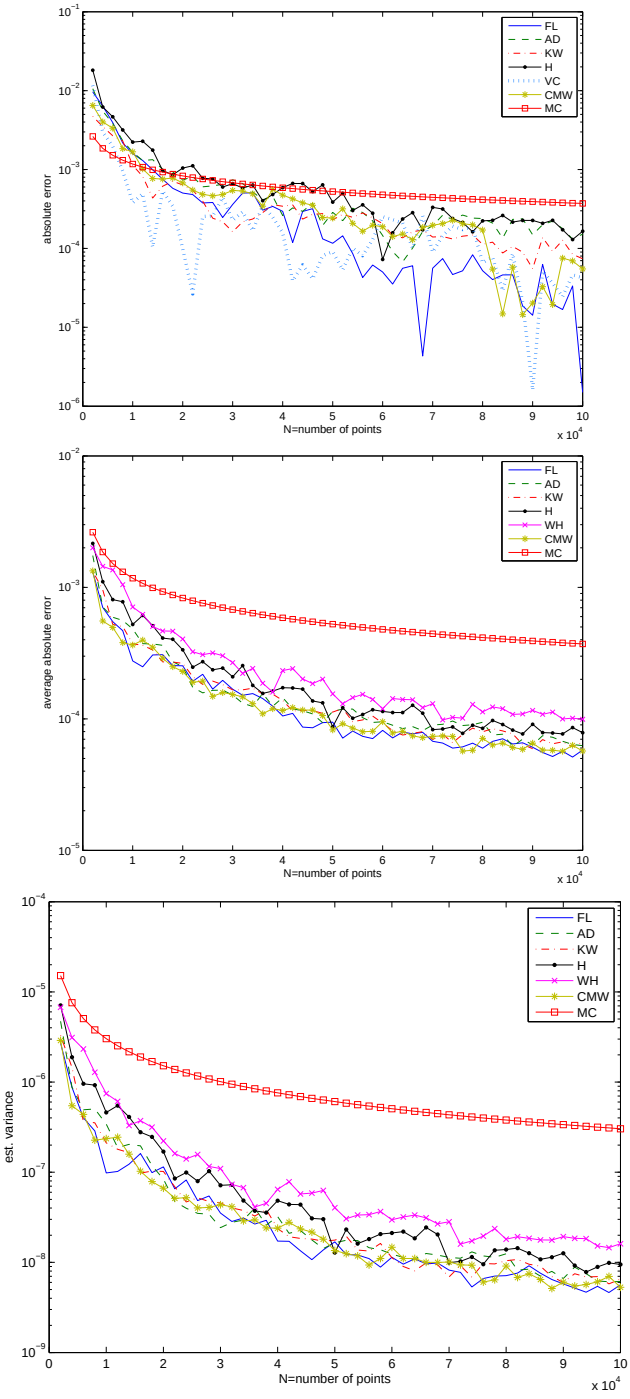

Fig. 29. Function g_3 with $s = 20$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

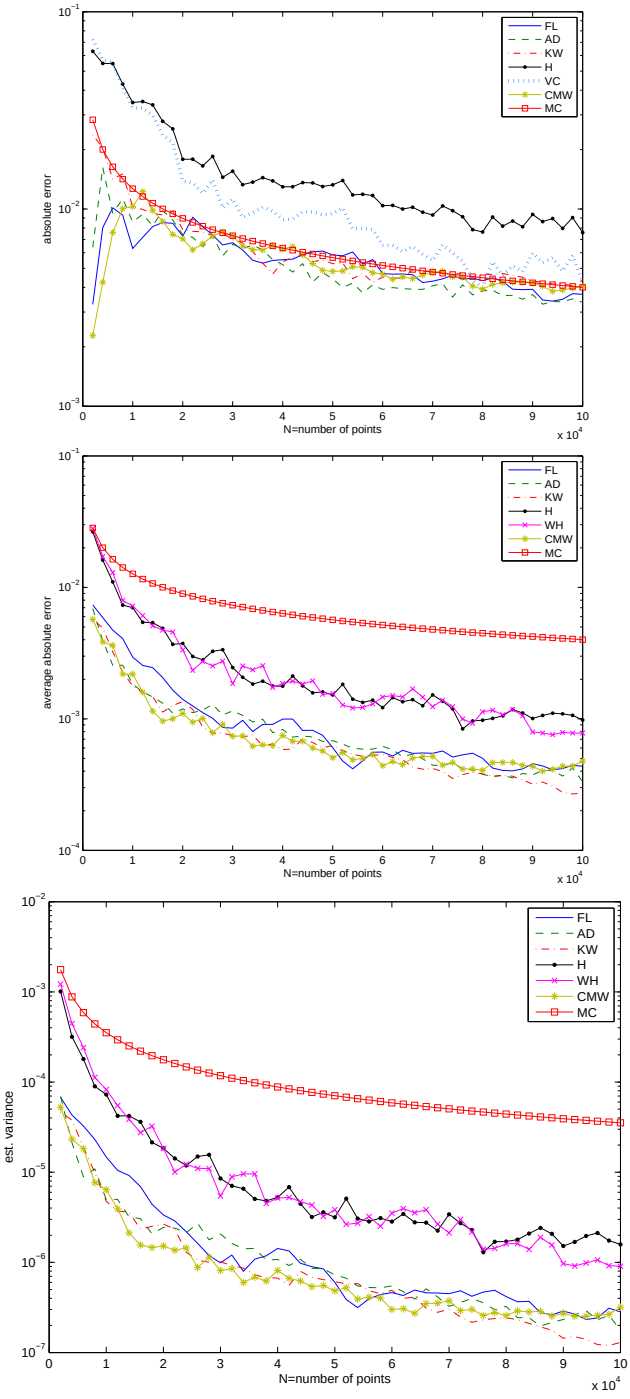

Fig. 30. Function g_3 with $s = 50$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

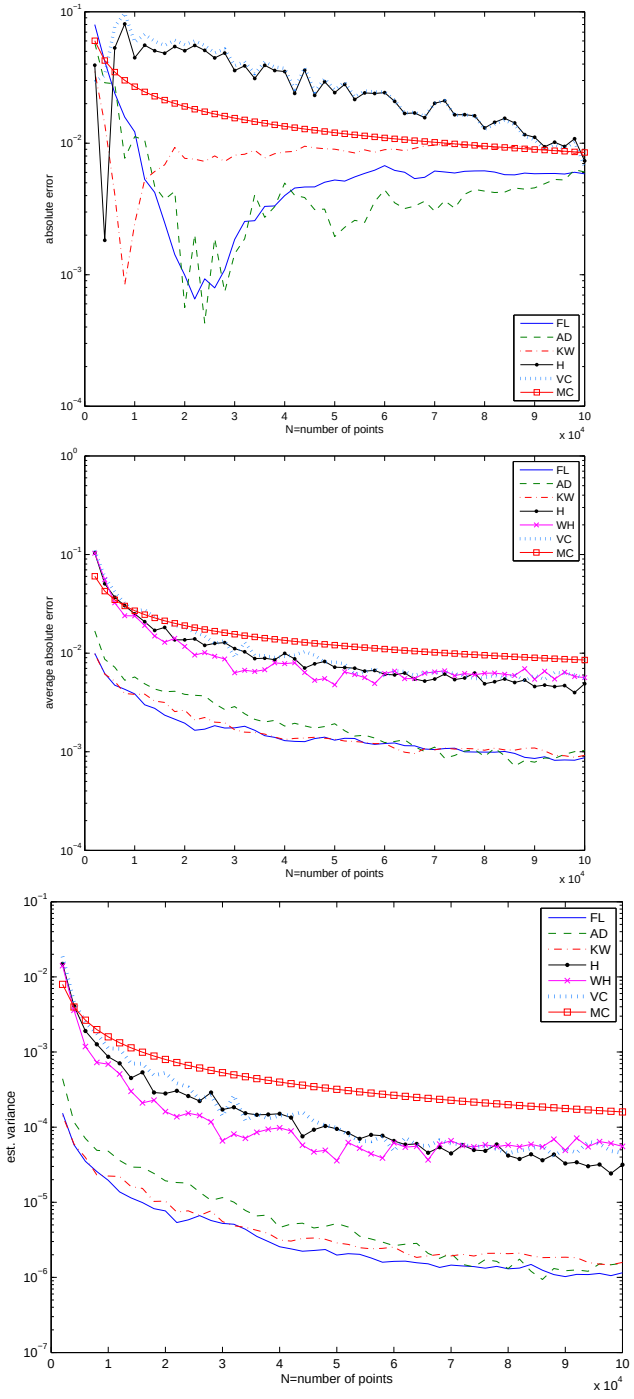

Fig. 31. Function g_3 with $s = 120$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

F. ASIAN OPTION

For the Asian option, for both $s = 40$ and $s = 75$, the deterministic versions of H and VC are worse than MC for all strike prices (top graph in Figures 32 to 38). With the randomization, when $s = 40$ the methods are quite comparable, but H and WH are a bit worse than the others. When we increase the dimension (Figures 36 to 38), even with the randomization, H, WH and VC are clearly worse than FL, AD, and KW.

For the last case ($K = 55$ and $s = 75$), we have extended the results up to $N = 10^6$, graphing the error and variance at each multiple of 5000. The results are shown on Figure 39.

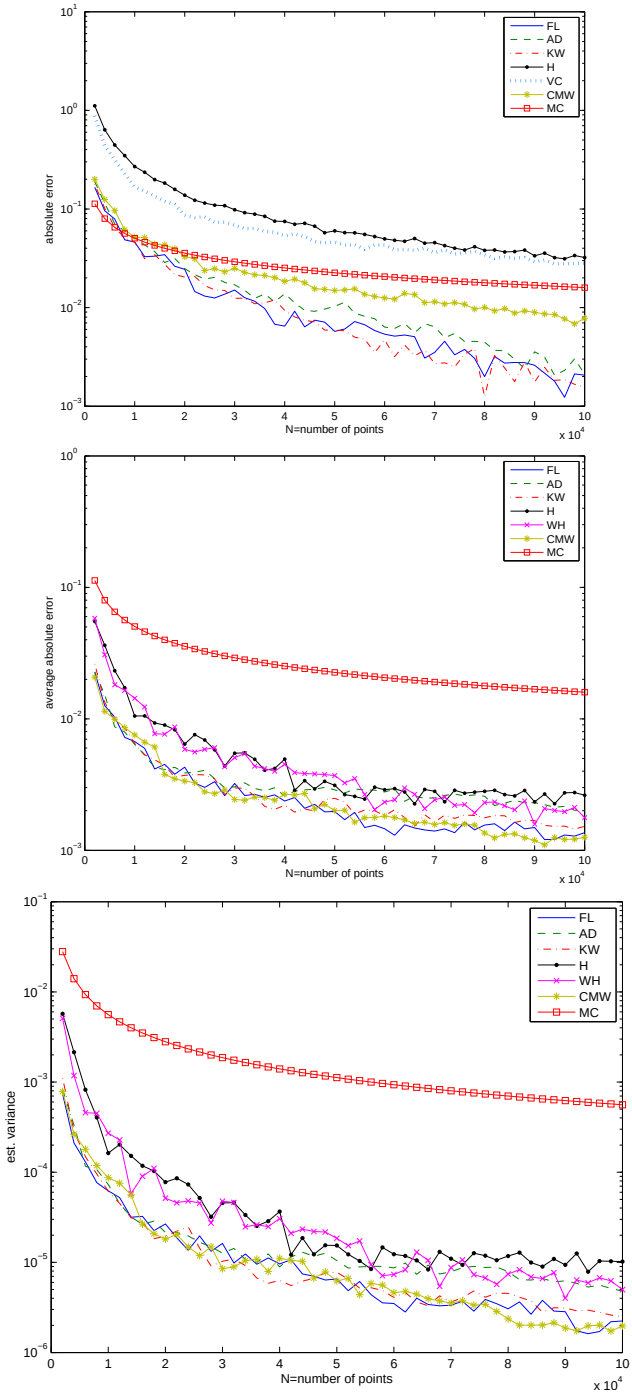

Fig. 32. Asian option with $K = 45$ and $s = 40$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

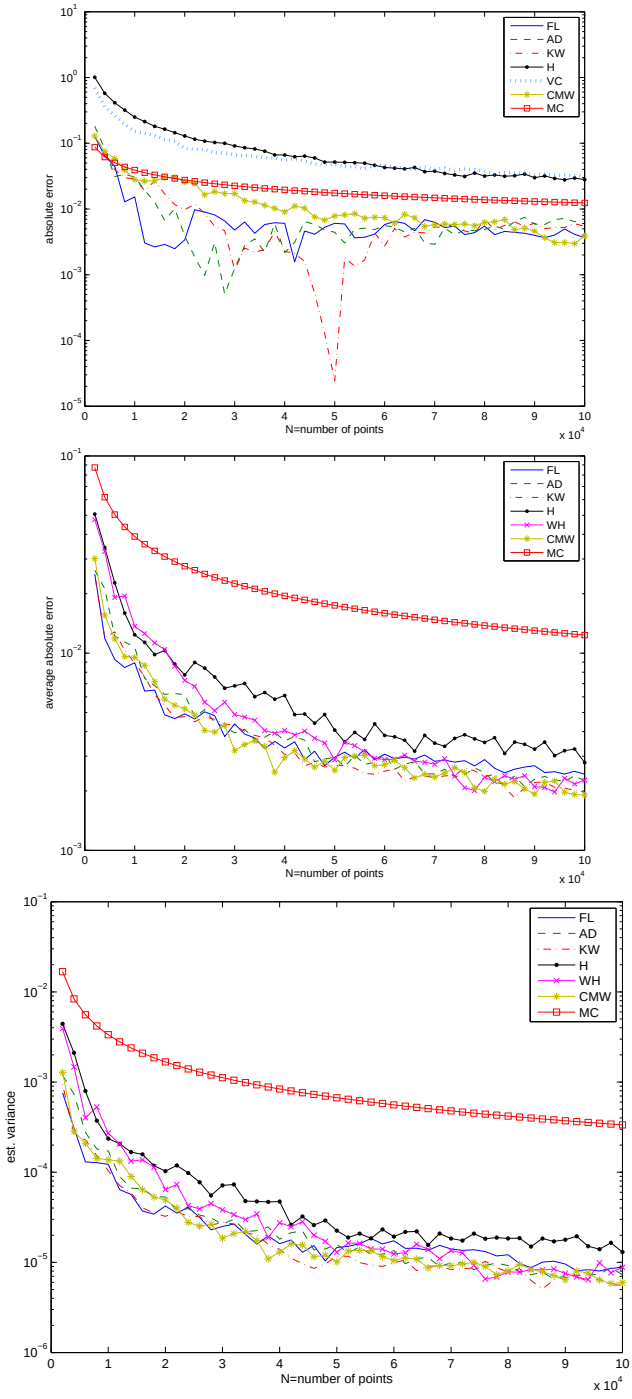

Fig. 33. Asian option with $K = 50$ and $s = 40$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

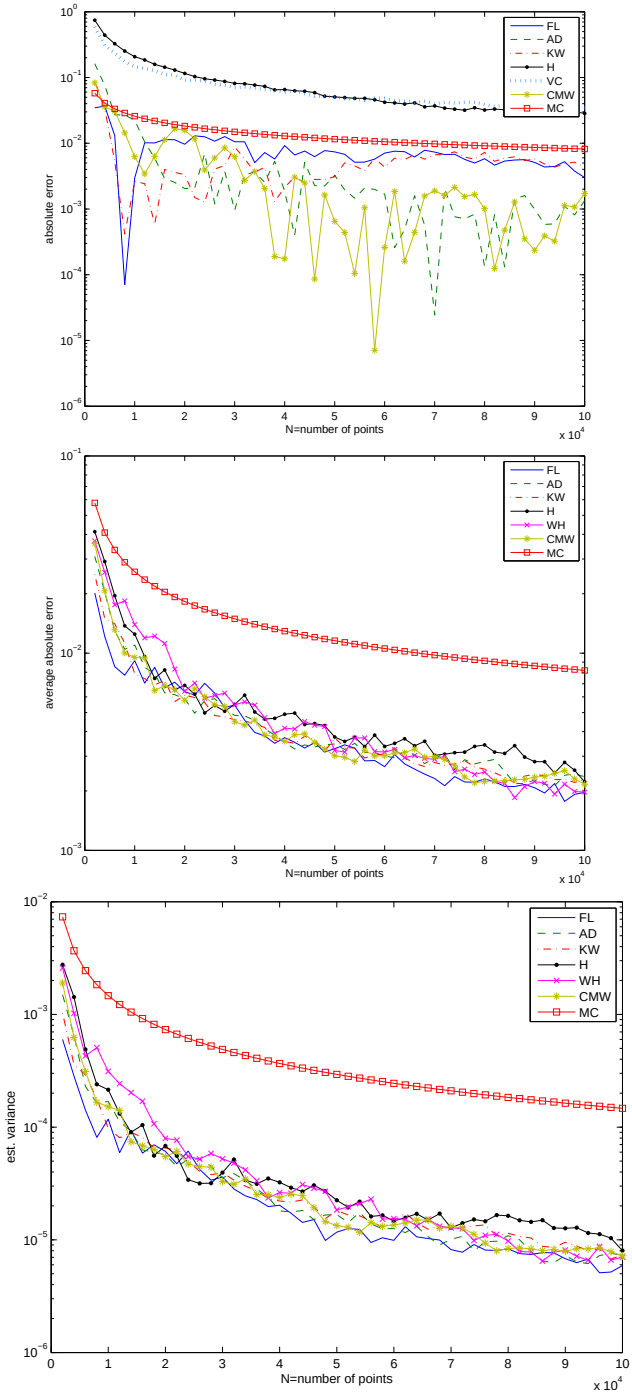

Fig. 34. Asian option with $K = 55$ and $s = 40$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

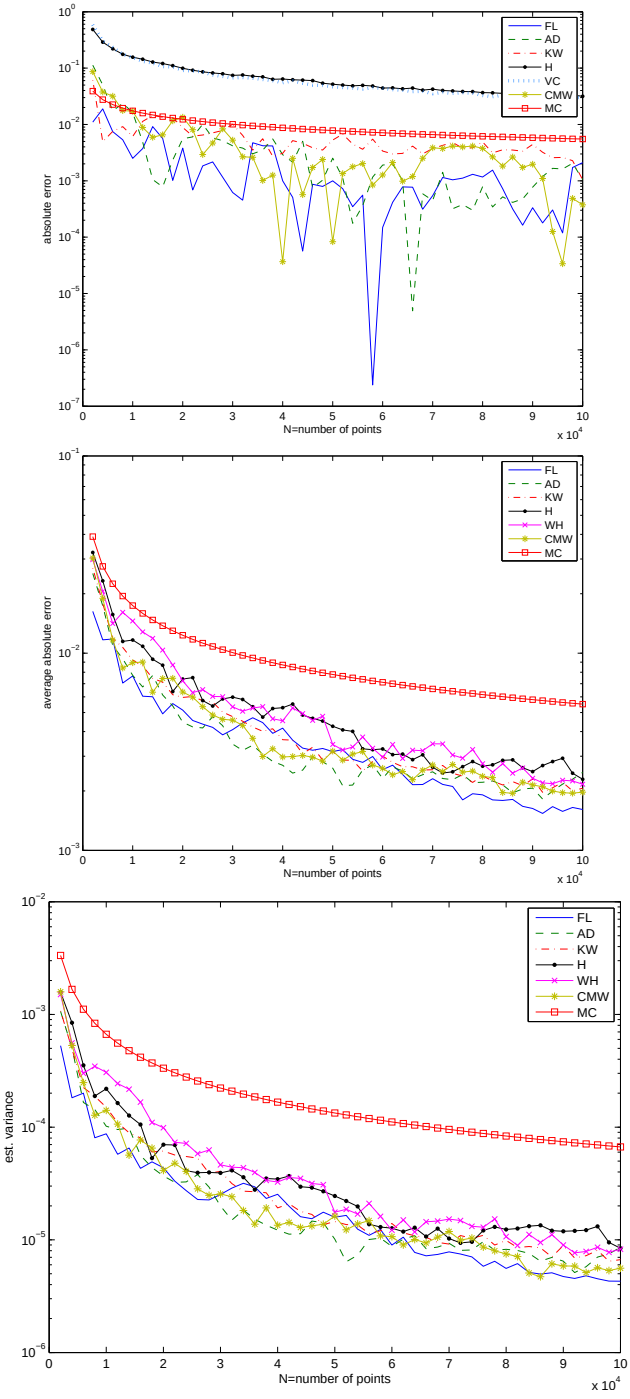

Fig. 35. Asian option with $K = 60$ and $s = 40$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

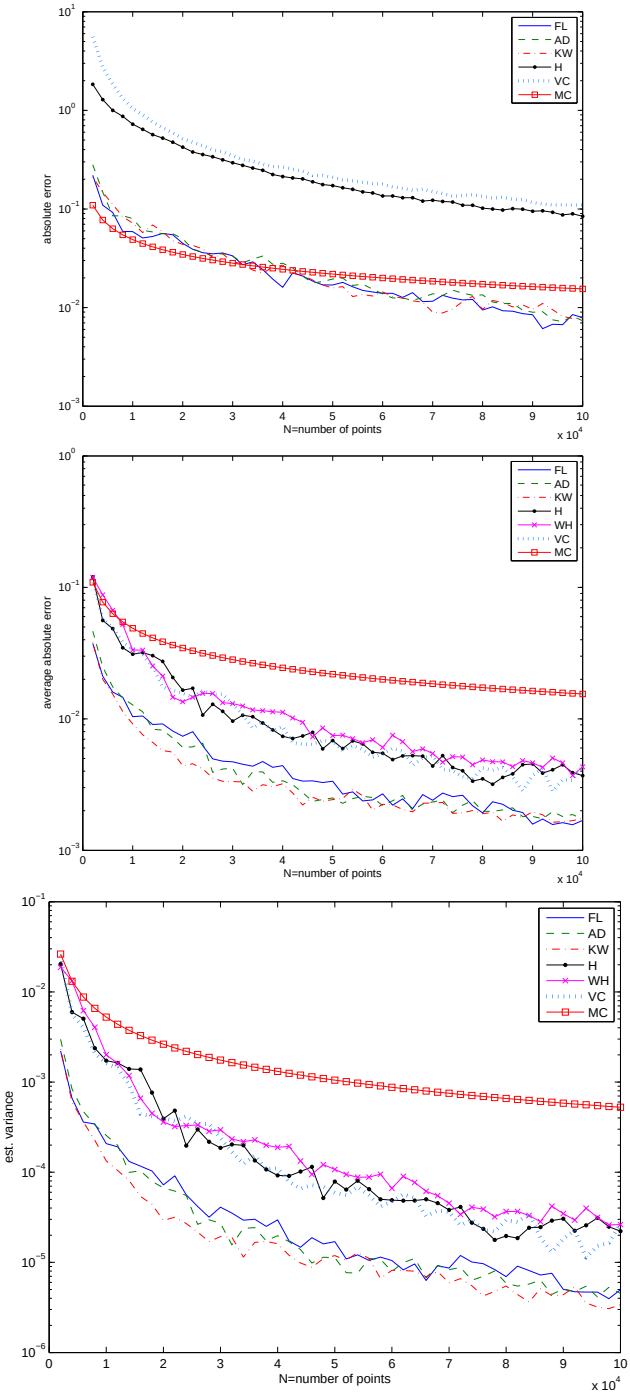

Fig. 36. Asian option with $K = 45$ and $s = 75$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

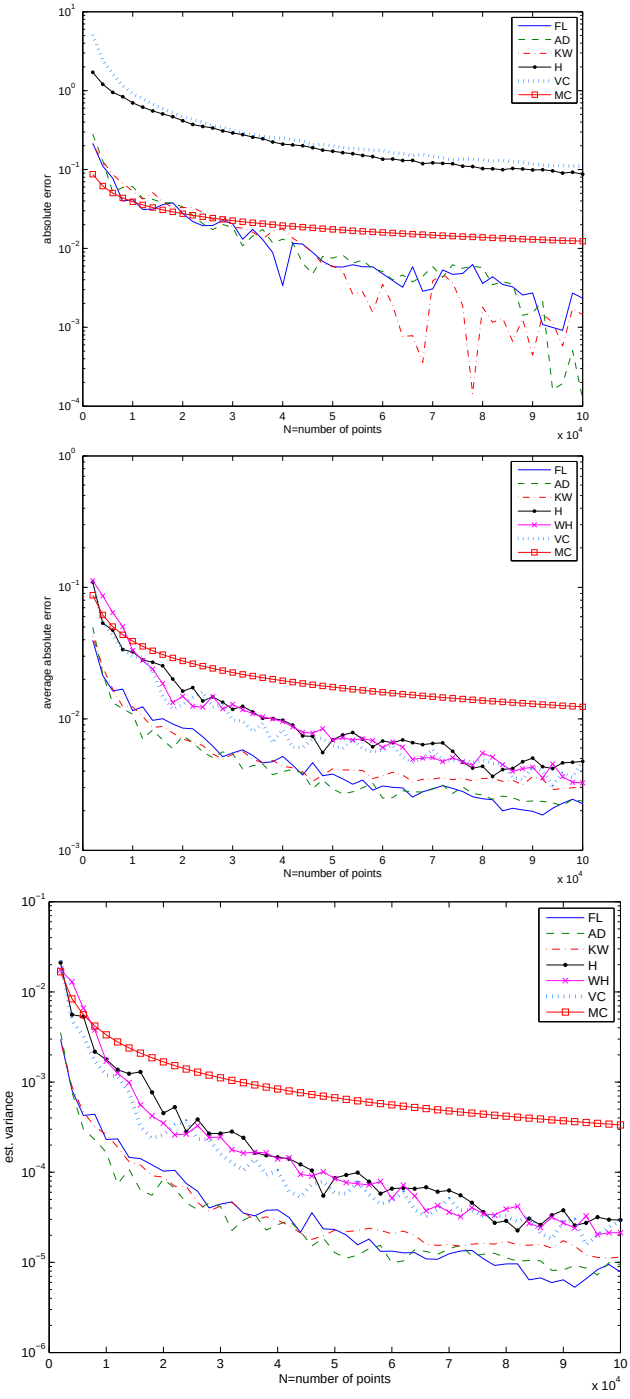

Fig. 37. Asian option with $K = 50$ and $s = 75$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

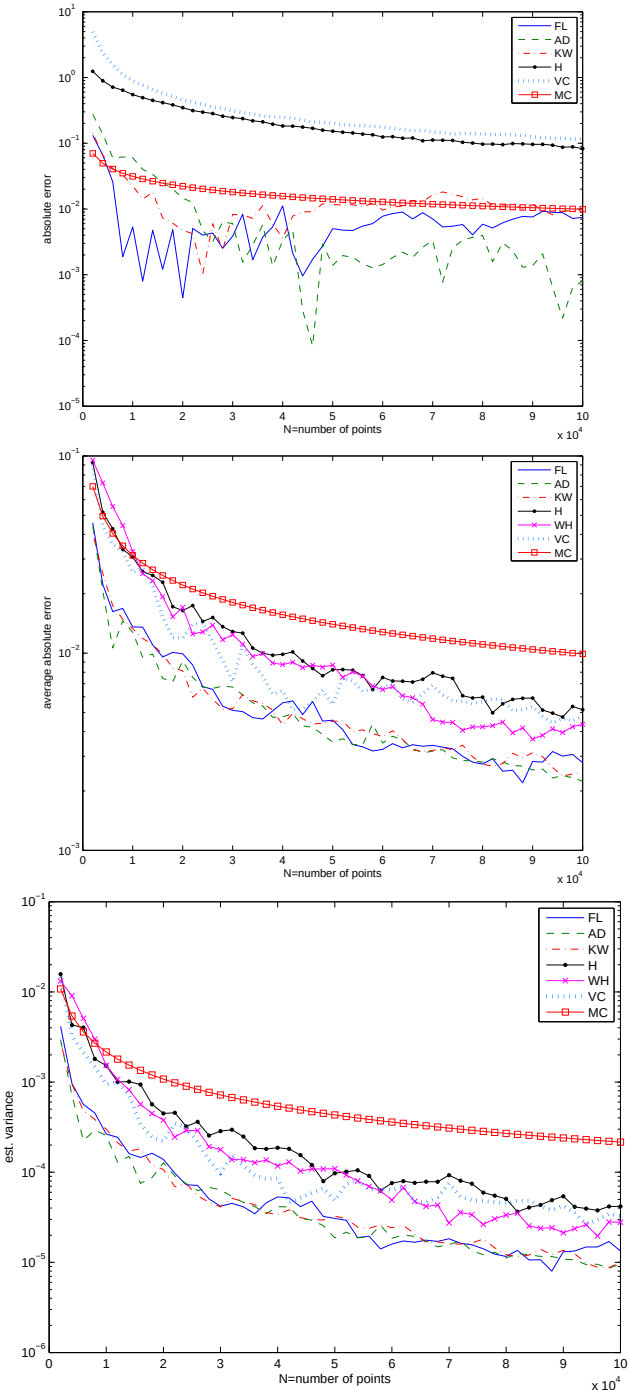

Fig. 38. Asian option with $K = 55$ and $s = 75$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

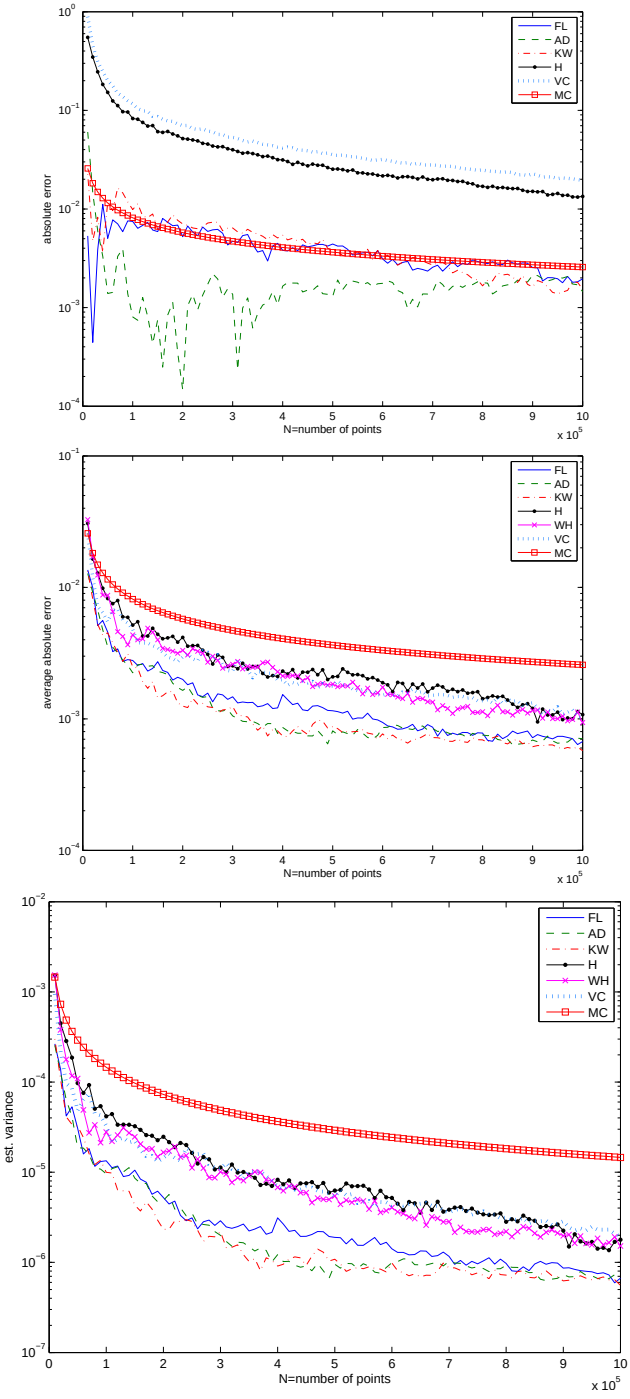

Fig. 39. Asian option with $K = 55$ and $s = 75$ for up to $N = 10^6$; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

G. EXPERIMENTS ON MORTGAGE-BACKED SECURITY

We start by describing the problem in details, following Caflisch et al. [1997]. The goal here is to estimate an expectation of the form

$$M_0 = \mathrm{E}\left(\sum_{j=1}^{s} v_j c_j\right),$$

which represents the time-0 value of this contract. Here v_j is the discount factor for month j, and c_j is the cash-flow for month j. Both of these quantities depend on the interest rate process in the following way: let i_l be the interest rate for month l. As in Caflisch et al. [1997], we use the interest-rate model

$$i_l = K_0 e^{\xi_l} i_{l-1},$$

where $\xi_l \sim N(0, \sigma^2)$. Then

$$v_l = \prod_{k=0}^{l-1} (1 + i_k)^{-1},$$

and

$$c_l = c r_l((1 - w_l) + w_l \alpha_l),$$

where

$$\begin{aligned}
c &= \text{monthly mortgage payment},\\
w_l &= \text{fraction of remaining mortgages prepaying in month } l,\\
&= K_1 + K_2 \arctan(K_3 i_l + K_4),\\
r_l &= \text{ fraction of remaining mortgages at month } l,\\
&= \prod_{k=1}^{l-1} (1 - w_l),\\
\alpha_l &= (\text{remaining annuity at month } l)/c,\\
&= \sum_{k=0}^{s-l} (1 + i_0)^{-k}.
\end{aligned}$$

Hence the problem is completely specified by the parameters (i_0, K_0, σ^2) for the interest rate model, and (K_1, K_2, K_3, K_4) for the prepayment model. As in [Caflisch et al. 1997], we choose $K_0 = \exp(-\sigma^2/2)$ so that $\mathrm{E}(i_k) = i_0$. Hence overall, we need to specify $(K_1, K_2, K_3, K_4, \sigma, i_0)$. In [Caflisch et al. 1997], two sets of parameters are chosen: the first one is given by

$$(K_1, K_2, K_3, K_4, \sigma, i_0) = (0.01, -0.005, 10, 0.5, 0.02, 0.007),$$

and is such that the 360-dimensional function $g(\cdot)$ satisfying

$$M_0 = \int_{[0,1)^{360}} g(x)dx,$$

where $x = (x_1, \ldots, x_s)$, is almost linear in its 360 inputs $x_1, \ldots, x_s$. The second choice

$$(K_1, K_2, K_3, K_4, \sigma, i_0) = (0.04, 0.0222, -1500, 7, 0.02, 0.007)$$

does not have such a strong linear component. Following [Caflisch et al. 1997], they are referred to as "almost linear" and "non-linear", respectively, in what follows. The other set of parameters that was tested is the one from Ninomiya and Tezuka [1996], given by

$$(K_1, K_2, K_3, K_4, \sigma, i_0) = (0.24, 0.134, -261.17, 12.72, 0.2, 0.00625).$$

For the mortgage-backed security, with the nearly linear set of parameters (Figure 40), FL and KW are the best, followed by AD, and then H, WH and VC. Just like for the Asian option, H and VC are much worse than MC when there is no randomization (top graph). Similar observations can be made for the non-linear case (Figure 41), except now AD is as good as FL and KW, and also all deterministic versions are slightly worse than MC, except for VC which becomes better than MC after $N = 50,000$. All methods do very well with the Nin-Tez set of parameters, as can be seen in Figure 42. As discussed in the paper, is probably due to the very low truncated effective dimension of this problem.

H. STUDYING THE ADMISSIBLE INTEGERS' POWER FOR HALTON-ATANASSOV

As mentioned in the paper, for the deterministic Halton-Atanassov sequence, although the theoretical support via the improved bounds on c_s holds for the version of that sequence where the powers of the admissible integers start at $r = 0$, starting the powers at $r = 1$ is much better in practice. Here, we show on Figure 43 (right) the 49th and 50th coordinates of the first 1000 points for the Halton-Atanassov sequence with powers starting at 0. It is practically identical to the plot for the Halton sequence, shown on the left-hand side.

From Figure 43, one can infer that this "power-0-AD" sequence is unlikely to perform better than Halton on integration problems where the number of points N is relatively small. To illustrate this, we chose to show the results obtained by this (deterministic) sequence on two problems where the AD sequence performed very well: the function g_2 with $c = 0.1$ and $s = 120$, and for the Asian call option with $K = 55$ and $s = 75$. Results are shown in Figures 44 and 45, respectively. On these two figures, for the Halton sequence we only show markers since otherwise, the (green) line for AD is not visible as it is superimposed onto the line for the Halton sequence. This suggests that although the "power-0-AD" sequence has better theoretical asymptotic properties than the Halton sequence, the number of points required for this advantage to show is much larger than the number of points N typically used in practical applications.

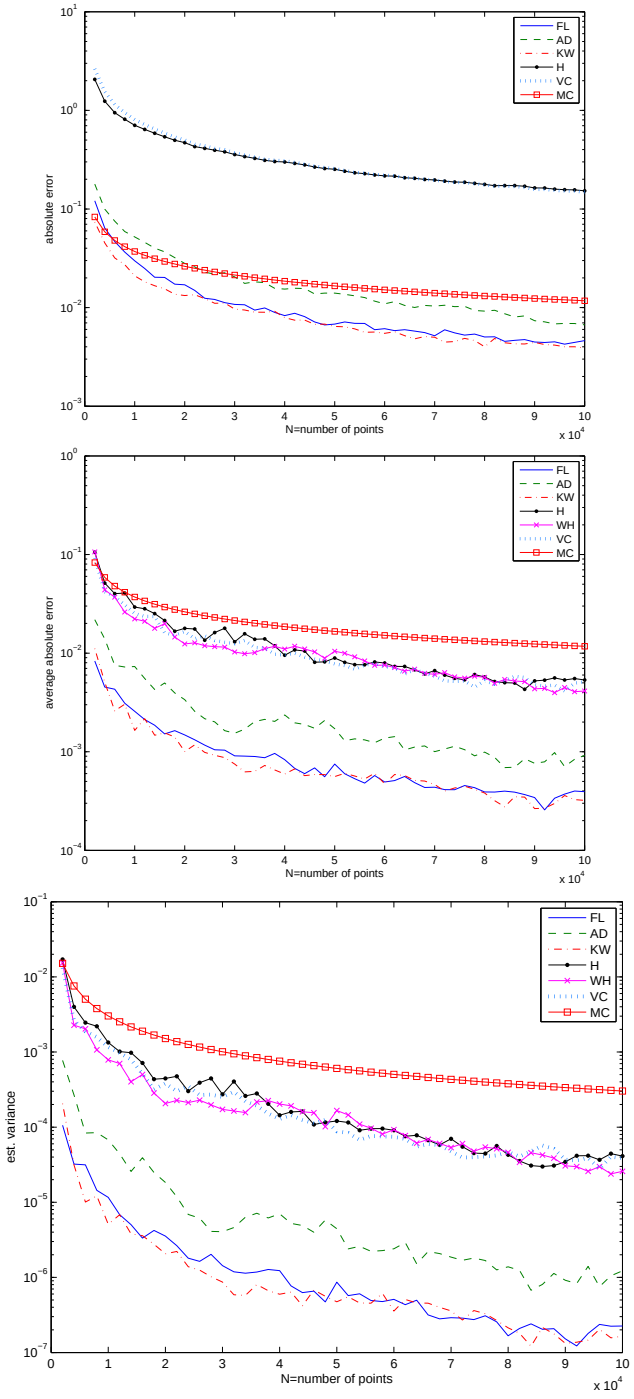

Fig. 40. Mortgage-backed security, nearly linear set of parameters; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

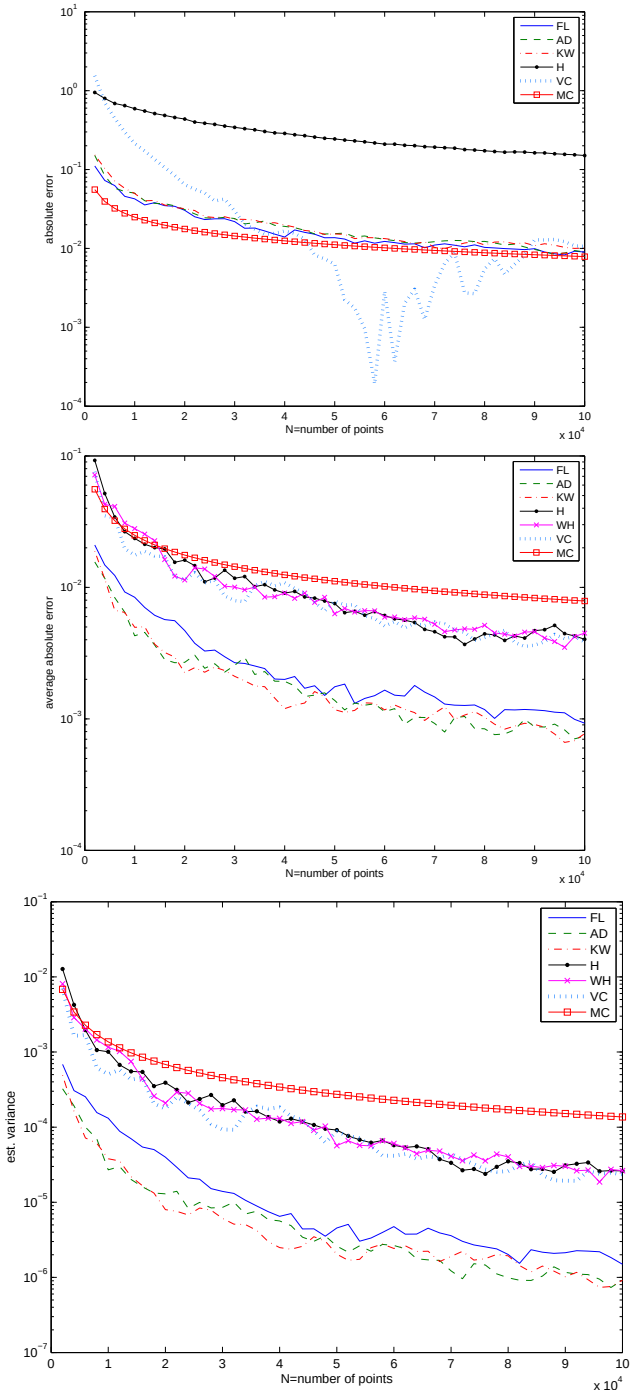

Fig. 41. Mortgage-backed security, non-linear set of parameters; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

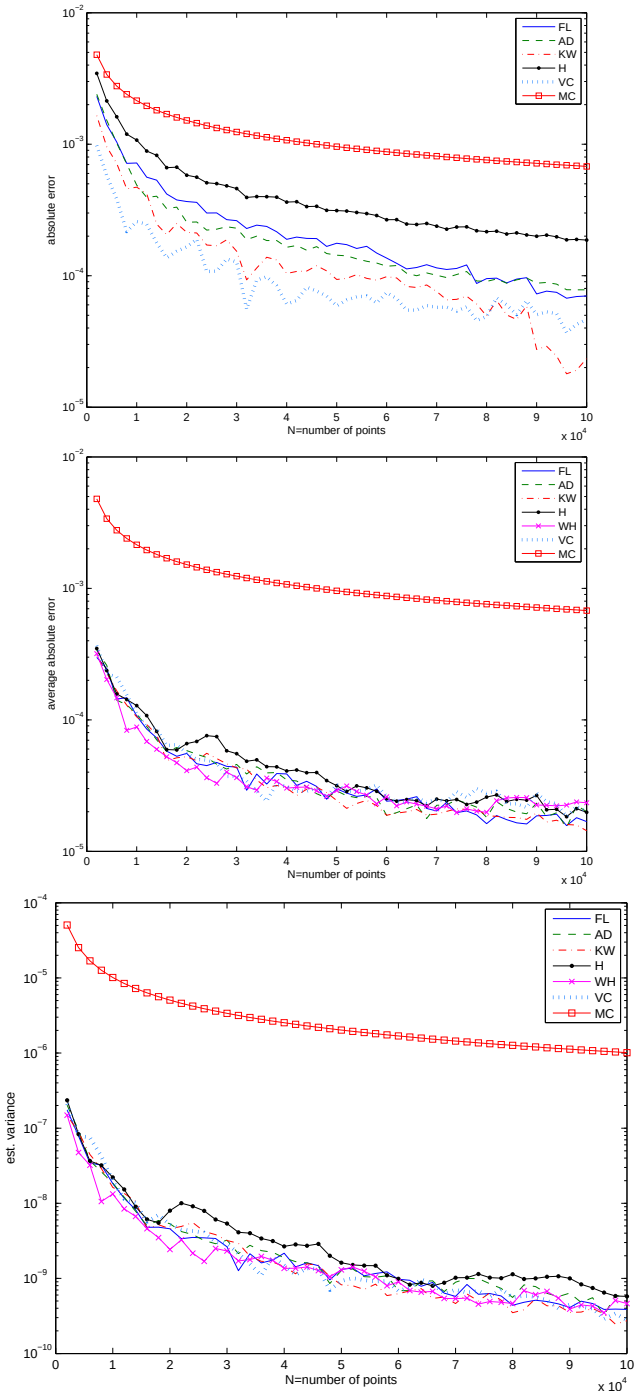

Fig. 42. Mortgage-backed security, set of parameters used by Ninomiya and Tezuka; deterministic error (top), average absolute error (middle), and estimated variance (bottom)

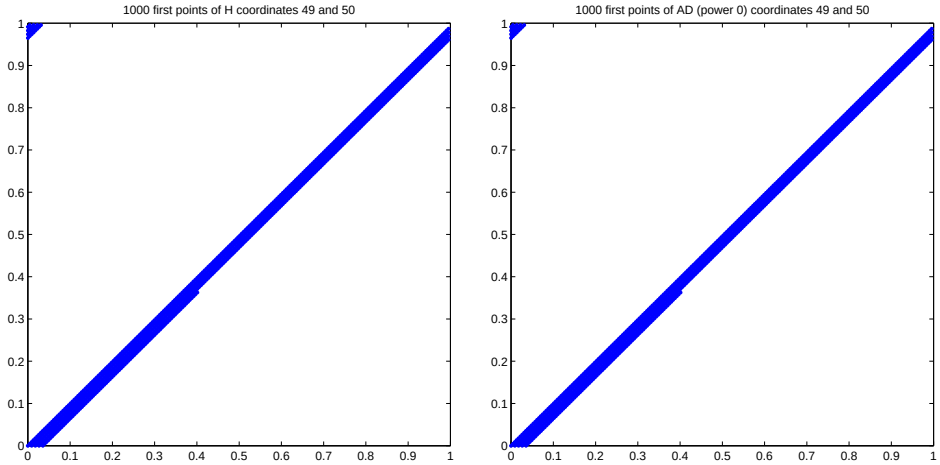

Fig. 43. 49th and 50th coordinates of the 1000 first points of Halton (left) and AD when the powers start at $r = 0$ (right)

Fig. 44. Performance of the power-0-AD sequence on the function g_2 with $c = 0.1$ and $s = 120$

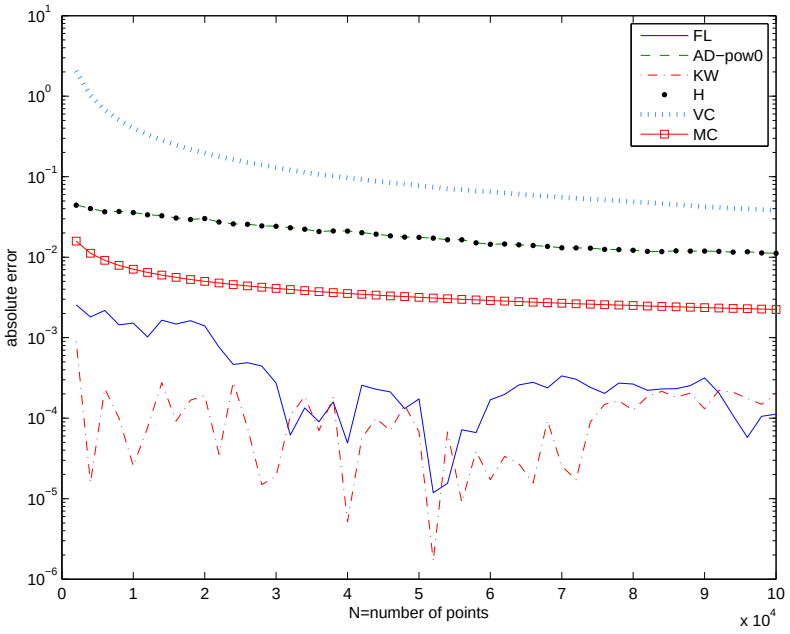

Fig. 45. Performance of power-0-AD sequence on the Asian option problem with $K = 55$ and $s = 75$