# Quasi-Monte Carlo methods

SE 702813 (Seminar zur Numerik und Stochastik)

Lukas Einkemmer

March 29, 2010

## 1 Introduction

The general problem we are interested in is to numerically compute the integral

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

where $f : [0,1]^d \to \mathbb{R}$ and $d \in \mathbb{N}$ is large (usually $d \geq 4$). As is shown in [3] classical methods of numerical integration are ill-suited for such problems. A number of related methods that alleviates this problem are known as Monte Carlo integration methods. Such methods can be shown to have a probabilistic error bound of order $\mathcal{O}\left(n^{-1/2}\right)$ (where $n$ is the number of function evaluations). In this term paper we will introduce other methods of approximating the before-mentioned integral, namely **quasi-Monte Carlo methods**, which are superficial similar to Monte Carlo methods but have certain advantages, especially for moderately sized $d$ (e.g. $d < 20$).

Monte Carlo methods use a sequence of (indepenent) random numbers to determine the points at which $f$ is evaluated. Random numbers, however, have, e.g., the disadvantage that they exhibit clustering behavior. Quasi-Monte Carlo methods, on the other hand, use a fully deterministic sequences that tries to "evenly covers" the unit cube. This rather vague notion of "evenly covers" will be made more precise in the following section by the so called discrepancy of a sequence.

The development in this term paper mainly follows [1] and [6].

## 2 Quasi-Monte Carlo Integration

First, let us precisely define what we mean by the discrepancy of a sequence with $n$ elements.

**Definition 2.1. (Discrepancy).** Suppose $E$ is the set of all rectangles in $[0,1]^d$ (a rectangle is assumed to be aligned with respect to the coordinate axes). Then, the **discrepancy of the sequence $\boldsymbol{x}_0, \dots, \boldsymbol{x}_{n-1}$** is defined by

$$D_n := \sup_{J \in E} \left| \frac{\#\{\boldsymbol{x}_i \in J \, : \, i \in \{0, \dots, n-1\}\}}{n} - \mathrm{vol} J \right|.$$

Surprisingly an alternative measure of discrepancy, the so called star discrepancy, will prove more useful.

**Definition 2.2. (Star discrepancy).** Suppose $E^*$ is the set of all rectangles $J$ in $[0,1]^d$ such that $\mathbf{0}$ is a vertex of the rectangle. Then, the **star discrepancy of the sequence $\boldsymbol{x}_0, \dots, \boldsymbol{x}_{n-1}$** is defined by

$$D_n^* := \sup_{J \in E^*} \left| \frac{\#\{\boldsymbol{x}_i \in J \, : \, i \in \{0, \dots, n-1\}\}}{n} - \mathrm{vol} J \right|.$$

In addition, we need the concept of the variation of a function (in the Hardy-Krause sense). It should be noted that this definition is not equivalent to the usual definition of variation. It differs in that we consider variation at the "upper" boundary as well.

**Definition 2.3. (Bounded variation).** Suppose $f : [0,1] \to \mathbb{R}$ is continuously differentiable. Then, the **variation of** $f$ is defined by

$$V(f) := \int_0^1 \left| \frac{df}{dx} \right| \mathrm{d}x.$$

Suppose $f : [0,1]^d \to \mathbb{R}$ is a differentiable function. Then, the **variation of** $f$ is defined by

$$V(f) := \int_{[0,1]^d} \left| \frac{\partial^d f(\boldsymbol{x})}{\partial \boldsymbol{x}_1 \dots \partial \boldsymbol{x}_d} \right| \mathrm{d}\boldsymbol{x}_1 \dots \mathrm{d}\boldsymbol{x}_d + \sum_{i=1}^{d} V(f|_{A_i}),$$

where $A_i = \{\boldsymbol{x} \in [0,1]^d \, : \, \boldsymbol{x}_i = 1\}$. We say a function $f$ has **bounded variation** if $V(f)$ exists and is finite.

For numerical integration purposes the following result is of central importance.

**Theorem 2.4.** ***(Koksma-Hlawka theorem).*** *Suppose $f : [0,1]^d \to \mathbb{R}$ is a function of bounded variation and $(\boldsymbol{x}_i)_{i=1}^{\infty}$. Then,*

$$|I - I_n| \leq V(f) D_n^*,$$

*where*

$$I_n := \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{x}_i).$$

*Proof.* Suppose $J \in E^*$ (i.e. the position of one vertex is $\mathbf{0}$). Then, $J$ is completely characterized by the antipodal vertex, denoted by $\boldsymbol{x}$, of the vertex $\mathbf{0}$. Therefore, we use the notation $J(\boldsymbol{x})$. Set

$$R(\boldsymbol{x}) := \int_{J(\boldsymbol{x})} \frac{1}{n} \sum_{i=1}^{n} \delta(\boldsymbol{y} - \boldsymbol{x}_i) - 1 \, d\boldsymbol{y} = \frac{\#\{\boldsymbol{x}_i \in J(\boldsymbol{x}) \, : \, i \in \{1, \dots, n\}\}}{n} - \mathrm{vol} J.$$

Clearly

$$\frac{\partial R(\boldsymbol{x})}{\partial \boldsymbol{x}_1 \dots \partial \boldsymbol{x}_d} = \frac{1}{n} \sum_{i=1}^{n} \delta(\boldsymbol{y} - \boldsymbol{x}_i) - 1$$

Therefore, the following computation gives us an estimate of the approximation error.

$$
\begin{aligned}
|I - I_n| &= \left| \int_{[0,1]^d} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{x}_i) \right| \\
&= \left| \int_{[0,1]^d} \left[ 1 - \frac{1}{n} \sum_{i=1}^{n} \delta(\boldsymbol{x} - \boldsymbol{x}_i) \right] f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right| \\
&= \left| \int_{[0,1]^d} \frac{\partial R(\boldsymbol{x})}{\partial \boldsymbol{x}_1 \dots \partial \boldsymbol{x}_d} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right| \\
&= \left| \sum_{\boldsymbol{\alpha} \in Z_2^d} (-1)^{|\boldsymbol{\alpha}|} \int_{[0,1]^{|\boldsymbol{\alpha}|}} R(\boldsymbol{x}) \frac{\partial^{|\boldsymbol{\alpha}|} f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} |_{\boldsymbol{x}^{\boldsymbol{\alpha}}=0}^{1} \, \mathrm{d}\boldsymbol{x}^{\boldsymbol{\alpha}} \right| \\
&= \left| \sum_{\boldsymbol{\alpha} \in Z_2^d} (-1)^{|\boldsymbol{\alpha}|} \int_{[0,1]^{|\boldsymbol{\alpha}|}} R(\boldsymbol{x}) \frac{\partial^{|\boldsymbol{\alpha}|} f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} |_{\boldsymbol{x}^{\boldsymbol{\alpha}}=1} \, \mathrm{d}\boldsymbol{x} \right| \\
&\leq \left( \sup_{\boldsymbol{x} \in [0,1]^d} R(\boldsymbol{x}) \right) \left| \sum_{\boldsymbol{\alpha} \in Z_2^d} (-1)^{|\boldsymbol{\alpha}|} \int_{[0,1]^{|\boldsymbol{\alpha}|}} \frac{\partial^{|\boldsymbol{\alpha}|} f|_{\boldsymbol{x}^{\boldsymbol{\alpha}}=1}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} \, \mathrm{d}\boldsymbol{x}^{\boldsymbol{\alpha}} \right| \\
&\leq \left( \sup_{\boldsymbol{x} \in [0,1]^d} R(\boldsymbol{x}) \right) \sum_{\boldsymbol{\alpha} \in Z_2^d} \int_{[0,1]^{|\boldsymbol{\alpha}|}} \left| \frac{\partial^{|\boldsymbol{\alpha}|} f|_{\boldsymbol{x}^{\boldsymbol{\alpha}}=1}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} \right| \, \mathrm{d}\boldsymbol{x}^{\boldsymbol{\alpha}} \\
&= \left( \sup_{\boldsymbol{x} \in [0,1]^d} R(\boldsymbol{x}) \right) V(f) \\
&= D_n^* V(f),
\end{aligned}
$$

where we used the fact that $R(\boldsymbol{x}) = 0$ if any component of $\boldsymbol{x}$ is 0, as well as integration by parts. Furthermore, we used the fact that the recursive definition given in Definition 2.3 can be written as

$$
V(f) = \sum_{\boldsymbol{\alpha} \in Z_2^d} \int_{[0,1]^{|\boldsymbol{\alpha}|}} \left| \frac{\partial^{|\boldsymbol{\alpha}|} f|_{\boldsymbol{x}^{\boldsymbol{\alpha}}=1}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} \right| \, \mathrm{d}\boldsymbol{x}^{\boldsymbol{\alpha}}.
$$

□

# 3 Low discrepancy sequence

In the previous section we derived an error bound (Theorem 2.4) for quasi-Monte Carlo integration. Since the variation of $f$ is fixed for a given function our task is to find a sequence that minimizes $D_n^*$. The first sequence we discuss, the Van der Corput sequence, is a somewhat artificial example in a single dimension. However, it forms the basis for a class of important sequences in multi-dimensional space.

**Theorem 3.1.** *Every $n \in \mathbb{N}_{\geq 0}$ has a unique digit expansion in a base $b \in \mathbb{N}_{\geq 2}$ of the form*

$$
n = \sum_{j=0}^{\infty} a_j(n) b^j,
$$

*where $a_j(n) \in \{0, 1, \dots, b-1\}$ for every $j \geq 0$ and $a_j(n) = 0$ for all sufficiently large $j$.*

*Proof.* We proceed by induction. First, 0 has obviously the expansion $a_j(n) = 0$, $\forall j \in \mathbb{N}_{\geq 0}$. Suppose we can expand $n - 1$ as

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

| Basis | $i=0$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 2 | 0 | 0.5 | 0.25 | 0.75 | 0.125 | 0.625 | 0.375 | 0.875 |
| 3 | 0 | $0.\dot{3}$ | $0.\dot{6}$ | $0.\dot{1}$ | $0.\dot{4}$ | $0.\dot{7}$ | $0.\dot{2}$ | $0.\dot{5}$ |
| 4 | 0 | 0.25 | 0.5 | 0.75 | 0.0625 | 0.3125 | 0.5625 | 0.8125 |

Table 3.1: Van der Corput sequence.

Then,

$$n = (1 + a_0(n-1)) + \sum_{j=1}^{\infty} a_j(n-1)b^j.$$

If $1 + a_0 < b - 1$ we are done. Otherwise, $1 + a_0 = b$ and we write

$$n = 0 + (1 + a_1(n-1)) + \sum_{j=2}^{\infty} a_j(n-1)b^j.$$

This procedure terminates after a finite number of steps, as desired. □

**Definition 3.2. (Radical-inverse function).**

$$\begin{aligned} \phi_b :& \mathbb{N}_{\geq 0} \to [0,1) \\ & n \mapsto \sum_{j=0}^{\infty} a_j(n) b^{-j-1} \end{aligned}$$

**Definition 3.3. (Van der Corput sequence).** Suppose $b \in \mathbb{N}_{\geq 2}$. Then, the sequence $(x_i)_{i=0}^{\infty}$ defined by

$$x_i = \phi_b(i)$$

for every $i \geq 0$ is called the **van der Corput Sequence in base** $b$.

**Example 3.4.** The first seven elements of the van der Corput sequence in base $b = 2, 3, 4$ are given in table 3.1.

The van der Corput sequence can be extended in an obvious way to a sequence in $[0,1)^d$.

**Definition 3.5. (Halton sequence).** Suppose $b_1, \ldots, b_d \in \mathbb{N}_{\geq 2}$. Then, the sequence $(\boldsymbol{x}_i)_{i=0}^{\infty}$ defined by

$$\boldsymbol{x}_i = \begin{bmatrix} \phi_{b_1}(i) \\ \vdots \\ \phi_{b_d}(i) \end{bmatrix}$$

is called the Halton sequence in the bases $b_1, \ldots, b_d$.

Obviously, for $d = 1$ the Halton sequence reduces to the van der Corput sequence (in base $b_1$).

**Example 3.6.** A comparison between random numbers and the Halton sequence in bases $2, 3$ is shown in figure 1 and figure 2.

**Theorem 3.7.** *If $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{n-1}$ are the first $n$ elements of the Halton sequence in the bases $b_1, \ldots, b_d$, where $b_1, \ldots, b_d$ are pairwise coprime. Then,*

$$D_n^* < \frac{d}{n} + \frac{1}{n} \prod_{i=1}^{d} \left( \frac{b_i - 1}{2 \log b_i} \log n + \frac{b_i + 1}{2} \right).$$

*Therefore,*

$$D_n^* \leq A(b_1, \ldots, b_d) \frac{\log^d n}{n} + \mathcal{O}\left( \frac{\log^{d-1} n}{n} \right)$$

*with*

$$A(b_1, \ldots, b_d) = \prod_{i=1}^{d} \frac{b_i - 1}{2 \log b_i}.$$

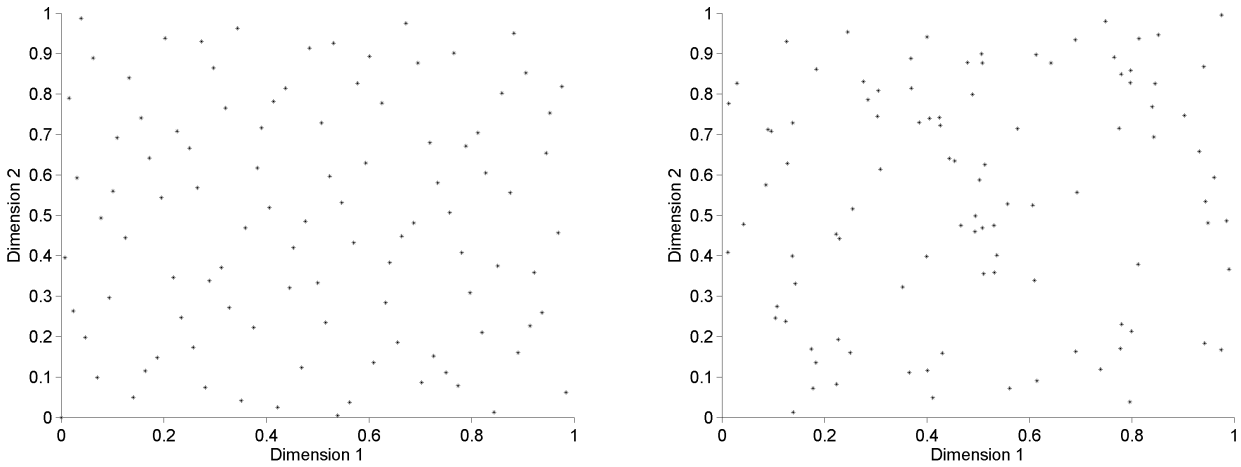


Figure 1: Halton sequence and random numbers (100 points each).

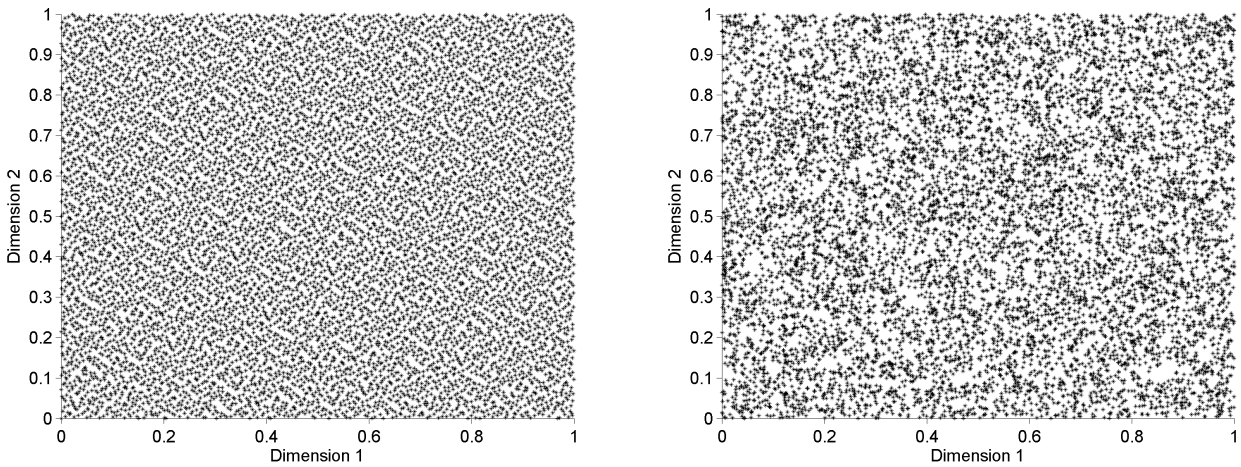


Figure 2: Halton sequence and random numbers (10000 points each).

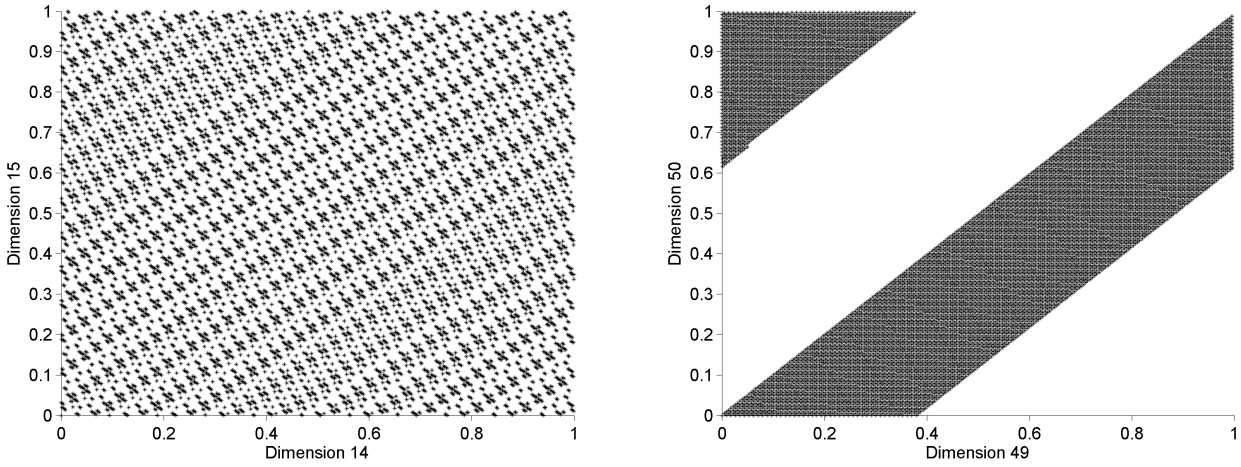


Figure 3: Halton sequence in 15 and 50 dimensions.

*Proof.* See e.g. [6, p. 29-31]. □

The question immediately arises in which way $b_1, \ldots, b_d$ has to be chosen such as to minimize $A(b_1, \ldots, b_d)$. Since $b_1, \ldots, b_d$ are pairwise coprime the obvious answer is given by Cololary 3.8.

**Corollary 3.8.** *If $b_1, \ldots, b_d$ is chosen such that $b_1 = p_1, \ldots, b_d = p_d$, where $p_i$ is the $i$-th prime number, then*

$$A(b_1, \ldots, b_d) = A_d = \prod_{i=1}^{d} \frac{p_i - 1}{2 \log p_i}$$

*is the (unique) global minimum under the constraints that $b_1, \ldots, b_d$ are pairwise coprime.*

*Proof.* Since

$$\frac{b_i - 1}{2 \log b_i}$$

is monotonically decreasing, we must choose $b_i$ as small as possible. Given these constraints the first $d$ prime numbers are the unique choice. □

It is widely believed that no sequence with better asymptotic properties (with respect to $n$) than the Halton sequence can be found (see e.g. [6, p. 32]). Therefore, we make the following definition.

**Definition 3.9. (Low-discrepancy sequence).** Suppose $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{n-1}$ are the first $n$ elements of a sequence $(\boldsymbol{x}_i)_{i=0}^{\infty}$. We call $(\boldsymbol{x}_i)_{i=0}^{\infty}$ a **low-discrepancy sequence** if

$$D_n^* \leq \mathcal{O}\left(\frac{\log^d n}{n}\right), \qquad \text{as } n \to \infty.$$

**Theorem 3.10.** *The Halton sequence for any base $b_1, \ldots, b_d$, where $b_1, \ldots, b_d$ are pairwise coprime, is a low-discrepancy sequence.*

*Proof.* Follows from 3.7. □

# 4 The need for additional low-discrepancy sequences

The Halton sequence discussed in the previous section has a flaw that renders it, almost, useless if the number of dimensions is large. This result can be anticipated from figure 3. The precise statement is given by Theorem 4.1.

**Theorem 4.1.** *Suppose $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{n-1}$ are the first $n$ elements of a Halton sequence. Then,*

$$\lim_{d\to\infty} \frac{\log A_d}{d \log d} = 1.$$

*Proof.* The prime number theorem states that (see e.g. [5])

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x} = 1,$$

where $\pi = \#\{p \leq x : p \text{ prime}\}$. Applying the logarithm we get

$$0 = \lim_{x\to\infty} \log\left(\frac{\pi(x)}{x/\log x}\right) = \lim_{x\to\infty} \log x \left(\frac{\log \pi(x)}{\log x} - 1 + \frac{\log\log x}{\log x}\right).$$

Therefore,

$$0 = \lim_{x\to\infty} \left(\frac{\log \pi(x)}{\log x} - 1 + \frac{\log\log x}{\log x}\right) = \lim_{x\to\infty} \left(\frac{\log \pi(x)}{\log x} - 1\right),$$

which implies

$$\lim_{x\to\infty} \frac{\log \pi(x)}{\log x} = 1.$$

Suppose $p_i$ is the $i$-th prime number. Then $\pi(p_i) = i$. Thus, for $x = p_i$ we get

$$1 = \lim_{x\to\infty} \frac{x}{\pi(x)\log x} = \lim_{x\to\infty} \frac{x}{\pi(x)\log \pi(x)} = \lim_{i\to\infty} \frac{p_i}{i \log i}.$$

This result can be applied to

$$\lim_{d\to\infty} \frac{\log A_d}{d\log d} = \lim_{d\to\infty} \frac{\sum_{i=1}^{d} \log \frac{p_i - 1}{2\log p_i}}{d\log d} = \lim_{d\to\infty} \frac{\sum_{i=1}^{d} [\log(i\log i - 1) - \log(2\log(i\log i))]}{d\log d} = \lim_{d\to\infty} \frac{\sum_{i=1}^{d} \log(i\log i - 1)}{d\log d},$$

since (for sufficiently large $d$)

$$\lim_{d\to\infty} \frac{\sum_{i=1}^{d} \log(2\log(i\log i))}{d\log d} \leq \lim_{d\to\infty} \frac{d\log(2\log(d\log d))}{d\log d} = \lim_{d\to\infty} \frac{\log(2\log(d\log d))}{\log d} = 0.$$

In addition,

$$\lim_{d\to\infty} \frac{\sum_{i=1}^{d} \log(i\log i - 1)}{d\log d} \leq \lim_{d\to\infty} \frac{d\log(d\log d)}{d\log d} \leq \lim_{d\to\infty} \frac{\log(d^{1+\epsilon})}{\log d} = 1 + \epsilon,$$

for every $\epsilon > 0$. Furthermore, since $\log(i\log i - 1)$ is monotonically increasing

$$\begin{aligned}\lim_{d\to\infty} \frac{\sum_{i=1}^{d} \log(i\log i - 1)}{d\log d} &\geq \lim_{d\to\infty} \frac{\sum_{i=1}^{d} \log(i^{1-\epsilon})}{d\log d} \geq \lim_{d\to\infty} \frac{\int_1^{d+1} (1-\epsilon)\log(x)\,\mathrm{d}x}{d\log d} \\ &= (1-\epsilon) \lim_{d\to\infty} \frac{d\log(d+1) + \log d - d}{d\log d} = 1 - \epsilon,\end{aligned}$$

for every $\epsilon > 0$ and sufficiently large $d$. Thus,

$$\lim_{d\to\infty} \frac{\log A_d}{d\log d} = 1,$$

as desired. □

The previous theorem implies that $A_d$ grows superexponentially. Therefore, for large $d$, the constant $A_d$ becomes extremely large and renders the Halton sequence unusable for most practical applications. The next section discusses a general theory of constructing low-discrepancy sequences that tries to overcome this difficulty.

# 5 General theory of low-discrepancy sequences

Now we turn to a property that defines a class of low-discrepancy sequences. Although rather technical at first sight, the following definitions can be understood by thinking of a sequence that covers the space at a given degree of fineness before moving to a higher degree.

**Definition 5.1.** An interval $E \subset [0,1]^d$ of the form

$$E = \prod_{i=1}^{d} [a_i b^{-e_i}, (a_i + 1) b^{-e_i})$$

with $a_i, e_i \in \mathbb{Z}$ and $e_i \geq 0$, $0 \leq a_i < b^{e_i}$ for $1 \leq i \leq d$ is called an **elementary interval in base** $b$.

**Definition 5.2.** Suppose $m, t \in \mathbb{Z}$ and $0 \leq t \leq m$. A $(t, m, d)$-**net** in base $b$ is a set $P$ with $\#(P) = b^m$ such that for every elementary interval of volume $b^{t-m}$ it holds that $\#(P \cap E) = b^t$.

**Definition 5.3.** Suppose $t \in \mathbb{Z}$, where $t \geq 0$. A sequence $(\boldsymbol{x}_i)_{i=0}^{\infty}$ is a $(t, d)$-**sequence in base** $b$ if, for all $k \in \mathbb{Z}, k \geq 0$ and $m > t$ the set

$$\{\boldsymbol{x}_i \,:\, kb^m \leq i \leq (k+1)b^m\}$$

is a $(t, m, d)$-**net in base** $b$.

**Theorem 5.4.** *Suppose $(\boldsymbol{x}_i)_{i=0}^{\infty}$ is a $(t, d)$-sequence in base $b$. Then,*

$$D_n^* \leq B_d b^t \frac{(\log N)^d}{N} + \mathcal{O}\left(\frac{b^t (\log N)^{d-1}}{N}\right),$$

*where*

$$B_d = \frac{1}{d}\left(\frac{b-1}{2\log b}\right)^d,$$

*if either $d = 2$, or $b = 2$ and $d = 3, 4$. Otherwise,*

$$B_d = \frac{1}{d!}\frac{b-1}{2\lfloor b/2 \rfloor}\left(\frac{\lfloor b/2 \rfloor}{\log b}\right)^d.$$

*Proof.* See e.g. [6, p. 49-60]. □

Various methods have been proposed to construct such sequences. The most prominent being Niederreiter and generalized Niederreiter sequences, which include many other sequences (e.g. Sobol sequences or Faure sequences) as a special case. From the estimate in Theorem 5.4 it is obvious that sequences with a small value of $t$ are sought. It can be shown, e.g., that Niederreiter sequences are optimal in this regard. However, a detailed treatment is beyond the scope of this term paper. The interested reader is referred to [6, Chap. 4.4-4.5] or [2].

# 6 Brownian bridge discretization

As already discussed, quasi-Monte Carlo methods offer an improvement over Monte Carlo methods only if the number if dimensions is moderate. We will now discuss a method to solve the heat equation (diffusion equation, imaginary time Schrödinger equation) by using quasi-Monte Carlo methods, even though the number of dimensions is large.

**Theorem 6.1.** ***(Feynman-Kac formula).*** *Suppose $u \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$ is bounded and satisfies*

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{1}{2}\Delta u + Vu, \\ u(0,x) &= f(x), \end{aligned}$$

*where $f : \mathbb{R}^d \to \mathbb{R}$ and $V : \mathbb{R}^d \to \mathbb{R}$. Then*

$$u(t,x) = \mathrm{E}\left(\int_0^t f(X_s) \mathrm{e}^{-\int_0^s V(X_\sigma)\,\mathrm{d}\sigma}\,\mathrm{d}s\right), \tag{1}$$

*where $X_t$ is a Brownian motion with $\mathrm{E}(X_t) = x$.*

*Proof.* See e.g. [4, p. 5-7]. □

The Feynman-Kac formula can be stated in a way that doesn't limit its application to simple heat equations with a diffusion term. However, for our development the theorem stated above is sufficient.

The first step is to approximate the integral in equation 1. The most basic method is to use a uniform grid.

**Definition 6.2. (Uniform grid discretization with $m$ nodes).**

$$\int_0^t f(X_s)\mathrm{e}^{-\int_0^s V(X_\sigma)\,\mathrm{d}\sigma}\,\mathrm{d}s \approx \frac{1}{m}\sum_{j=1}^{m}\left[f(Y_j)e^{-\frac{1}{j}\sum_{k=1}^{j}V(Y_k)}\right],$$

where $Y_0 = x$ and $Y_j = Y_{j-1} + \mathcal{N}(0, \sqrt{t/m})$.

To compute the expected value in equation 1 we need to evaluate a $m \cdot d$ dimensional integral. For a good approximation we choose $m$ as large as computational constraints permit. To this integral, we could easily apply Monte Carlo methods. However, due to the large number of dimensions usually needed in such an approximation quasi-Monte Carlo methods would lose most of their advantage. Therefore, we use the following approximation.

**Definition 6.3. (Brownian bridge discretization with $m$ nodes).**

$$\int_0^t f(X_s)\mathrm{e}^{-\int_0^s V(X_\sigma)\,\mathrm{d}\sigma}\,\mathrm{d}s \approx \frac{1}{m}\sum_{j=1}^{m}\left[f(Y_j)e^{-\frac{1}{j}\sum_{k=1}^{j}V(Y_k)}\right].$$

Suppose $m = 2^l$ for $l \in \mathbb{N}_{\geq 1}$. Given $Y_0 = x$ and $Y_m = x + \mathcal{N}(0, \sqrt{t})$ we use the following construction

$$\begin{aligned}
Y_{m/2} &= \frac{1}{2}Y_0 + \frac{1}{2}Y_m + \mathcal{N}(0, \sqrt{a \cdot m/2}),\\
Y_{m/4} &= \frac{1}{2}Y_0 + \frac{1}{2}Y_{m/2} + \mathcal{N}(0, \sqrt{a \cdot m/4}),\\
Y_{3m/4} &= \frac{1}{2}Y_{m/2} + \frac{1}{2}Y_m + \mathcal{N}(0, \sqrt{a \cdot m/4}).
\end{aligned}$$

Similar constructions are possible for $m \neq 2^l$ (See e.g. [1, p. 39]).

For Monte Carlo methods the above discretization doesn't change the error bound since the overall variance is still the same. However, in the Brownian bridge discretization the large time steps are filled in first, resulting in a number of dimensions with large variance and a number of dimensions with smaller variance. The basic idea is to use quasi-Monte Carlo integration on say the first 10 dimensions, which have large variance. In this case the error scales close to $\mathcal{O}(n^{-1})$. For the remaining dimensions Monte Carlo integration is employed. In this case the error scales as $\mathcal{O}(n^{-1/2})$.

# References


[1] Russel E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, pages 1–49, 1998. `http://dsec.pku.edu.cn/~tieli/notes/numer_anal/MCQMC_Caflisch.pdf`.

[2] Josef Dick and Harald Niederreiter. On the exact t-value of niederreiter and sobol' sequences. 2008.

[3] Lukas Einkemmer. Monte Carlo methods, University of Innsbruck (bachelor thesis), 2009.

[4] Nikola Kamburov. The feynman-kac formula. 2007.

[5] John Knopfmacher. Encyclopaedia of mathematics. 2002. `http://eom.springer.de/A/a130060.htm`.

[6] Harald Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods.* Society for Industrial Mathematics, 1992.