Non-uniform low-discrepancy sequence generation and integration of singular integrands

Jürgen Hartinger[1] and Reinhold Kainhofer[2]

[1] Graz University of Technology, Department of Mathematics, Steyrergasse 30, A-8010 Graz, Austria, hartinger@finanz.math.tugraz.at
[2] Vienna University of Technology, Department of Math. Methods in Economics, Wiedner Hauptstr. 8-10/105-1, A-1040 Wien, Austria, reinhold@kainhofer.com

Summary. In this article, we will first highlight a method proposed by Hlawka and Mück to generate low-discrepancy sequences with an arbitrary distribution H, and discuss its shortcomings. As an alternative, we propose an interpolated inversion method that is also shown to generate H-distributed low-discrepancy sequences, in an effort of order $\mathcal{O}(N \log N)$.

Finally, we will address the issue of integrating functions with a singularity on the boundaries. Sobol and Owen proved convergence theorems and orders for the uniform distribution, which we will extend to general distributions. Convergence orders will be proved under certain origin- or corner-avoidance conditions, as well as growth conditions on the integrand and the density. Our results prove that also non-uniform quasi-Monte Carlo methods can be well applied to integrands with a polynomial singularity at the integration boundaries.

1 Introduction

The numerical solution of several problems arising in financial mathematics require the use of non-uniformly distributed point sequences. In many cases, (pseudo-) random sequences for a given density are generated by some kind of transformation, possibly involving two or more independent random variables. As quasi-Monte Carlo sequence follow a given construction scheme, subsequent elements of the sequence do not satisfy the requirement of independence. A good overview over several other ways to generate non-uniformly distributed (pseudo-) random sequences can be found in the Devroye's monograph [2]. Unfortunately, almost none of them can be applied to QMC.

In 1972, Hlawka and Mück [8] proposed a method to generate H-distributed sequences with low discrepancy by using the (quasi)-empirical distribution function instead. Later, they also extended the method to multi-dimensional sequences [9].

The case of dependent random variates is more involved, and hardly anything about this case is known for quasi-Monte Carlo methods. For special distributions with given marginals and covariance matrix generation methods are known, like the NORTA and QUARTA methods [6]. Again, in this case, the inverse of the marginals needs to be known to apply the transformation, so this method is also not applicable in most cases.

In this article, we will first investigate the Hlawka-Mück method and highlight its shortcomings. We will then propose some adaptions to make the generated sequences more suitable in many practical cases and investigate sequences generated by an approximated inversion of the cumulative distribution function.

In the second part we will point our view to the non-uniform integration of singular integrands. QMC integration of functions with a singularity at the integration boundaries were already investigated by Sobol' [17], and later as non-uniform integration problems by Hartinger, Kainhofer, and Tichy [5]. Both publications give criteria for convergence of the singular integral, but do not explicitly prove error orders. Owen [15] proved these for uniform integration using certain growth conditions on the function near the singularity. In this paper we will expand Owen's results to integration with respect to arbitrary densities.

2 Basic Definitions

Remark 1. Although all results in the article will be formulated on the unit cube $U^s = [0,1]^s$, they are valid on any compact subinterval $[\mathbf{a}, \mathbf{b}] \subset \mathbb{R}^s$ by a simple affine transformation of the sequence and all corresponding entities.

2.1 Discrepancy and Koksma-Hlawka Inequality

When dealing with quasi-Monte Carlo sequences, the most common measure of their distribution properties is the discrepancy. For uniformly distributed sequences on U^s it measures the maximum error one can obtain on intervals parallel to the axes:

Definition 1 (uniform discrepancy). *The discrepancy $D_N(\omega)$ of a sequence $\omega = (\mathbf{x}_1, \mathbf{x}_2, \ldots)$ is defined as*

$$D_N(\omega) = \sup_{J \subseteq U^s} \left| \frac{1}{N} A_N(J, \omega) - \lambda(J) \right|,$$

where A_N counts the number of elements of $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ falling into the interval J, i.e. $A_N(J, \omega) = \sum_{n=1}^{N} \chi_J(\mathbf{x}_n)$, and λ denotes the Borel-measure of the interval J.

The best sequences known to date (e.g. Halton, Sobol, Faure sequences, and (t, s)-nets) have a discrepancy order of $\mathcal{O}\left(\log^s N/N\right)$, which is also conjectured to be optimal.

The notion of discrepancy is especially important in view of the famous Koksma-Hlawka inequality, which allows to bound the quasi-Monte Carlo integration error by the variation of f multiplied by the discrepancy of the sequence ω. A good discussion of variation can be found in [16], and a detailled overview on discrepancy and low-discrepancy sequences is given in the monographs by Niederreiter [14] and Drmota and Tichy [3].

A similar concept of discrepancy can be defined for non-uniformly distributed sequences, i.e. sequences with density h or distribution function H:

Definition 2 (non-uniform discrepancy). *The H-discrepancy of the sequence $\tilde{\omega} = (\mathbf{y}_1, \mathbf{y}_2, \ldots)$ measures its distribution properties with respect to the measure H on U^s. It is defined as*

$$D_{N,H}(\tilde{\omega}) = \sup_{J \subseteq U^s} \left| \frac{1}{N} A_N(J, \tilde{\omega}) - \mathrm{H}(J) \right|.$$

Theorem 1 (non-uniform Koksma-Hlawka Inequality, [1]). *Let f be a function of bounded variation on U^s, H a probability distribution with continuous density on U^s and $\tilde{\omega} = (\mathbf{y}_1, \mathbf{y}_2, \ldots)$ a sequence on U^s. Then the QMC integration error can be bounded by*

$$\left| \int_{U^s} f(\mathbf{x}) \mathrm{dH}(\mathbf{x}) - \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{y}_n) \right| \leq V(f) D_{N,H}(\tilde{\omega}). \tag{1}$$

2.2 Existing Methods for the Generation of Non-Uniform Sequences, and Their Problems

The Koksma-Hlawka inequality gives a convergence criterion for QMC integration, and shows that asymptotically QMC methods have to be preferred over Monte Carlo methods, due to their error order of $\mathcal{O}\left(\log^s N/N\right)$ compared to $1/\sqrt{N}$ for Monte Carlo integration.

For generating non-uniformly distributed low-discrepancy sequences, most desirable would be the direct transformation of uniformly distributed sequences to H-distributed sequences using the inverse of the distribution function H^{-1} (the conditional distribution functions or the marginal distributions for multi-dimensional sequences). Such a transformation preserves the discrepancy in one dimension, i.e.

$$D_N(\omega) = D_{N,H}(H^{-1}(\omega)),$$

and is independent of the value of N, so that it can be used to generate an arbitrary number of points. In most cases, however, the distribution function is not explicitly available, so this method is no easily applicable. In the

multi-dimensional case with dependencies between the dimensions, even the discrepancy is not preserved, as intervals are not transformed to intervals.

The most common practice for pseudo-random variates, the acceptance-rejection method, also fails for quasi-Monte Carlo sequences. The main reason is that the rejection effectively introduces discontinuities into the integrand, which leads to bad results of QMC methods, as several numerical investigations show (e.g. [12, 18]). Even the Koksma-Hlawka inequality fails to provide error bounds due to the unbounded variation of such functions. An additional problem is the high number of discarded points, so the cost of generating a sequence of N points is at least one order of magnitude higher.

To tackle the problem of integration with other densities, Wang proposed a smoothed rejection sampling method [18] by adapting the integrand. Thus, while his method avoids the jumps in the integrand, it cannot be used to directly generate H-distributed sequences. The same shortcoming appears with the approach of stratified sampling, where the integration domain is split into various areas, and in each area n_i uniformly distributed points are generated. While this methods works well for integration, the sequences used therein do not display very good distribution properties.

3 The Hlawka-Mück Method

The idea behind the Hlawka-Mück transformation [8, 9] is to use an approximation of the distribution function in the inversion method. Instead of directly taking this value, they again use the original sequence to count the relative number of elements below that value and use this number as the new point. This way, the quasi-Monte Carlo error will not only involve the discrepancy of the original sequence, but also the quality of the approximation. They prove a bound on the discrepancy of $D_{N,H}(\tilde{\omega}) \leq (1+4M)^s D_N(\omega)$, where M denotes the supremum of the density.

In 1996, Hlawka [7] gave a modification of the Hlawka-Mück method using the one-dimensional marginal distributions instead of the conditional distributions for the transformation.

Definition 3. *Let $h(\mathbf{x})$ be a density function on $[0,1]^s$. For a point $\mathbf{x} = \left(x^{(1)}, \ldots, x^{(s)}\right) \in [0,1]^s$ we define the marginal distribution functions as*

$$\begin{aligned}
H_1(x^{(1)}) &= \int_0^{x^{(1)}} \int_0^1 \cdots \int_0^1 h(\mathbf{u}) \mathrm{d}\mathbf{u} \\
H_2(x^{(2)}) &= \int_0^1 \int_0^{x^{(2)}} \cdots \int_0^1 h(\mathbf{u}) \mathrm{d}\mathbf{u} \\
&\vdots \\
H_s(x^{(s)}) &= \int_0^1 \int_0^1 \cdots \int_0^{x^{(s)}} h(\mathbf{u}) \mathrm{d}\mathbf{u}\,.
\end{aligned}$$

As each of the functions H_i is invertible, Hlawka defines a transformation and bounds the discrepancy of the transformed sequence as follows:

Lemma 1 (Hlawka [7]). *Let $H(\mathbf{x})$ denote a cumulative distribution function with density $h(\mathbf{x}) = h_1(x^{(1)})h_2(x^{(2)})\cdots h_s(x^{(s)})$ defined on U^s and $M_h = \sup h(\mathbf{x})$. Let furthermore $\omega = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ be a sequence in U^s with discrepancy $D_N(\omega)$. Then the point set $\tilde{\omega} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$ with*

$$y_k^{(j)} = \frac{1}{N}\sum_{r=1}^{N}\left\lfloor 1 + x_k^{(j)} - H_j\left(x_r^{(j)}\right)\right\rfloor = \frac{1}{N}\sum_{r=1}^{N}\chi_{[0,x_k^{(j)}]}\left(H_j\left(x_r^{(j)}\right)\right) \tag{2}$$

has an H-discrepancy of

$$D_{N,H}(\tilde{\omega}) \leq 2(1+3M_h)^s D_N(\omega)\ .$$

The advantage of this approach is that the approximation quality of the distribution function increases with the number of sampled points N, so that the resulting H-distributed sequence is again a low-discrepancy sequence $\tilde{\omega}$, at least for independent marginals. For dependent marginals, in particular if the distribution function does not factor, the discrepancy can only be proved to satisfy the inequality $D_{N,H}(\tilde{\omega}) \leq c(D_N(\omega))^{1/s}$.

Observe, however, that in these integration problems at least in principle one can always avoid dependent sequences by hiding the dependence in the integrand through an incorporation of an appropriate copula (see [13] for an introduction into copulas).

Applying the Hlawka-Mück method to singular integrands, we found [4, 5] that these low-discrepancy sequences also work well with singular integrands, but only with an additional shift of all components with $y_k^{(j)} < 1/N$ to a value of $1/N$. By this shift the order of the discrepancy is preserved, and the resulting sequence is a low-discrepancy sequence with density h.

However, the Hlawka-Mück method also has several disadvantages:

1. The resulting sequence is generated only on a grid with spacing $\frac{1}{N}$. While the resulting sequence displays the required distribution properties, for several applications finer-grained sequences are of desire.
2. Several points might have identical coordinates, in particular for highly peaked distributions. Consequently, the minimum distance principle, which is desired in several applications in computer graphics (see e.g. [11]), is no longer fulfilled.
3. The construction of each point involves a sum over all other points, so the cost is $\mathcal{O}(N^2)$, and the (numerically expensive) distribution function has to be evaluated sN times.
4. One has to fix the number N beforehand, and the resulting set will heavily depend on it. This also means that when adding some points to the sequence, all other elements have to be regenerated.

In the sequel we will present several ways to solve or at least considerably improve these problems for most practical uses.

4 Interpolation

For some applications the Hlawka-Mück methods have the drawback that all points of a set with cardinality N lie on the lattice $\{\mathbf{x} \in U^s \big| x^{(l)} = i/N \text{ for } 0 \leq i \leq N, 1 \leq l \leq s\}$. Rather than using the non-continuous quasi-empirical distribution, we therefore propose to use a smoothed approximation, where the values between the jumps are interpolated in the empirical distribution function. The idea is to avoid the lattice structure and improve the approximation of the inverse distribution function.

Theorem 2. *Let $\omega_N = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ be a sequence in U^s with discrepancy $D_N(\omega_N)$, and $H(\mathbf{x})$ a distribution function with bounded, continuous density $h(\mathbf{x}) = \prod_{i=1}^{s} h_i(x^{(i)})$ and $h_i(x^{(i)}) \leq M < \infty$ for all i. Furthermore, let $H_i(x) = \int_0^x h_i(u)\mathrm{d}u$ and define for $k = 1, \ldots, N$ and $l = 1, \ldots, s$ the values*

$$x_k^{(l)-} = \max_{\mathcal{A}=\left\{\mathbf{x}_i \in \omega_N \;\middle|\; H_l(x_i^{(l)}) \leq x_k^{(l)}\right\}} x_i^{(l)}, \qquad \text{and} \qquad x_k^{(l)-} = 0 \text{ for } \mathcal{A} = \emptyset,$$

$$x_k^{(l)+} = \min_{\mathcal{B}=\left\{\mathbf{x}_i \in \omega_N \;\middle|\; H_l(x_i^{(l)}) \geq x_k^{(l)}\right\}} x_i^{(l)}, \qquad \text{and} \qquad x_k^{(l)+} = 1 \text{ for } \mathcal{B} = \emptyset.$$

Then the discrepancy of the set $\bar{\omega}_N = (\mathbf{y}_k)_{1 \leq k \leq N}$ generated by

$$y_k^{(l)} = \frac{H_l\left(x_k^{(l)+}\right) - x_k^{(l)}}{H_l\left(x_k^{(l)+}\right) - H_l\left(x_k^{(l)-}\right)} x_k^{(l)-} + \frac{x_k^{(l)} - H_l\left(x_k^{(l)-}\right)}{H_l\left(x_k^{(l)+}\right) - H_l\left(x_k^{(l)-}\right)} x_k^{(l)+} \tag{3}$$

can be bounded by

$$D_{N,H}(\bar{\omega}) \leq (1 + 2M)^s D_N(\omega).$$

For the proof, we need to recall a lemma from the original paper of Hlawka and Mück [9]:

Lemma 2. *Let $\omega_1 = (\mathbf{u}_1, \ldots, \mathbf{u}_N)$ and $\omega_2 = (\mathbf{v}_1, \ldots, \mathbf{v}_N)$ be two sequences in U^s. If for all $1 \leq j \leq s$ and all $1 \leq i \leq N$ the condition*

$$|u_i^{(j)} - v_i^{(j)}| \leq \varepsilon_j$$

holds for some values ε_j, we get the following bound on the difference of the discrepancies

$$|D_N(\omega_1) - D_N(\omega_2)| \leq \prod_{j=1}^{s}(1 + 2\varepsilon_j) - 1. \tag{4}$$

Proof (Proof of Theorem 2). We first start with the one-dimensional case.

The set $\bar{\omega}_N = \{y_1, \ldots, y_N\}$ is H-distributed, so that the set $H(\bar{\omega}_N) = \{H(y_1), \ldots, H(y_N)\}$ is uniformly distributed and their respective discrepancies are equal.

We want to apply Lemma 2 with $\omega_1 = \omega_N$ and $\omega_2 = H(\bar{\omega}_N)$, so for $1 \leq k \leq N$ we obtain

$$|H(y_k) - x_k| = \left|H(y_k) - H(H^{-1}(x_k))\right| = \left|\int_{H^{-1}(x_k)}^{y_k} h(t)\mathrm{d}t\right| \leq M\left|y_k - H^{-1}(x_k)\right| \leq MD_N(\omega_N). \quad (5)$$

The last inequality can be proved as follows: By the definition of x_k^- and x_k^+ we have $H(x_k^-) \leq x_k \leq H(x_k^+)$, and by the monotonicity of H (assuming it is continuous, otherwise similar arguments can be used) we get

$$x_k^- \leq H^{-1}(x_k) \leq x_k^+.$$

Furthermore, y_k as constructed in (3) is just a linear interpolation between x_k^- and x_k^+, so we have the same bounds: $x_k^- \leq y_k \leq x_k^+$.

Subtracting these two, we get the estimate

$$\left|H^{-1}(x_k) - y_k\right| \leq \left|x_k^+ - x_k^-\right| \leq \max_{1\leq j\leq N} \min_{i\neq j} |x_i - x_j| \leq D_N(\omega_N),$$

where the last inequality can easily be seen via the definition of the discrepancy, or via the notion of dispersion (see e.g. [3]).

Applying Lemma 2 with $\varepsilon = MD_N(\omega_N)$ finally gives:

$$|D_N(\omega_N) - D_{N,H}(\bar{\omega}_N)| \leq 2MD_N(\omega_N)$$

and thus $D_{N,H}(\bar{\omega}_N) \leq (1 + 2M)D_N(\omega_N)$.

For the multi-dimensional version we can bound the one-dimensional projections like in the one-dimensional case (5), so again applying Lemma 2 we get

$$D_{N,H}(\bar{\omega}_N) \leq D_N(\omega_N) + (1 + 2MD_N(\omega_N))^s - 1$$

Expanding the binomial term and using $(D_N)^k \leq D_N$, since $D_N \leq 1$, we finally get the desired result $D_{N,H}(\bar{\omega}_N) \leq (1 + 2M)^s D_N(\omega_N)$. □

Remark 2. From the proof, it can readily be seen that this bound holds for every construction that leads to

$$y_k^{(l)} \in \left[x_k^{(l)-}, x_k^{(l)+}\right].$$

Thus the kind of interpolation is not relevant for the discrepancy bound, as the smoothness of the interpolation is not taken into account. Using some additional restrictions on the interpolation, one might find even better bounds.

In order to integrate functions with singularities at the boundary it will be convenient to shift the interpolated sequence in an appropriate way to avoid regions that lie too close to the singularity.

Corollary 1. *Let $(\bar{\omega}_N) = \{y_1, \ldots, y_n\}$ be constructed as in Theorem 2. Then the sequence $\hat{\omega}_N = \{\hat{y}_1, \ldots, \hat{y}_n\}$ defined by*

$$\hat{y}_k^{(l)} = \begin{cases} x_k^{(l)+} & \text{if } \mathcal{A} = \emptyset, \\ x_k^{(l)-} & \text{if } \mathcal{B} = \emptyset, \\ y_k^{(l)} & \text{otherwise,} \end{cases}$$

has an H-discrepancy of order

$$D_{N,H}(\hat{\omega}) \leq (1 + 2M)^s D_N(\omega)$$

and the same distance

$$\min_{k=1,\ldots,N} \min_{1 \leq j \leq s} \min(\hat{y}_k^{(j)}, 1 - \hat{y}_k^{(j)})$$

to the boundaries as the original sequence ω.

In the construction one might question why one does not use the point set $\left\{\frac{k}{N}\right\}_{0 \leq k \leq N}$ to approximate the distribution function. However, in that case, adding one single point to the set would then also change the whole set of support points. As a result, the distribution function for all support points would have to be reevaluated, which in practice is the numerically expensive part of the calculation. If one uses the points of the original low-discrepancy sequence, the distribution function only has to be evaluated at the new point, although all $\mathbf{y}_k$ will still have to be readjusted.

4.1 Using Different Sequences for Approximation and Inversion

In the previous section we investigated a transformation, where the distribution function in each dimension was approximated using the corresponding low-discrepancy sequence for that dimension. Thus the distribution function has to be evaluated sN times. As this evaluation is the numerically expensive part of the generation for moderate values of N it is of advantage for practical applications to lower the number of evaluations.

The idea now is to use the same one-dimensional low-discrepancy sequence $\hat{\omega} = (z_i)_{0 \leq i \leq N}$ for all dimensions to approximate the distribution function. If two or more dimensions share the same marginal distribution, the cumulative distribution function has to be evaluated only N times instead of a multiple of N. Again, the resulting sequence displays the low-discrepancy property:

Theorem 3. *Let $\hat{\omega} = (z_i)_{1 \leq i \leq N}$ be a one-dimensional sequence with discrepancy $D_N(\hat{\omega})$, and $\omega = (\mathbf{x}_i)_{1 \leq i \leq N}$ an s-dimensional sequence with discrepancy $D_N(\omega)$. Let furthermore $H(\mathbf{x})$ like in Theorem 2, and similarly define*

$$z_k^{(l)-} = \max_{\mathcal{A} = \left\{ z_i \in \hat{\omega} \;\middle|\; H_l(z_i) \leq x_k^{(l)} \right\}} z_i \quad \text{and} \quad z_k^{(l)+} = \min_{\mathcal{B} = \left\{ z_i \in \hat{\omega} \;\middle|\; H_l(z_i) \geq x_k^{(l)} \right\}} z_i.$$

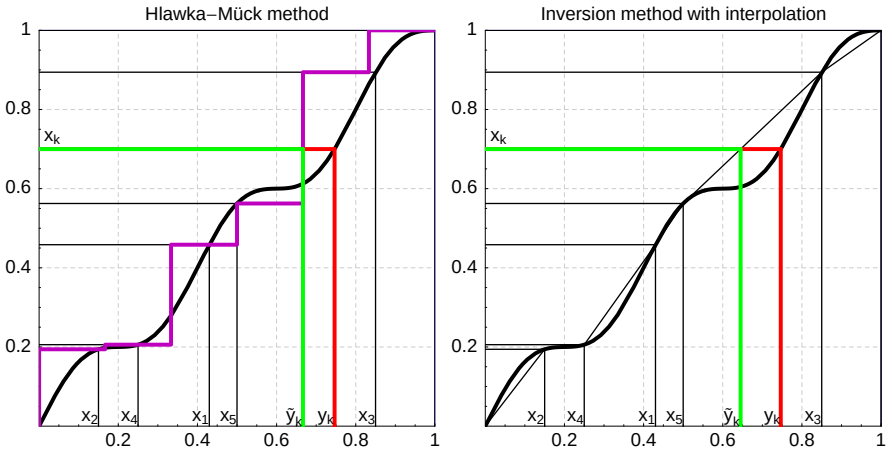

Fig. 1. Hlawka-Mück and our interpolation construction

Again, we set $z_k^{(l)-} = 0$ *if* $\mathcal{A} = \emptyset$ *and* $z_k^{(l)+} = 1$ *if* $\mathcal{B} = \emptyset$.

Then the H*-discrepancy of any transformed sequence* $\bar{\omega} = (\mathbf{y}_k)_{1 \le k \le N}$ *with the property* $y_k^{(l)} \in \left[z_k^{(l)-}, z_k^{(l)+}\right]$ *for all* $0 \le k \le N$ *and* $0 \le l \le s$ *can be bounded by*

$$D_{N,H}(\bar{\omega}) \le D_N(\hat{\omega}) + D_N(\omega)(1+2M)^s \,.$$

Proof. Similar to the proof of Theorem 2 we obtain in one dimension

$$\left|y_k - H^{-1}(x_k)\right| \le \left|z_k^+ - z_k^-\right| \le D_N(\hat{\omega}_N) \,,$$

and from this $|H(y_k) - x_k| \le M D_N(\hat{\omega}_N)$. As a result, we have $D_{N,H}(\bar{\omega}_N) = D_N(H(\bar{\omega}_N)) \le 2M D_N(\hat{\omega}_N) + D_N(\omega_N)$.

Applying the same steps to the one-dimensional projections, and using the same relations as in the previous theorem, we get the bound

$$D_{N,H}(\bar{\omega}_N) \le D_N(\omega_N) + D_N(\hat{\omega}_N)(1+2M)^s$$

for the multi-dimensional case. □

To get a better understanding of the differences in the Hlawka-Mück method and our interpolation method, both are depicted in figure 1.

In many applications, like the evaluation of Asian options with a given distribution of the stock prices (see e.g. [4]), all dimensions share the same one-dimensional distribution, and thus the distribution function can be factored into a product of s identical factors: $H(\mathbf{x}) = \prod_{i=1}^{s} H^{(1)}(x^{(i)})$, where $H^{(1)}(x)$ denotes the one-dimensional distribution function. In that case, the distribution function has to be evaluated only N times, instead of sN time as in other methods.

Since the sequence to approximate the distribution function and the sequence used for inversion are now decoupled, we can lower the generation effort even more by pre-sorting the support points $\left(H^{(1)}(z_k)\right)_{1\le k\le N}$:

Lemma 3. *Let $H(\mathbf{x}) = \prod_{i=1}^{s} H^{(1)}(x^{(i)})$. The numerical cost of generating an N-element, H-distributed low-discrepancy sequence as defined in Theorem 3 has a numerical cost of $\mathcal{O}(N \log N)$.*

Proof. The generation of the H-distributed sequence consists of several steps:

(1) Generation of the uniformly distributed sequence $(\hat{\omega}_N)$,
(2) Generation of the uniformly distributed sequence (ω_N),
(3) Calculation of the distribution function $\hat{H}_N = H^{(1)}(\hat{\omega}_N)$,
(4) Pre-sorting the support points $\hat{H}_N$,
(5) For each $1 \le n \le N$
   a) finding the corresponding values $\mathbf{z}_k^-$ and $\mathbf{z}_k^+$, and
   b) calculating the resulting point $\mathbf{y}_k$.

Clearly, (1), (2), and (3) are of order $\mathcal{O}(N)$. Sorting an N-element set of numerical values is of order $\mathcal{O}(N \log N)$ using Merge Sort or Heap Sort (see [10]). Finally, for each of the N elements, finding the values of $\mathbf{z}_k^{(l)-}$ and $\mathbf{z}_k^{(l)+}$ is of order $\mathcal{O}(\log N)$ since the $\hat{H}_N$ are already sorted. The actual calculation of $\mathbf{y}_k^{(l)}$ from the $\mathbf{z}_k^{(l)\pm}$ is of constant order for each of the N elements. Thus we obtain an asymptotic order of

$$3\mathcal{O}(N) + \mathcal{O}(N \log N) + \mathcal{O}(N \log N) + \mathcal{O}(N) = \mathcal{O}(N \log N). \qquad \square$$

Remark 3. If one does not use an N-element sequence as $\hat{\omega}_N$, but an $\tilde{N} = p^{\lceil \log_p N \rceil}$-element sequence, one can always add new points to the sequence in linear effort, until the number of elements gets larger then $\tilde{N}$. Only for these logarithmically many points all N points created so far need to be readjusted. For all other cases, however, the already existing points do not need to be touched. This is of advantage and will also lower the total simulation effort if one does not know the exact number of required points a priori.

4.2 Comparing the Actual Discrepancy

As a quick check of our theoretical results, we compared the discrepancy of the sequences generated by the Hlawka-Mück and by our interpolated transformation method with the discrepancy of the original sequence. Unfortunately, the L^∞- (or extreme) discrepancy $D_N(\bar{\omega}_N)$ cannot be calculated explicitly in dimensions higher than 2, so we will compared the L^2-discrepancy, which describes the mean error instead of the maximum error over all subintervals of $[0,1)^s$ containing the origin. As one can already expect from the discrepancy bounds proved above, for both transforms we do not see any effect in the L^2-discrepancy compared to the untransformed sequences.

5 Non-Uniform Integration of Singular Functions

Using the results from the previous sections, one can bound the QMC integration error for functions of finite variation. However, for functions with a singularity - which appear for example in many problems from finance - these bounds are infinite.

Sobol' [17] proved a convergence theorem for singular integration using uniformly distributed low-discrepancy sequences, Owen [15] proved the corresponding error orders under some growth conditions on the function. Hartinger, Kainhofer, and Tichy [5] proved a similar convergence theorem for non-uniformly distributed sequences, albeit using an L-shaped region for cutting off the integral. Sobol and Owen, in contrast, mainly looked at hyperbolic regions, which require more sophisticated proof techniques, but can give better error orders in general. Both their proofs make use of the so-called low-variation extension of a function (see [16], although the idea is due to Sobol'). In the following we will use Owen's notations, where $(a : b)$ denotes the set of integers a through b, while for $u \subset (1 : s)$ and $\mathbf{x}, \mathbf{y} \in U^s$ we denote by $\mathbf{x}^u : \mathbf{y}^{-u}$ the point where the coordinates u are taken from $\mathbf{x}$, while the coordinates $(1 : s) \setminus u$ are taken from $\mathbf{y}$. Also, we will use two special types of regions that exclude a certain volume around the origin or all corners:

$$K_{min}^{orig}(\varepsilon) = \left\{ \mathbf{x} \in U^s \middle| \min_{1 \le j \le s} x^{(j)} > \varepsilon \right\} \tag{6}$$

$$K_{prod}^{orig}(\varepsilon) = \left\{ \mathbf{x} \in U^s \middle| \prod_{j=1}^{s} x^{(j)} > \varepsilon \right\} \tag{7}$$

$$K_{min}^{corner}(\varepsilon) = \left\{ \mathbf{x} \in U^s \middle| \min_{1 \le j \le s} \min(x^{(j)}, 1 - x^{(j)}) > \varepsilon \right\} \tag{8}$$

$$K_{min}^{corner}(\varepsilon) = \left\{ \mathbf{x} \in U^s \middle| \prod_{j=1}^{s} \min(x^{(j)}, 1 - x^{(j)}) > \varepsilon \right\} \tag{9}$$

K_{min}^{orig} avoids the origin and the lower boundaries via an L-shaped region, while K_{prod}^{orig} avoids it via a hyperbolic region. K_{min}^{corner} and K_{prod}^{corner} have similar avoidance patterns, but for all corners at the same time.

Definition 4 (low-variation extension). *Let $f : U^s \mapsto \mathbb{R}$ be an s-times differentiable function (possibly unbounded at the definition boundaries, but bounded inside). Furthermore, let $K \subseteq U^s$ a region with anchor $\mathbf{c} \in U^s$. That is, for each $\mathbf{x} \in K$ we have $[\mathbf{x}, \mathbf{c}] \subseteq K$. Then the low-variation extension $\tilde{f}$ of f from K to U^s is defined by*

$$\tilde{f}(\mathbf{x}) = f(\mathbf{c}) + \sum_{\emptyset \ne u \subseteq (1:s)} (-1)^{|u|} \int_{[\mathbf{x}^{(u)}, \mathbf{c}^{(u)}]} \mathbb{1}_{\mathbf{z}^{(u)} : \mathbf{c}^{(-u)} \in K} \partial^u f \left(\mathbf{z}^{(u)} : \mathbf{c}^{(-u)} \right) \mathrm{d}\mathbf{z}^{(u)} . \tag{10}$$

Owen [16] showed that its Vitali and Hardy-Krause variation are bounded by

$$V_{U^s}(\tilde{f}) \leq \int_K \left|\partial^{(1:d)} f(\mathbf{x})\right| \mathrm{d}\mathbf{x} \tag{11}$$

$$V_{\mathrm{HK}}(\tilde{f}) \leq \sum_{u\neq\emptyset} \int_{K_u(\mathbf{1}^{(-u)})} \left|\partial^u f\left(\mathbf{x}^{(u)} : \mathbf{1}^{(-u)}\right)\right| \mathrm{d}\mathbf{x}^{(u)}\,, \tag{12}$$

using the definition $K_u(\mathbf{b}^{(-u)}) = \left\{\mathbf{x}^{(u)} \in U^{|u|} | \mathbf{x}^{(u)} : \mathbf{b}^{(-u)} \in K\right\}$.

We will in the sequel only consider singular functions that fulfill one of the growth conditions for some $A_j > 0$, $B < \infty$, and all $u \subseteq (1:s)$:

$$|\partial^u f(\mathbf{x})| \leq B \prod_{j=1}^{s} \left(x^{(j)}\right)^{-A_j - \mathbb{1}_{j\in u}}, \text{ or} \tag{13}$$

$$|\partial^u f(\mathbf{x})| \leq B \prod_{j=1}^{s} \min\left(x^{(j)}, 1-x^{(j)}\right)^{-A_j - \mathbb{1}_{j\in u}}\,. \tag{14}$$

5.1 *L*-shaped Regions

As a first case we will consider sequences that lie in $K^{orig}_{min}(\varepsilon)$, and thus avoid the origin in an L-shaped region. This property can easily be seen for the Halton and general $(0,s)$ sequences, and also for non-uniform low-discrepancy sequences that are generated by the Hlawka-Mück transformation (with shift, as shown in [5]). This case was already investigated by the authors in [5], but no explicit error bounds were given. In [4] it was applied to the special example of pricing an Asian option, and error bounds were given for that specific problem.

The error bounds given by Owen [16] for the uniform distribution are easily generalized.

Theorem 4. *Let $f : U^s \mapsto \mathbb{R}$, and $\omega_N = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be a sequence with $\mathbf{x}_j \in K^{orig}_{min}(\varepsilon_N)$ for $1 \leq j \leq N$. Let furthermore $H(\mathbf{x})$ be a distribution on U^s with density $h(\mathbf{x})$ and $M_\varepsilon = \sup_{\mathbf{x}\in U^s \setminus K^{orig}_{min}(\varepsilon)} h(\mathbf{x}) \leq \infty$. If f fulfills growth condition* (13)*, and $0 < \varepsilon_N = CN^{-r} < 1$, then*

$$\left|\int_{U^s} f(\mathbf{x})\mathrm{dH}(\mathbf{x}) - \frac{1}{N}\sum_{n=1}^{N} f\left(\mathbf{x}_n\right)\right| \leq C_1 D_{N,H} N^{r\sum_{j=1}^{s} A_j} + C_2 N^{r(\max A_j - 1)} M_{\varepsilon_N} \tag{15}$$

with some explicitly computable, finite constants C_1, and C_2.

Also, if $\mathbf{x}_j \in K^{corner}_{min}(\varepsilon_N)$ for all j with $0 < \varepsilon_N = CN^{-r} < 1/2$, and f is a real-valued function on $(0,1)^s$ that fulfills growth condition (14)*, then* (15) *holds. M_ε has to be taken as the supremum over $U^s \setminus K^{corner}_{min}(\varepsilon)$ in that case.*

The proof is obtained by replacing the Koksma-Hlawka bound in [15, Proof of Theorem 5.2] by Chelson's non-uniform bound (1) and factoring out the supremum of the density M_ε when necessary.

Proof. Using a $3\,\varepsilon$-argument, we have

$$\left|\int_{U^s} f(\mathbf{x})\mathrm{dH}(\mathbf{x}) - \frac{1}{N}\sum_{n=1}^{N} f\left(\mathbf{x}_n\right)\right| \leq \left|\int_{U^s}\left(f(\mathbf{x}) - \tilde{f}(\mathbf{x})\right)\mathrm{dH}(\mathbf{x})\right| + \left|\int_{U^s}\tilde{f}(\mathbf{x})\mathrm{dH}(\mathbf{x}) - \frac{1}{N}\sum_{n=1}^{N}\tilde{f}\left(\mathbf{x}_n\right)\right| + \left|\frac{1}{N}\sum_{n=1}^{N}\tilde{f}\left(\mathbf{x}_n\right) - \frac{1}{N}\sum_{n=1}^{N} f\left(\mathbf{x}_n\right)\right| \quad (16)$$

The last term vanishes, since $f(\mathbf{x}) = \tilde{f}(\mathbf{x})$ on K.

The second term can be bounded by $V_{HK}(\tilde{f})D_{N,H}(\omega)$ using the non-uniform Koksma-Hlawka inequality (1), and using Owen's inequality (12) for $V_{HK}(\tilde{f})$ even further by $C_1 N^{r\sum_{j=1}^{s} A_j} D_{N,H}(\omega)$ with $C_1 = B\prod_{j=1}^{s} C^{-A_j} A_j^{-1}$.

Finally, for the first term we use Lemma 5.1 of [16]: If $K \subseteq U^s$ with anchor $c = \mathbf{1}$, and f fulfills growth condition (13), then for all $\mathbf{x} \in \mathbf{U}_K = U^s - K$ we have $\left|f(\mathbf{x}) - \tilde{f}(\mathbf{x})\right| \leq \tilde{B}\prod_{j=1}^{s}\left(x^{(j)}\right)^{-A_j}$ with $\tilde{B} = B\prod_{j=1}^{s}\left(1 + \frac{1}{A_j}\right)$.

Thus, the first term can be bounded by

$$\left|\int_{U^s}\left(f(\mathbf{x}) - \tilde{f}(\mathbf{x})\right)\mathrm{dH}(\mathbf{x})\right| \leq \int_{\mathbf{U}_K}\left|f(\mathbf{x}) - \tilde{f}(\mathbf{x})\right| h(\mathbf{x})\mathrm{d}\mathbf{x} \leq$$

$$M_{\varepsilon_N}\tilde{B}\int_{\mathbf{U}_K}\prod_{j=1}^{s}\left(x^{(j)}\right)^{-A_j}\mathrm{d}\mathbf{x} \leq M_{\varepsilon_N}\tilde{B}\prod_{j=1}^{s}\left(\frac{1}{1-A_j}\right) s\, C^{1-\min A_k} N^{r(\max A_k - 1)}.$$

The last inequality follows from direct integration, similar to [15, Proof of Theorem 5.2]. Thus we have $C_2 = \tilde{B}\left(\prod_{j=1}^{s}\frac{1}{1-A_j}\right) sC^{1-\min A_j}$.

For the corner-case, we note that the unit cube can be partitioned into 2^s cubes with anchor $\frac{\mathbf{1}}{\mathbf{2}} = \left(\frac{1}{2},\ldots,\frac{1}{2}\right)$, and each of them can be bounded like above. Furthermore, the variation on each of them sums up to the variation on the whole unit interval, thus we get the same bound with an additional factor 2^s in the constants. □

Remark 4. Suppose that one uses some classical low discrepancy construction (e.g. Sobol, Faure, or Halton sequences) in combination with (shifted) Hlawka-Mück or the (shifted) interpolation method. Then $r = 1$ and $D_{N,H} \leq CN^{-1+\varepsilon}$, and the obtained error will be of the order

$$\mathcal{O}\left(N^{-1+\varepsilon+\sum_{j=1}^{s} A_j}\right).$$

When using importance sampling with a distribution that has different tail behavior than the original distribution, one often ends up with a singular integral, where the density also has a singularity at the boundary. In this case, M_ε is not finite, and the bound from above does not give any sensible result.

On the other hand, if the density tends to zero, one has to expect that the effect of the singularity of the functions should be somehow lightened.

Thus we will now look at densities that fulfill another "growth condition" in a region $U^s \setminus K$ around the origin:

$$\forall \mathbf{x} \in U^s \setminus K: \; h(\mathbf{x}) \leq C_h \prod_{j=1}^{s} \left(x^{(j)}\right)^{-\tilde{A}_j}, \quad \text{for some } \tilde{A}_j < 1, C_h \in \mathbb{R} \quad (17)$$

If $\tilde{A}_j = 1$, the bound is not integrable any more. However, since h is a distribution density function and thus integrable, one should be able to find an integrable bound. We will also assume that $A_j + \tilde{A}_j < 1$, as otherwise the bound for the whole integral would be infinite.

Using this growth condition, one can now prove a version of the theorem that takes into account the behavior of $h(\mathbf{x})$ near the origin (or all corners):

Theorem 5. *Let ω_N, H, and f be the sequence, distribution, and integrand from Theorem 4. If furthermore the density $h(\mathbf{x})$ satisfies the growth condition* (17)*, then*

$$\left| \int_{U^s} f(\mathbf{x}) \mathrm{dH}(\mathbf{x}) - \frac{1}{N} \sum_{n=1}^{N} f\left(\mathbf{x}_n\right) \right| \leq C_1 D_{N,H} N^{r \sum_{j=1}^{s} A_j} + \tilde{C}_2 N^{r\left(\max(A_j+\tilde{A}_j)-1\right)}$$

with C_1 from Theorem 4, and $\tilde{C}_2 = \tilde{B}\, C_h \prod_{j=1}^{s} \frac{1}{1-A_j-\tilde{A}_j} s\, C^{1-\min(A_j+\tilde{A}_j)}$.

Proof. The proof follows along the lines of Theorem 4, the major difference being in the bound for the first term:

$$\begin{aligned} \int_{U^s \setminus K} \left| f(\mathbf{x}) - \tilde{f}(\mathbf{x}) \right| h(\mathbf{x}) \mathrm{d}\mathbf{x} &\leq \tilde{B} C_h \int_{U^s \setminus K} \prod_{j=1}^{s} \left(x^{(j)}\right)^{-(A_j+\tilde{A}_j)} \mathrm{d}\mathbf{x} \\ &\leq \tilde{B} C_h \prod_{j=1}^{s} \frac{1}{1-(A_j+\tilde{A}_j)} C^{1-\min(A_j+\tilde{A}_j)} s N^{r\left(\max(A_j+\tilde{A}_j)-1\right)}. \quad \square \end{aligned}$$

5.2 Hyperbolic Regions

A serious improvement in the bound for the error order can be obtained by choosing sequences that avoid the origin in a hyperbolic sense (i.e. sequences that lie in $K^{orig}_{prod}(\varepsilon)$) and thus more strongly, as Owen [15] showed for the uniform distribution. In that case, the $\sum A_j$ in the bound can be replaced by $\max A_j$. We will now state a similar theorem for arbitrary distributions H:

Theorem 6. *Let $f(\mathbf{x})$ be a real-valued function on U^s (possibly unbounded at the lower boundary) which satisfies growth condition* (13)*. Let furthermore $\omega_N = (\mathbf{x}_i)_{1 \leq i \leq N}$ be a point set with $\mathbf{x}_i \in K^{orig}_{prod}(\varepsilon_N)$ and $0 < \varepsilon_N = CN^{-r} < 1$ for some constants $C, r > 0$. Finally, let $H(\mathbf{x})$ be a distribution on U^s with density $h(\mathbf{x})$ that satisfies growth condition* (17)*. Then for all $\eta, \tilde{\eta} > 0$ we have*

$$\left|\int_{U^s} f(\mathbf{x})\mathrm{dH}(\mathbf{x}) - \frac{1}{N}\sum_{n=1}^{N} f(\mathbf{x}_n)\right| \leq C_\eta^{(1)} D_{N,H}(\omega)\, N^{\eta + r\max_j A_j} + C_{\tilde{\eta}}^{(2)} N^{\tilde{\eta} + r\max_j (A_j + \tilde{A}_j) - r} \quad (18)$$

for constants $C_\eta^{(1)}$ and $C_{\tilde{\eta}}^{(2)}$. A similar bound holds for the corner case when $\varepsilon_N < 2^{-s}$. The bound holds with $\eta = 0$ if the maximum among the A_j is unique, and with $\tilde{\eta} = 0$ if the maximum among the $A_j + \tilde{A}_j$ is unique.

Proof. We again denote by $\tilde{f}$ the low-variation extension of f from $K_{prod}^{orig}(\varepsilon_N)$ to U^s with anchor $\mathbf{1}$. Again (16) holds, and the first term can be bounded by

$$\int_{U^s\setminus K} \left|f(\mathbf{x}) - \tilde{f}(\mathbf{x})\right| h(\mathbf{x})\mathrm{d}\mathbf{x} \leq \tilde{B}C_h \int_{U^s\setminus K} \prod_{j=1}^{s} \left(x^{(j)}\right)^{-(A_j+\tilde{A}_j)} \mathrm{d}x = \mathcal{O}\left(\varepsilon^{1-\max_j (A_j+\tilde{A}_j)}\right) \quad (19)$$

using a lemma of Sobol ([17, Lemma 3] or [16, Lemma 5.4]) if all $A_j + \tilde{A}_j$ are distinct. If any two of the $A_j + \tilde{A}_j$ are equal, and they are not the maximum, one can increase one of them by a small value without affecting the max. If the maximum is not distinct, one has to increase some of them and thus the maximum by no more than $\tilde{\eta}/r$.

The variation of $\tilde{f}$ was already proved by Owen to be bounded by $V_{HK}(\tilde{f}) \leq C_1 N^{r\max_j A_j}$ if the maximum among the A_j is distinct. If this is not the case a similar argument like before brings in the η in the bound.

Combining these two bounds, we arrive at (18).

The corner case can be argued similarly (see [16, Proof of Theorem 5.5]) by splitting the unit cube into 2^s subcubes and investigating each separately. □

Remark 5. Determination of the asymptotics of ε_N for hyperbolic regions is more delicate than for the L-shaped regions. In particular for the corner case not much is known even for the classical sequences and the uniform distribution (see e.g. [15]). Neverthenless, it is obvious that sequences obtained by Hlawka-Mück's construction or by interpolation do not result in better asymptotics than N^{-s}. Thus, in combination with classical low discrepancy sequences one will get error estimates of the order

$$\mathcal{O}\left(N^{-1+\varepsilon+s\max_{j=1,\ldots,s} A_j}\right).$$

Acknowledgments

The authors want to thank Art Owen for several inspiring discussions and suggestions. This research was supported in part by the Austrian Science Fund Project S-8308-MAT.

References

1. P. Chelson. *Quasi-Random Techniques for Monte Carlo Methods*. PhD. Dissertation, The Claremont Graduate School, 1976.
2. L. Devroye. *Nonuniform random variate generation*. Springer-Verlag, New York, 1986.
3. M. Drmota and R. F. Tichy. *Sequences, Discrepancies and Applications*, volume 1651 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1997.
4. J. Hartinger, R. Kainhofer, and M. Predota. Option pricing as an unbounded non-uniform Quasi-Monte Carlo integration problem. Submitted, 2004.
5. J. Hartinger, R. Kainhofer, and R. Tichy. Quasi-Monte Carlo algorithms for unbounded, weighted integration problems. *Journal of Complexity*, 20(5):654–558, 2004.
6. S. G. Henderson, B. A. Chiera, and R. M. Cooke. Generating "dependent" quasi-random numbers. In *Proceedings of the 2000 Winter Simulation Conference*, pages 527–536. IEEE, 2000.
7. E. Hlawka. Gleichverteilung und Simulation. *Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II*, 206:183–216, 1997.
8. E. Hlawka and R. Mück. A transformation of equidistributed sequences. In *Applications of number theory to numerical analysis*, pages 371–388. Academic Press, New York, 1972.
9. E. Hlawka and R. Mück. Über eine Transformation von gleichverteilten Folgen. II. *Computing*, 9:127–138, 1972.
10. D. E. Knuth. *The art of computer programming. Volume 3. Sorting and searching*. Addison-Wesley, Reading, Massachusetts, 1973.
11. T. Kollig and A. Keller. Efficient multidimensional sampling. *Computer Graphics Forum*, 21(3):557 – 563, 2002.
12. W. J. Morokoff and R. E. Caflisch. Quasi-Monte Carlo integration. *J. Comput. Phys.*, 122(2):218–230, 1995.
13. R. B. Nelsen. *An introduction to copulas*, volume 139 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1999.
14. H. Niederreiter. *Random number generation and Quasi-Monte Carlo methods*, volume 63 of *SIAM Conf. Ser. Appl. Math.* SIAM, Philadelphia, 1992.
15. A. B. Owen. Halton sequences avoid the origin. Technical report, Stanford University, 2004.
16. A. B. Owen. Multidimensional variation for quasi-monte carlo. Technical report, Stanford University, 2004.
17. I. M. Sobol'. Calculation of improper integrals using uniformly distributed sequences. *Soviet Math. Dokl.*, 14(3):734 – 738, 1973.
18. X. Wang. Improving the rejection sampling method in quasi-Monte Carlo methods. *J. Comput. Appl. Math.*, 114(2):231–246, 2000.