A Pragmatic view on Numerical Integration of Unbounded Functions

Heinz Hofbauer[1], Andreas Uhl[1], and Peter Zinterhof[2]

[1] Salzburg University, Department of Computer Sciences
{hhofbaue,uhl}@cosy.sbg.ac.at
[2] Salzburg University, Department of Computer Sciences
peter.zinterhof@sbg.ac.at

Summary. We take a pragmatic approach to numerical integration of unbounded functions. In this context we discuss and evaluate the practical application of a method suited also for non-specialists and application developers. We will show that this method can be applied to a rich body of functions, and evaluate it's merits in comparison to other methods for integration of unbounded integrals. Furthermore, we will give experimental results to illustrate certain issues in the actual application and to confirm theoretic results.

1 Introduction

The basic concept of any QMC method for numerical integration is to approximate the integral by a finite sum, such that

$$I(f) := \int_{U^s} f(x)dx \approx \frac{1}{N}\sum_{n=1}^{N} f(x_n) =: I'_N(f)$$

where x_n are suitably chosen and U^s is the unit cube. To identify suitable, i.e. uniformly distributed, points x_n the star discrepancy is defined as

$$D_N^* := D_N^*(x_1, \ldots, x_n) = \sup_{J\in\mathcal{F}} \left\| \frac{\#\{x|x \in J\}}{N} - m(J) \right\|$$

where $\mathcal{F}$ is the family of all subintervals of the form $J = \prod_{i=1}^{d}[0, t_i) \in U^s$ with volume $m(J)$. The approximation error

$$E_N(f) := |I'_N(f) - I(f)|$$

depends on D_N^* and the variation $V(f)$ in the sense of Hardy and Krause, see [Nie92] or [Owe04b] for details, of the function f. The dependency is stated in the Koksma-Hlawka inequality

$$E_N(f) \leq V(f)D_N^*(f)$$

which is the fundamental error bound for quasi-Monte Carlo methods.

A big problem with numerical integration is the fact that the variation is rather restrictive. Even simple functions like $m(x) = max(x_1 + x_2 + x_3 - 1, 0)$ are of unbounded variation $V(m) = \infty$, see [Owe04b]. Also if a function is unbounded the variation is unbounded resulting in an error estimate $E_N(f) \leq \infty$.

There have already been a number of methods proposed in literature which aim at tackling the problem of numerically integrating unbounded functions, these will be described in Section 2. The method proposed in this work is discussed in detail in Section 3. In Section 4 we give experimental results which indicate that the method may even be applied to functions not contained in the restrictive class of functions it is proved for, associated problems are also discussed.

2 Methods for the Numerical Integration of Unbounded Functions

In the case of singularities the Koksma-Hlawka inequality becomes meaningless since functions containing singularities are unbound and thus of infinite variation. When examining methods of numerical integration for integrands with singularities usually the distinction is made whether the singularities are in the interior of the unit cube or on the boundary.

2.1 Singularities on the Boundary

Sobol' [Sob73] investigated a number of functions which have singularities in the origin. By restricting the growth of the integral by

$$D_N \int_{a_N}^{1} |f(x)|dx = o(N)$$

for $N \to \infty$, where $a_N = \min_{1 \leq i \leq N} x_i$, he shows that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{\mu=1}^{N} f(P_\mu) = \int_{U^s} f(P)dP$$

holds, but unfortunately fails to give an error bound. He allows one dimensional functions $f(x)$ to have a rational singularity $0 < \xi < 1$ but reduces them to functions with singularities in the origin. In the multi dimensional case Sobol's test function is $f(x) = x_1^{-\beta_1} \cdots x_s^{-\beta_s}$ where the growth condition holds if $\forall i \, \beta_i < 1$.

An important class of methods deals with point generation sequences which avoid the corners. Owen [Owe04a] deals with singularities by replacing the part of the function which is not touched by the numerical integration, i.e. the part lying in the hyperbolic or L-shaped region avoided by the Halton sequence, by a bounded extension of the function. This way he attains a finite variation for the function and can prove error bounds for the numerical integration.

Definition 1. *Let* $1 > \epsilon > 0$ *then the following regions are subsets of the* s*-dimensional unit cube* $U^s = [0,1]^s$

$$H_o(\epsilon) = \{x \in U^s | \prod_{i=1}^{s} x_i \geq \epsilon\} \qquad H(\epsilon) = \{x \in U^s | \prod_{i=1}^{s} \min(x_i, 1 - x_i) \geq \epsilon\}$$

$$L_o(\epsilon) = \{x \in U^s | \min_{1 \leq i \leq s}^{s} x_i \geq \epsilon\} \quad L(\epsilon) = \{x \in U^s | \min_{1 \leq i \leq s} \min(x_i, 1 - x_i) \geq \epsilon\}.$$

The region H_o *excludes a* hyperbolic region *near the origin and* H *excludes hyperbolic regions near all corners of the unit cube. Likewise,* L_o *excludes a* L-shaped region *near the origin and* L *near all corners.*

Definition 2. *Let* $1 > \epsilon > 0$*, then if a sequence of* N *points* P_N *which fulfills*

$$\forall x \in P_N \Rightarrow x \in H_o(\epsilon)$$

we say the sequence avoids the origin in a hyperbolic fashion. *The same holds for* H *(*avoids all corners*),* L_o *(*in an L-shaped fashion*) and* L *(*avoids all corners*).*

Remark 1. It is not unusual to differentiate between corners since a given sequence usually doesn't avoid all corners to the same degree, i.e. with the same ϵ.

Owen shows that the Halton sequences avoid all corners in a hyperbolical sense. He also shows that for a finite $C > 0$ the Halton points $x_1, \ldots, x_n$ avoid the hyperbolic region $\{x | \prod_j x^j \leq Cn^{-1}\}$, while independent uniform points x_n enter that region infinitely often, with probability one. It is also shown that while points from the Halton sequence avoid the 0 and 1 corner stronger than independent uniform points do this doesn't hold for all other corners. To show an error bound for numerical integration with point sequences which avoid the origin in a hyperbolic $\{x \in [0,1]^d | \prod_{1 \leq i \leq d} x^i \geq \epsilon\}$ or L-shaped $\{x \in [0,1]^d | \min_{1 \leq i \leq d} x^i \geq \epsilon\}$ way Owen also imposes growth conditions on the functions.

Hartinger et al. show in [HKZ05] that generalized Niederreiter sequences possess corner avoidance properties similar to Halton sequences around the origin. They also show the corner avoidance rates for Halton and Faure sequences for corners different than the origin. To get efficient QMC rules for the

integrands one has to find point sets satisfying the condition $\prod_{i=1}^{s} x_n^{(i)} \geq cN^{-r}$ with small r as stated in [Owe04a] or for an all corner case when the avoidance condition is written as $\min_{1\leq n\leq N} \prod_{i=1}^{s} \min(1-x_n^{(i)}, x_n^{(i)}) \geq cN^{-r}$.

They show that for each point $\mathbf{x}_n$, $0 \leq n < b^l$, of a generalized Niederreiter (t,s)-sequence in base b the bound $\prod_{i=1}^{s} x_n^{(i)} \geq b^{-l-t-s}$ holds.

Kainhofer, Hartinger, and Tichy [HKT04] also dealt with QMC methods for multidimensional integrals with respect to a measure other than the uniform distribution. They allow the integrand to be unbounded on the lower boundary of the interval and justify the "strategy of ignoring the singularity" by using weighted integration with a non-uniform distribution. This means integration problems of the form $I_{[\mathbf{a},\mathbf{b}]} := \int_{[\mathbf{a},\mathbf{b}]} f(\mathbf{x}) dH(\mathbf{x})$ where H denotes a s-dimensional distribution with support $K = [\mathbf{a},\mathbf{b}] \subset \mathbb{R}^s$ and f is a function with singularities on the left boundary of K. To use a generalized version of the Koksma-Hlawka inequality they have to define a H-discrepancy of $\omega = (y_1, \ldots)$ which measures the distribution properties of the sequence. It is defined as $D_{N,H}(\omega) = \sup_{J\subset K} |N^{-1}A_N(J,\omega) - H(J)|$ where A_N counts the number of elements in $(y_1, \ldots, y_N)$ falling into the interval J, e.g. $A_N(J,\omega) = \sum_{n=1}^{s} \chi_J(y_n)$, and $H(J)$ denotes the probability of $J \subset K$ under H. With this $D_{N,H}$ they can define the Koksma-Hlawka inequality for this case as $|I_K - N^{-1}\sum_{n=1}^{N} f(y_n)| \leq V(f)D_{N,H}(\omega)$.

While the authors state that there is a certain lack of sequences with low H-discrepancy they also propose a technique for constructing such sequences by using the Hlawka and Mück method [HM72]. However, for such a sequence $\tilde{\omega}$ there might be some elements $\tilde{y}_k$ which attain 0. Since the singularities of $f(x)$ are on the lower boundary these sequences are not directly suited to be used with the numerical integration, however a simple change is proposed, generating a new sequence $\bar{\omega}$.

For the multi-dimensional case the idea is basically the same, Kainhofer et al. state a convergence criterion as follows. Similar to the one-dimensional case the construction of H-distributed sequences leads to problems when using Hlawkas method. However an adjustment to the generated sequence is given by the authors.

In [HK05] Hartinger and Kainhofer deal with the problem of generating low discrepancy sequences with an arbitrary distribution H. While they did so before ([HKT04]) they identify some disadvantages which carry over to the transformed sequence they proposed. They specifically deal with the property of the Hlawka-Mück method that for some applications the points of the generated sequence of a set with cardinality N lie on a lattice with spacing $1/N$. Their solution is to use a smoothed approximation where the values between the jumps are interpolated in the empirical distribution function.

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

generate a new sequence $\hat{\omega}$ from the constructed sequence $\bar{\omega}$ which has the same distance to the boundaries as the original sequence ω.

They show, by utilizing the same techniques as Owen in [Owe04a], that the sequence can be used to integrate improper integrals which have a singularity in the corner. They also show an error estimate for the L-shaped and hyperbolic corner avoidance cases.

De Doncker [dDG03] reduces the error rate of the methods of Klinger and Sobol' by constructing extensions which reduce the approximation error. She looks at the leading asymptotic order of the error and generates extrapolations for such functions in such a way that error terms vanish. She has shown that for one dimensional functions with algebraic end point singularities her method works very well. Furthermore, it gains significant convergence acceleration when applied to some logarithmic and interior algebraic singularities. Additionally an asymptotic error expansion was derived for integrands with algebraic singularities at the boundaries of the d-dimensional unit cube.

The improvements were found to occur in stages, as each error term vanishes. She also states that further research is needed to determine conditions for which an exact order of leading error terms can be established, and thus a proper extrapolation can be made.

2.2 Singularities in the Interior

In [Owe04c] Owen applied and extended his results from [Owe04a] to singularities $z \in [1,0]^d$ inside the unit cube. However since the sequences do not avoid the region of the singularity, which can be in the interior of the unit cube, he proposes using the extended function $\tilde{f}$ instead of the original function f for numerical integration. Like in [Owe04a] he requires the function to obey a growth condition. He then defines an extendible region K around singularity z for which $||x - z||_p \geq \nu$ holds for some $\nu > 0$, additionally he defines an anchor $c \in K$ for which $\text{rect}[c, y] \subset K \ \forall y \in K$ holds, where $\text{rect}[x, y] = \prod_{i=1}^{s} [\min(x_i, y_i), \max(x_i, y_i)]$ is the rectangular hull of x and y, thus he can use Sobol's extension $\tilde{f}(x)$. With the help of the extendible region, $\tilde{f}$ and the growth condition he gives an error estimate for any Lebesgue measurable function f for the integration.

However, Owen states in the conclusion that it is not clear if such a good extension to f can be found for arbitrary level sets.

Klinger [Kli97] shows that the numerical integration of a function is still possible when it has a singularity in the origin, or can be transformed such that the singularity is in the origin, by removing the point closest to the origin from the integration. This basically excludes an elemental interval containing the origin from the estimation, for Halton sequences he defines similar intervals. Only Halton and $(0, s)$-sequences are used and Klinger uses the properties of elemental intervals to find points which are near the singularity and thus not included in the numerical integration. While the $(0, s)$-sequence is a Niederre-

iter sequence, and thus a notion of elemental intervals already exits, Klinger needs to define a similar notion for Halton sequences:

Let a_k be positive rational numbers which satisfy $\sum_{k=1}^{s} 1/a_k \geq 1$ and define positive numbers, coprime integers p_k, q_k by $p_k/q_k := a_k$. Now let $R = \prod_{k=1}^{s}[\delta_k(\delta_k - CN^{-1/a_k}), (1-\delta_k)CN^{-1/a_k} + \delta_k)$ where $\delta_k \in \{0,1\}$ and $C = \min_{1\leq k\leq s} b_k^{-q_k}$. Then at most one of the first N points $x_0, \ldots, x_N$ of Halton's sequence falls into R.

Consequently Klinger states that since $x_0 = \mathbf{0}$, this argument also shows that none of the first N points of a Halton sequence x_n fall into the above interval when $\delta_k = 0$ for all $1 \leq k \leq s$. He gives the error bounds for Halton and $(0, s)$-sequences.

The computational experiments included in the paper compare the error bounds of the Halton, Sobol' and Niederreiter sequences. The error bounds are shown to be reasonable and also that Halton sequences have, for low dimension, basically the same characteristics as Sobol' or Niederreiter sequences but are less computationally expensive. For high dimension, shown for dimension 10 in the experiments, Halton sequences are worse for at least moderate N.

The method proposed in the next section is rather simple in application, and can deal with arbitrary patterns of singularities. However, this entails a rather problematic (at least theoretical) restriction to the class of functions to which it can be applied.

3 A Pragmatic Approach

A number of people, starting with Sobol' in [Sob73], have conducted research for error bounds for improper integrals. One of the more recent results is by Zinterhof [Zin02].

Definition 3. *For a function $f(x)$ and a $B > 0$ the functions $f_B(x)$ and $\hat{f}_B(x)$ are defined as*

$$f_B(x) = \begin{cases} f(x) & |f(x)| \leq B \\ 0 & |f(x)| > B \end{cases}$$

$$\hat{f}_B(x) = \begin{cases} 0 & |f(x)| \leq B \\ f(x) & |f(x)| > B. \end{cases}$$

Definition 4. *Consider the class of s-variate functions, $f(x_1, \ldots, x_s)$ $0 \leq x_i \leq 1$ $i = 1, \ldots, s$, consisting of all functions which fulfill*

$$\begin{aligned} &(a)\ I(|\hat{f}_B|) = O(B^{-\beta}) \text{ for some } \beta > 0 \\ &(b)\ V(f_B) = O(B^{\gamma}) \text{ for some } \gamma \geq 1. \end{aligned}$$

This class will be called $C(\beta, \gamma)$.

Theorem 1. *Let $f \in C(\beta,\gamma)$, D_N^* be the discrepancy of the set of nodes $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and $B = {D_N^*}^{-1/(\beta+\gamma)}$, then the estimate*

$$I(f) = \frac{1}{N}\sum_{n=1}^{N} f_B(\mathbf{x}_n) + \mathrm{O}({D_N^*}^{\beta/(\beta+\gamma)})$$

holds, where $I(f) = I(f_B) + I(\hat{f}_B)$.

Proof. From

$$I(f) = I(f_B) + I(\hat{f}_B)$$

using the Hlawka-Koksma inequality we get

$$\left| I(f_B) - \frac{1}{N}\sum_{n=1}^{N} f_B(x_n) \right| \le V(f_B) D_N^* \le C_1(f) B^\gamma D_N^*$$

and from condition (a) we get

$$|I(\hat{f}_B)| \le C_2(f) B^{-\beta}.$$

Consequently

$$E_N = \left| I(f) - \frac{1}{N}\sum_{n=1}^{N} f_B(x_n) \right| \le C_1(f) B^\gamma D_N^* + C_2(f) B^{-\beta},$$

which takes it's minimum of order when using

$$B = {D_N^*}^{\frac{-1}{\beta+\gamma}}.$$

Thus for an error estimate we get

$$E_N \le C(f) {D_N^*}^{\beta/(\beta+\gamma)}$$

where $C(f)$ is a constant depending on f.

Remark 2. Zinterhof [Zin02] also shows that the error bound is optimal.

Remark 3. Since the optimal B is depending on D_N^*, β and γ it is in any case depending on N. Also, if either of D_N^*, β or γ is depending on s then B is also depending on s.

3.1 The Class $D(\beta,\gamma)$

An important issue remains: the richness of the class $C(\beta,\gamma)$. Generally if the jump line, i.e. $f(x) = B$, $x \in U^s$, is not axis parallel the variation is

unbounded and consequently $f \notin C(\beta,\gamma)$ since condition (b) (in Definition 4) is violated [3]. It can clearly be seen that the class $C(\beta,\gamma)$ is very restrictive.

Consider the class T of bounded step functions $t \in T$, $t(x_1,...,x_s) : U^s \to \mathbb{C}$. These are functions which are piecewise constant on U^s where U^s is partitioned into a finite number of, pairwise disjoint, intervals of the form $\prod_{i=1}^{s}[a_i,b_i)$. All functions $t \in T$ are of bounded variation.

Definition 5. *Let the class $D(\beta,\gamma)$ be defined as $D(\beta,\gamma) := \{f | f \in C(\beta,\gamma)$ and $f_B \in T\}$. If it is clear from context we will abbreviate and write D.*

Remark 4. Using the definition of D we can easily state $T \subset D \subset C := C(\beta,\gamma)$.

It is well known that if $g \in L_1$, which implies $\int_{U^s} |g(\mathbf{x})| d\mathbf{x} < \infty$, then there exists for every $\epsilon > 0$ a $t_\epsilon(x) \in T$ such that

$$\int_{U^s} |g(\mathbf{x}) - t_\epsilon(\mathbf{x})| d\mathbf{x} < \epsilon.$$

Also since $T \subset D$ we can easily write

$$\int_{U^s} |g(\mathbf{x}) - d_\epsilon(\mathbf{x})| d\mathbf{x} < \epsilon,$$

with $d_\epsilon(x) \in D$. Generally, the functions $\in D$ will approximate a given function $g \in C \subset L_1$ better than functions $\in T$.

In any case, C is rich since T is rich and $T \subset C$. The restrictiveness of $C(\beta,\gamma)$ is a direct result of the restrictiveness of the variation in the sense of Hardy and Krause.

3.2 The function $f(\mathbf{x}) = \max(x_1,...,x_s)^{-\beta}$

Now let us consider the function $f = \frac{1}{\max(x_1,...,x_s)^\beta}$ with $0 < \beta < 1$. Certainly $f \notin T$ and $\lim_{\mathbf{x}\to 0} f(\mathbf{x}) = \infty$, thus if we can show that $f \in C$ we have $T \subsetneq C$. To do this we need to estimate the integral value and variation of the function f to see if conditions (a) and (b) in Definition 4 are met.

Integral Value

Theorem 2. *Let $f = \frac{1}{\max(x_1,...,x_s)^\beta}$ with $0 < \beta < 1$, then $\int_{I^s} f(\mathbf{x}) d\mathbf{x} = \frac{s}{s-\beta}$ $0 < \beta < 1$.*

Proof. With induction. For $s = 1$ the claim holds since $\int_0^1 x_1^{-\beta} dx_1 = \frac{1}{-\beta+1} x_1^{-\beta+1}|_0^1 = \frac{1}{1-\beta}$.

Now

[3] Thanks to Reinhold Kainhofer for pointing this out.

$$\int_0^1 \dots \int_0^1 \max(x_1, \dots, x_s)^{-\beta} dx_1 \dots dx_s =$$
$$= \int_0^1 \left(\int_0^1 \dots \int_0^1 \max(x_1, \dots, x_s)^{-\beta} dx_1 \dots dx_{s-1} \right) dx_s =$$
$$= \int_0^1 \dots \int_0^1 dx_1 \dots dx_{s-1} \int_0^1 \max(x_1, \dots, x_s)^{-\beta} dx_s.$$

Let us consider $\int_0^1 \max(x_1, \dots, x_{s-1}, x_s)^{-\beta} dx_s$, and let $\hat{x}_s := \max(x_1, \dots, x_{s-1})$, thus $\int_0^1 \max(x_1, \dots, x_{s-1}, x_s)^{-\beta} dx_s = \int_0^1 \max(\hat{x}_s, x_s)^{-\beta} dx_s$ where

$$\max(\hat{x}_s, x_s) = \begin{cases} x_s & x_s \geq \hat{x}_s \\ \hat{x}_s & x_s < \hat{x}_s. \end{cases}$$

Thus

$$\int_0^1 \max(\hat{x}_s, x_s)^{-\beta} dx_s = \int_0^{\hat{x}_s} \hat{x}_s^{-\beta} dx_s + \int_{\hat{x}_s}^1 x_s^{-\beta} dx_s = \frac{-\beta \hat{x}_s^{-\beta+1} + 1}{1-\beta}.$$

Now we have

$$\int_0^1 \dots \int_0^1 dx_1 \dots dx_{s-1} \int_0^1 \max(x_1, \dots, x_s)^{-\beta} dx_s =$$
$$= \int_0^1 \dots \int_0^1 dx_1 \dots dx_{s-1} \frac{1 - \beta \hat{x}_s^{-\beta+1}}{1-\beta} =$$
$$= \frac{1}{1-\beta} \left[1 - \beta \int_0^1 \dots \int_0^1 \max(x_1, \dots, x_{s-1}^{-\beta+1} dx_1 \dots dx_{s-1} \right].$$

Now using the induction hypothesis we get

$$\int_0^1 \dots \int_0^1 dx_1 \dots dx_{s-1} \int_0^1 \max(x_1, \dots, x_s)^{-\beta} dx_s =$$
$$= \frac{1}{1-\beta} \left[1 - \beta \frac{s-1}{s-1-\beta+1} \right] = \frac{s - \beta - s\beta + \beta}{(1-\beta)(s-\beta)} = \frac{s}{s-\beta}.$$

Remark 5. In a similar fashion we obtain $\int_{I^s} \hat{f}_B(\mathbf{x}) d\mathbf{x} = \frac{s}{s-\beta} (\frac{1}{B})^{\frac{s-\beta}{\beta}}$.

Variation

Definition 6. *Let* $\mathbf{P}$ *be the set of all partitions of the* s*-dimensional unit cube* I^s *then the variation of a function* f *in the sense of Vitali is defined as*

$$V_V(f) := \sup_{P \in \mathbf{P}} \sum_{p \in P} |\Delta(f; p)|,$$

where $\Delta(f; p)$ *is the* s*-fold alternate sum, i.e. adjacent corners have opposite sign, of the function values on the corners of the interval* p*.*

Definition 7. *Let $n \in \mathbf{N}$ and $0 = t_0 < \ldots < t_{n-1} < t_n = 1$, $t_i \in I^s$ $0 \leq i \leq n$. Now let $Z_s(t_0, \ldots, t_n) = \{\{t_1, \ldots, t_n\}^s\}$ be the set of s-tuples formed by $t_0, \ldots, t_n$. A partition P of s-dimensional unit cube $I^s = [0,1)^s$ is called* valid partition *if there is exists a $Z_s(t_0, \ldots, t_n)$ such that $P = \{\{[t_0, t_1), [t_1, t_2), \ldots, [t_{n-1}, t_n)\}^s\}$ where $[t_{k-1}, t_k)[t_{l-1}, t_l) = [t_{k-1}, t_k) \times [t_{l-1}, t_l)$. We say the partition P belongs to $Z_s(t_0, \ldots, t_n)$ and write $P(Z_s(t_0, \ldots, t_n))$.*

Remark 6. From the construction of $Z_s(t_0, \ldots, t_n)$ it follows that for a valid partition only the intervals of the form $[t_{k-1}, t_k) \times \cdots \times [t_{k-1}, t_k)$ $1 \leq k \leq n$ cross the principal diagonal, i.e. the restriction of the principal diagonal of the unit cube to such an interval is the principal diagonal of the interval.

Remark 7. Every partition of I^s can be refined to a valid partition.

Lemma 1. *$V_V(\max(x_1, \ldots, x_s)) = 1$ for $(x_1, \ldots, x_s) \in I^s$.*

Proof. The function $f(x_1, \ldots, x_s) = \max(x_1, \ldots, x_s)$ fulfills $\max(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_s) = x_k$ for $\max(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_s) \leq x_k$. Let $I_{n_1,\ldots,n_s} = [t_{n_1-1}, t_{n_1}) \times \cdots \times [t_{n_s-1}, t_{n_s})$ be an interval of the valid partition $P(Z_s(t_0, \ldots, t_n))$, which doesn't cross the principal diagonal $(x_1, \ldots, x_s) = t(1, \ldots, 1)\, 0 \leq t \leq 1$. Now we can write $V_V(f; I_{n_1,\ldots,n_s}) = |\sum_{\tau_1=0}^{1} \cdots \sum_{\tau_s=0}^{1} (-1)^{\tau_1+\cdots+\tau_s} \max(t_{n_1-1} + \tau_1(t_{n_1} - t_{n_1-1}), \ldots, t_{n_s-1} + \tau_s(t_{n_s} - t_{n_s-1})|$. Since $I_{n_1,\ldots,n_s}$ is not on the principal diagonal of I^s there is a k_0, $1 \leq k_0 \leq s$ such that

$$\max\big(t_{n_1-1} + \tau_1(t_{n_1} - t_{n_1-1}), \ldots, t_{n_{k_0}-1} + \tau_{k_0}(t_{n_{k_0}} - t_{n_{k_0}-1}), t_{n_{k_0}}, t_{n_{k_0}+1} + \tau_{k_0+1}(t_{n_{k_0}} - t_{n_{k_0}-1}), \ldots\big) = t_{n_{k_0}}$$

and

$$\max\left(t_{n_1-1} + \tau_1(t_{n_1} - t_{n_1-1}), \ldots, t_{n_s-1} + \tau_s(t_{n_s} - t_{n_s-1})\right) = t_{n_{k_0}-1} + \tau_{k_0}(t_{n_{k_0}} - t_{n_{k_0}-1})$$

for all $\tau_1, \ldots, \tau_{k_0-1}, \tau_{k_0+1}, \ldots, \tau_s$, $\forall i : \tau_i \in \{0, 1\}$.

It follows that

$$V_V(f; I_{n_1,\ldots,n_s}) = \left| \sum_{\tau_1=0}^{1} \cdots \sum_{\tau_{k_0-1}=0}^{1} \sum_{\tau_{k_0+1}=0}^{1} \cdots \sum_{\tau_s=0}^{1} (-1)^{\tau_1+\cdots+\tau_{k_0-1}+\tau_{k_0+1}+\cdots+\tau_s} (t_{n_{k_0}} - t_{n_{k_0}-1}) \right| =$$

$$= (t_{n_{k_0}} - t_{n_{k_0}-1}) \left| \sum_{\tau_1=0}^{1} \cdots \sum_{\tau_{k_0-1}=0}^{1} \sum_{\tau_{k_0+1}=0}^{1} \cdots \sum_{\tau_s=0}^{1} (-1)^{\tau_1+\cdots+\tau_{k_0-1}+\tau_{k_0+1}+\cdots+\tau_s} \right| = 0.$$

If on the other hand $I_{n_1,\ldots,n_s}$ lies on the principal diagonal of I^s, then $n_1 = \cdots = n_s = n_0$ and $I_{n_1,\ldots,n_s}$=$I_{n_0,\ldots,n_0}$= $[t_{n_0-1}, t_{n_0}) \times [t_{n_0-1}, t_{n_0}) \times \ldots \times [t_{n_0-1}, t_{n_0})$, then

$$\begin{aligned}
V_V(f; I_{n_0,\ldots,n_0}) = \Bigg| & \sum_{\tau_1=0}^{1} \cdots \sum_{\tau_s=0}^{1} (-1)^{\tau_1+\cdots+\tau_s} \\
& \max\left(t_{n_0-1} + \tau_1(t_{n_0} - t_{n_0-1}), \ldots, t_{n_0-1} + \tau_s(t_{n_0} - t_{n_0-1})\right)\Bigg| = \\
= & \left| \sum_{\tau_1,\ldots,\tau_s=0}^{1} (-1)^{\tau_1+\cdots+\tau_s} t_{n_0-1} \max(\tau_1, \ldots, \tau_s)(t_{n_0} - t_{n_0-1}) \right| = \\
= & \Bigg| t_{n_0-1} \sum_{\tau_1,\ldots,\tau_s=0}^{1} (-1)^{\tau_1+\cdots+\tau_s} + \\
& (t_{n_0} - t_{n_0-1}) \left(\sum_{\tau_1,\ldots,\tau_s=0}^{1} (-1)^{\tau_1+\cdots+\tau_s} - \sum_{\tau_1,\ldots,\tau_s=0}^{1} (-1)^{\tau_1+\cdots+\tau_s} \right) \Bigg| = \\
= & \left| t_{n_0-1} 0 + (t_{n_0} - t_{n_0-1})(0-1) \right| = t_{n_0} - t_{n_0-1}.
\end{aligned}$$

Then holds $V_V(f; I^s) = \sum_{n_0=1}^{n} (t_{n_0} - t_{n_0-1}) = 1$, where $V_V(f; I^s)$ is attained already at the principal diagonal of I^s.

Remark 8. If $g(x)$ in $[0, 1]$ is monotone or of finite variation $\mathrm{Var}(g)$ then $V_V(g(\max(x_1, \ldots, x_s); I^s)) = |g(1) - g(0)|$ or $V_V(g(\max(x_1, \ldots, x_s); I^s)) = \mathrm{Var}(g)$.

Remark 9. Let $0 \le a_k < b_k$, $1 \le k \le s$ then for $I_{a,b} = \prod_{k=1}^{s} [a_k, b_k)$ we analogously get $V_V(g(\max(x_1, \ldots, x_s)); I_{a,b}) = |g(b) - g(a)|$ for $a_1 = \cdots = a_s$ and $b_1 = \cdots = b_s$, otherwise $V_V(g(\max(x_1, \ldots, x_s)); I_{a,b}) = 0$. The variation in the sense of Vitali of functions $g(\max(x_1, \ldots, x_s))$ is concentrated on the principal diagonal of the unit cube.

Remark 10. For functions $f = g(\max(x_1, \ldots, x_s))$

$$V_{HK}(f; I^s) = V_V(f; I^s)$$

holds. The variation in the sense of Hardy and Krause is defined as

$$V_{HK}(f; I_{a,b}) = \sum_{J \subset I^s} V_V(f; J)$$

where J are all k-dimensional faces $\{(u_1, \ldots, u_s) \in I^s | u_j = 1, j \ne i_1, \ldots, i_k\}$ with $1 \le k \le s$ and $1 \le i_1 < \cdots < i_k \le s$. Since for $k < s$ there are some $u_j = 1$, the function $f = g(\max(x_1, \ldots, x_{t-1}, 1, x_{t+1}, \ldots, x_s)) = g(1)$, $t \ne i_1, \ldots, i_k$, is constant and consequently $V_V(f; J) = 0$, $\forall J \subsetneq I^s$.

Let now $g(x) = 1/x^\beta$, $0 < x \leq 1$, $0 < \beta < 1$ and $f_\beta(x_1, \ldots, x_s) = 1/\max(x_1, \ldots, x_s)^\beta$, $(x_1, \ldots, x_s) \in I^s$. Now let

$$\hat{f}_{\beta,B} = \begin{cases} 0 & f_\beta(x_1, \ldots, x_s) > B,\ \max(x_1, \ldots, x_s) < 1/B^\beta = B' \\ f_\beta(x_1, \ldots, x_s) & f_\beta(x_1, \ldots, x_s) \leq B,\ \max(x_1, \ldots, x_s) \geq 1/B^\beta = B', \end{cases}$$

and

$$\tilde{f}_{\beta,B} = \begin{cases} f_\beta(B', \ldots, B') = B & f_\beta(x_1, \ldots, x_s) > B \\ f_\beta(x_1, \ldots, x_s) & f_\beta(x_1, \ldots, x_s) \leq B, \end{cases}$$

and

$$\chi_{\beta,B} = \begin{cases} B & f_\beta(x_1, \ldots, x_s) > B \\ 0 & f_\beta(x_1, \ldots, x_s) \leq B, \end{cases}$$

and clearly $\tilde{f}_{\beta,B} = \hat{f}_{\beta,B} + \chi_{\beta,B}$. It can be easily seen that $V_V(\chi_{\beta,B}; I^s) = B$ and from the remarks before we know that $V_V(\tilde{f}_{\beta,B}) = |g(0) - g(1)| = B - 1$. Consequently, $V_V(\hat{f}_{\beta,B}; I^s) = V_V(\tilde{f}_{\beta,B} - \chi_{\beta,B}; I^s) \leq V_V(\tilde{f}_{\beta,B}; I^s) + V_V(\chi_{\beta,B}; I^s) = 2B - 1$.

Thus we have finally shown that $f(\mathbf{x})(= \max(\mathbf{x})^{-\beta})$, $0 < \beta < 1$ is in $C(\frac{s-\beta}{\beta}, 1)$.

Remark 11. Since $\max(\mathbf{x})^{-\beta} \in C(\frac{s-\beta}{\beta}, 1)$, $0 < \beta < 1$, we also know that the error takes it's minimum when $B = D_N^{*\frac{-\beta}{s}}$ (c.f.: proof of Theorem1).

4 Experimental Results

First, we want to investigate the behavior of function $f(\mathbf{x}) = \max(x_1, ..., x_s)^{-\beta}$ as discussed in the last section in numerical experiments. As point sequence we used the Zinterhof sequence [Zin69], which is a special case of Weyl sequences defined as follows

$$x_n = (\{ne^{1/1}\}, \ldots, \{ne^{1/s}\}), \quad n = 1, 2, 3, \ldots,$$

for points $n = 1, 2, ...$ and dimension s. Note that the Zinterhof sequence has certain corner avoidance properties as well, which is due to the high degree of irrationality of the generated points. Caused by corresponding diophantine properties this is true not only for the origin but for all rational points as well. For the calculation of the bound B we use the bound of the discrepancy given by LeVeques inequality [KY81] and the diaphony of the Zinterhof sequence [Zin69].

Figure 1 displays the results for the original and integral preserving transformed function (transformed in such a way as to get singularities in the interior of the unit interval as well as on the border)

$$\max = \max(x_1, ..., x_s)^{-0.5}, \qquad \max\,' = \max(\{5x_1\}, ..., \{5x_s\})^{-0.5}$$

respectively, where $\{x\}$ is the remainder of x. We let N run and hold $B = 2$ (according to Remark 11) fixed for dimensions 10 and 15 (labeled `d10` and `d15` respectively).

Remark 12. We hold B fixed at a value which is calculated for $N = 10^7$ so that we will get the best result towards the end of calculation. If we wanted the lowest error for each N we would have to let B vary accordingly.

As expected, the the error rates are very good, especially towards higher N.

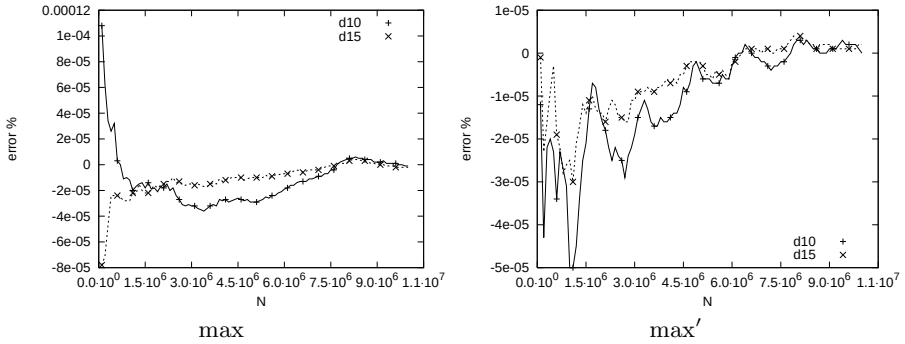


Fig. 1. Functions max and max′ for dimension 10 and 15 with $B = 2$, relative error over N.

Theoretically, we are restricted to functions of class C, practically however the method can be applied to a wider range of functions. Consider the functions

$$f^1 = \prod_{i=1}^{s} \frac{1}{x_i^{0.5}}, \qquad f^2 = \prod_{i=1}^{s} \frac{1}{\ln(\frac{1}{x_i})^{0.5}}$$

where f_B^1 and f_B^2 both have non axis parallel jump curves and consequently infinity variation.

Remark 13. The integral values over the s-dimensional unit cube for $f_\alpha^1(\mathbf{x}) = \prod_{i=1}^{s} x_i^{-\alpha}$ and $f_\alpha^2(\mathbf{x}) = \prod_{i=1}^{s} \ln(1/x_i)^{\alpha-1}$, $0 < \alpha < 1$, are $\int_{U^s} f_\alpha^1(\mathbf{x})d\mathbf{x} = (1/(1-\alpha))^s$ and $\int_{U^s} f_\alpha^2(\mathbf{x})d\mathbf{x} = \Gamma(\alpha)^s$ respectively.

Experimentally, functions f_1 and f_2 can be integrated using our technique, even though their bound representations for this method have infinite variation, c.f. Definition 4 condition (b). For a test we used a fixed $B = 10^9$, which also hints at a serious problem with this method if $f \notin C$. Since the variations

of f_B^1 and f_B^2 are infinite we can not obtain a β and thus no optimal bound B using dimension 10 and 15.

Figure 2 left hand side shows the results for function f^1 over the number of points N, and the right side shows the results for the integral preserving transformation

$$f'^1 = \prod_{i=1}^{s} \frac{1}{\{5x_i\}^{0.5}}$$

where $\{x\}$ is again the remainder of x.

The figures show that the estimation converges toward a fixed error, this is to be expected since we will by construction always miss $I(\hat{f}_B)$ (see Theorem 1) since we kept B fixed while it is actually a function of the discrepancy and thus of N. The difference in error between dimension 10 and 15 is a well known phenomenon (curse of dimensionality). However, given that we can somehow obtain the proper bound B for the number of points N used for the integration the error converges even though f^1 (and f_B^1) is of unbounded variation.

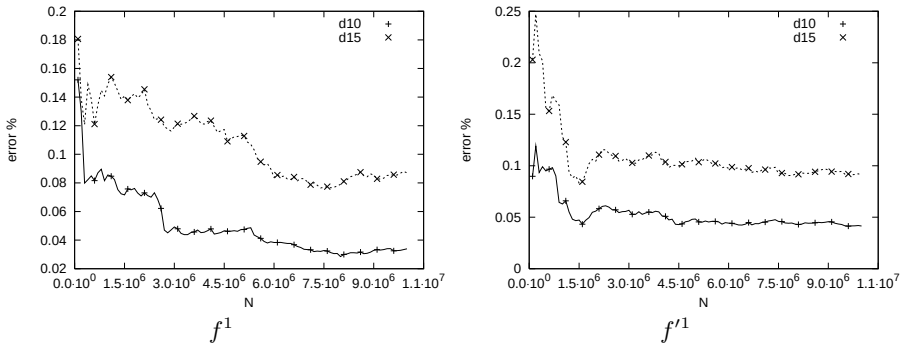


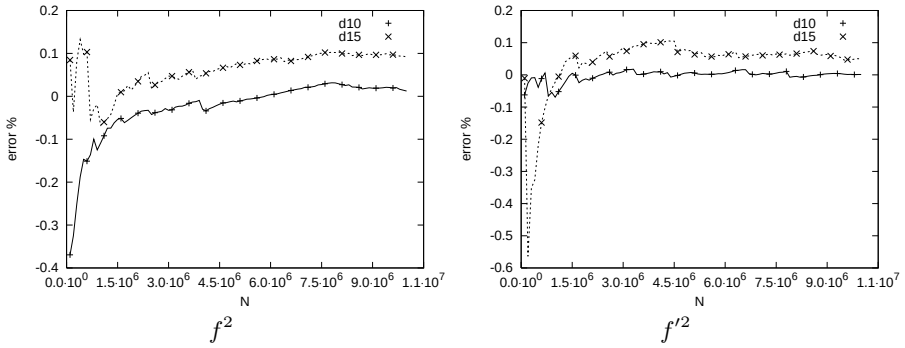
Fig. 2. Functions f^1 and f'^1 for dimension 10 and 15 with $B = 10^9$, relative error over N.

Keeping the bound B fixed and again using dimensions 10 and 15 we will likely experience problems in the integration when we turn to another function. To illustrate this we used function f^2, and an integral preserving transformation as follows

$$f'^2 = \prod_{i=1}^{s} \frac{1}{\ln(\frac{1}{\{5x_i\}})^{0.5}},$$

shown in Fig. 3 left and right hand side respectively. When the bound is chosen too low the results usually becomes stable quickly with a high error, stemming again from $\hat{f}_B^2$. In this case the bound was chosen too high, i.e. we would need to use more points N to get to the region where B is optimal. This

can be seen from the overshoots, usually high error rates at the beginning, due to points falling near the jump curve, thus early introducing high values, i.e. close to B, to the estimation. These will usually vanish when the number of points is high enough to get a fine grained sampling of the unit cube but will stay visible a long time. So while the method works, experimentally, even for functions not in C this poses the problem of estimating a proper B to be used in the integration.



f^2 f'^2

Fig. 3. Functions f^2 and f'^2 for dimension 10 and 15 with $B = 10^9$, relative error over N.

To illustrate the effect B has on the integration we use a fixed number of points $N = 10^6$ and let B vary. The result of this test, again for functions f^1 and f^2 in dimensions 10 and 15, is given in Fig. 4. What can be seen is that the bound depends not only on the number of points but also on the dimension, which is not surprising since it depends on the discrepancy. Also, for f^2, there is an interval of B where the integration holds, while on the other hand we get an increase in error as we move away from that interval. Also, since the bound is depending on the discrepancy, which in turn depends on the number of points and the dimension, the bound is a function of the dimension leading to quite some error in dimension 15 where the approximation was very close for dimension 10.

Finally, we consider a function which can not be reduced to singularities along the border or in the corner. Consider the function (again with integral preserving transformation)

$$m = \sum_{i=1}^{s-1} \frac{1}{|x_i - x_{i+1}|^{0.5}}, \qquad m' = \sum_{i=1}^{s-1} \frac{1}{|\{5x_i\} - \{5x_{i+1}\}|^{0.5}}.$$

Remark 14. The integral value of $m_\alpha(\mathbf{x}) = \sum_{i=1}^{s-1} |x_i - x_{i+1}|^{-\alpha}$, $0 < \alpha < 1$ over the s-dimensional unit cube U^s is $\int_{U^s} m_\alpha(\mathbf{x}) d\mathbf{x} = 2(s-1)/(1-\alpha)(2-\alpha)$.

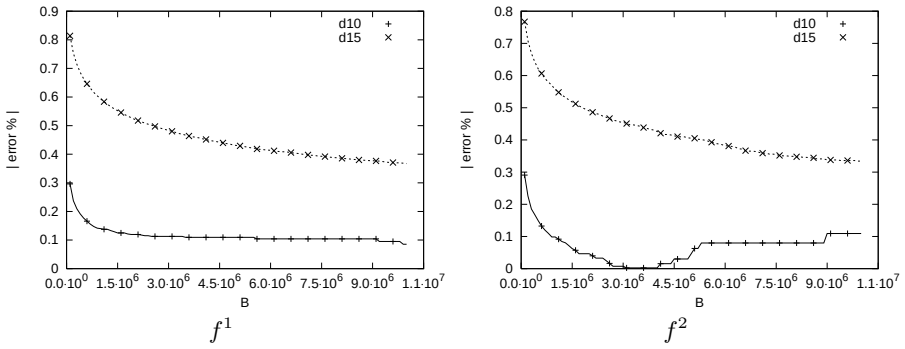


Fig. 4. Functions f^1 and f^2 for dimension 10 and 15 with $N = 10^6$, relative error over B.

Remark 15. The function m_α has a singularity along the $s-1$ dimensional manifold $R = \cup_{i=1,\ldots,s-1} R_i$, where $R_i = \{\mathbf{x} | \mathbf{x} \in U^s, x_i = x_{i+1}\}$, $i = 1, \ldots, s-1$.

Remark 16. If we use a step function m_M to approximate m with M intervals we can give a bound for the variation as $V_{HK}(m_M) \leq M 2^s s \binom{s}{\lfloor s/2 \rfloor} B = \mathrm{O}(B)$. Furthermore the integral of $\hat{m}_B$ can easily calculated as $\int_{U^s} \hat{m}_{\alpha B}(\mathbf{x}) d\mathbf{x} = (2(s-1)/(1-\alpha)(2-\alpha)) B^{-2\alpha+\alpha^2} = \mathrm{O}(B^{-(2\alpha-\alpha^2)})$ thus $m \in D(\alpha^2 - 2\alpha, 1)$.

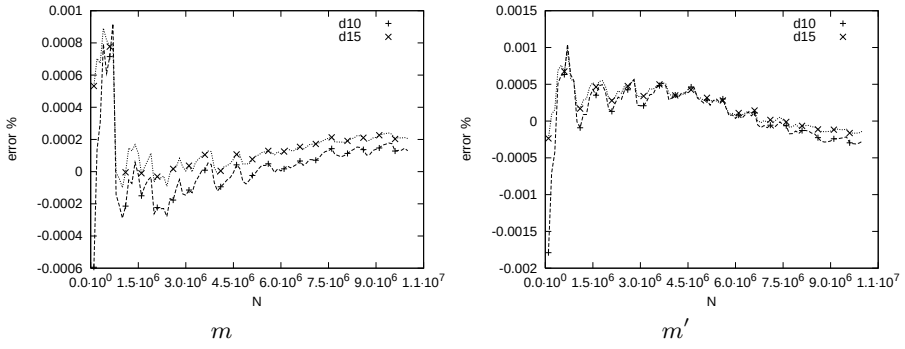


Fig. 5. Functions m and m' for dimension 10 and 15 with $B = 10^4$, relative error over N.

For m and m' we used $B = 10^4$ (according to Remark 16), again for both dimension 10 and 15, and the results are given in Fig. 5. As expected (since we use a single fixed B) the error at the beginning is quite high but

swiftly falls to "normal" levels. Somewhat more important is the rather good convergence when considering that this functions singularity is more severe than the singularity of the two previous functions.

Since we only used an estimation for B we can not expect the numerical integration to be optimal, so now we will assess how good the approximation actually is. For this purpose, we integrated m in dimension 10 for different values of B. The results are given in Fig. 6, the left side gives the relative error over N for different values of B and the right side gives the number of points for which the function value exceeded B for a fixed $N = 10^7$. On the left side we see that for a lower B the error is increased and if we set B to high the error increases again. Overall our estimated B seems to be a bit too low and some value between 10^4 and $2\ 10^4$ would have been the best fit. This is affirmed by the right hand side of the figure, where we see that the difference between 10^4 and $2\ 10^4$ is more than one point (otherwise it would not be possible to get results between these two values on the left hand side). Overall we see that with a function approximation using class $D(\beta, \gamma)$ it is hard to get the best approximation but it is certainly possible to get a good approximation. Also, when we look at the left hand side we see that for $B = 10^4$ and $B = 2\ 10^4$ the error increases again after about $N = 6\ 10^6$. This is a further evidence that the choice of B is vital for the integration.

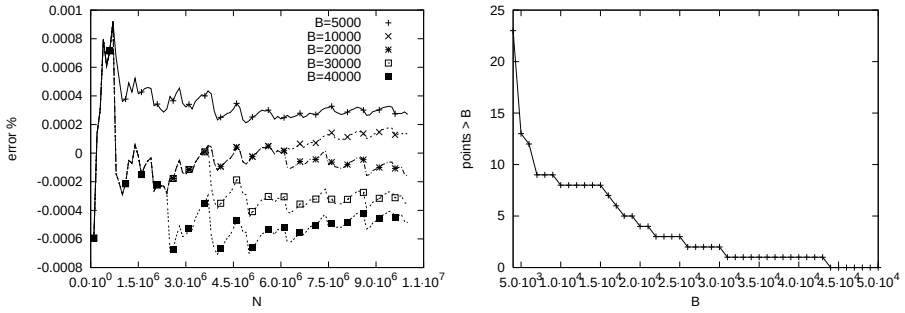


Fig. 6. The effects of different values of B for the function m.

5 Conclusion

We have shown that the proposed method can be used to numerically integrate over a rich class of functions $C(\beta, \gamma)$. The method also works experimentally on an even bigger class of functions with the problem that some parameters, i.e. the bound B, cannot be chosen specifically for the function. Furthermore, the bound can be applied during runtime, and thus the method can be applied

to the function directly. Also, the method is not restricted to singularities on the boundary or in the corner. Thus this method is extremely easy to implement and apply, even for non specialists.

However, even if a function is of class C we face the problem that we have to know the number of points beforehand to choose an optimal B. Also, since we need β and γ to choose optimal parameters for the numerical integration the function must be well known. This is theoretically of no importance but practically can prevent (optimal) integration.

References

[dDG03] Elise de Doncker and Yuqiang Guan, *Error bounds for integration of singular functions using equidistributed sequences*, Journal of Complexity **19** (2003), no. 3, 259–271.

[HK05] Jürgen Hartinger and Reinhold F. Kainhofer, *Non-uniform low-discrepancy sequence generation and integration of singular integrands*, Proceedings of MC2QMC2004, Juan-Les-Pins France, June 2004 (H. Niederreiter and D. Talay, eds.), Springer Verlag, June 2005.

[HKT04] Jürgen Hartinger, Reinhold F. Kainhofer, and Robert F. Tichy, *Quasi-monte carlo algorithms for unbound, weighted integration problems*, Journal of Complexity **20** (2004), no. 5, 654–668.

[HKZ05] Jürgen Hartinger, Reinhold Kainhofer, and Volker Ziegler, *On the corner avoidance properties of various low-discrepancy sequences*, in press.

[HM72] E. Hlawka and R. Mück, *Über eine transformation von gleichverteilten folgen*, Computing **9** (1972), 127–138.

[Kli97] B. Klinger, *Numerical Integration of Singular Integrands Using Low-Discrepancy Sequences*, Computing **59** (1997), 223–236.

[KY81] Hua Loo Keng and Wang Yuan, *Applications of Number Theory to Numerical Analysis*, Springer Verlag, Science Press, 1981.

[Nie92] Harald Niederreiter, *Random number generation and quasi-monte carlo methods*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 62, Society for Industrial and Applied Mathematics (SIAM), 1992.

[Owe04a] Art B. Owen, *Halton Sequences Avoid the Origin*, Tech. report, Stanford University, 2004.

[Owe04b] ———, *Multidimensional variations for quasi-Monte Carlo*, Tech. report, Stanford University, January 2004.

[Owe04c] ———, *Quasi-Monte Carlo for integrands with point singularities at unknown locations*, Tech. report, Stanford University, 2004.

[Sob73] I.M. Sobol', *Calculation of improper integrals using uniformly distributed sequences*, Soviet Math Dokl. **14** (1973), no. 3, 734–738.

[Zin69] Peter Zinterhof, *Einige zahlentheoretische methoden zur numerischen Quadratur und Interpolation*, Sitzungsberichte der Österreichischen Akademie der Wissenschaften, math.-nat.wiss. Klasse Abt. II **177** (1969), 51–77.

[Zin02] ———, *High dimensional improper integration procedures*, Proceedings of Contributions of the 7th International Scientific Conference (Hroncova 5, 04001 Košice, SLOVAKIA) (Civil Engineering Faculty Technical University of Košice, ed.), TULIP, May 2002, pp. 109–115.