MATHEMATICS OF COMPUTATION
Volume 73, Number 246, Pages 827–841
S 0025-5718(03)01569-2
Article electronically published on August 7, 2003

QUASI–MONTE CARLO INTEGRATION OVER $\mathbb{R}^d$

PETER MATHÉ AND GANG WEI

Abstract. In this paper we show that a wide class of integrals over $\mathbb{R}^d$ with a probability weight function can be evaluated using a quasi–Monte Carlo algorithm based on a proper decomposition of the domain $\mathbb{R}^d$ and arranging low discrepancy points over a series of hierarchical hypercubes. For certain classes of power/exponential decaying weights the algorithm is of optimal order.

1. Introduction, formulation of the problem

Usually, quasi–Monte Carlo integration is used to evaluate integrals

$$\mathrm{Int}_\rho(f) := \int f(\mathbf{x})\rho(\mathbf{x})\ d\mathbf{x},$$

where it is assumed that ρ is the uniform probability density function on $[0,1]^d$. In practical applications, probabilities are more general and are rather given on $\mathbb{R}^d$, such that preliminary transformations are required to fit the quasi–Monte Carlo integration setup. This can be done either by a change of variables or by splitting $\mathbb{R}^d$ into subcubes. Change of variables may lead to functions on $[0,1]^d$ with unbounded variation, even if the initial function had bounded variation.

Thus we propose to decompose

$$\mathbb{R}^d = \bigcup_{j=0}^{m+1} I_j, \qquad I_j := Q_j \backslash Q_{j-1},\ \ j = 1,\dots,m,$$

$$Q_j = \left\{\mathbf{x} \in \mathbb{R}^d,\ |x_l| < 2^j,\ l = 1,\dots,d\right\},\quad j = 0,1,\dots,m,$$

$$I_0 = Q_0,\ \ I_{m+1} = Q_m^c;$$

and we rewrite

$$\mathrm{Int}_\rho(f) = \int f(\mathbf{x})\rho(\mathbf{x})\ d\mathbf{x} = \sum_{j=0}^{m+1} \int_{I_j} f(\mathbf{x})\rho(\mathbf{x})\ d\mathbf{x}.$$

If ρ is bounded and smooth and if f has bounded variation and is integrable, then each $f(\mathbf{x})\rho(\mathbf{x})$ has bounded variation over I_j, as can be seen below. The above decomposition can be used to base quadrature formulas by replacing the integrals

Received by the editor June 11, 2002 and, in revised form, October 10, 2002.
2000 *Mathematics Subject Classification.* Primary 65C05; Secondary 68Q25.
Key words and phrases. Quasi–Monte Carlo integration, elliptically contoured distributions.

by quadrature formulas using low-discrepancy points, which results in

$$S_{m,\mathbf{n}}(f,\rho) := \sum_{j=0}^{m} \frac{q_j}{n_j} \sum_{i=1}^{n_j} f(\mathbf{y}_{ij})\rho(\mathbf{y}_{ij})\chi_{I_j}(\mathbf{y}_{ij}), \tag{1}$$

where $m+1$ is any number of cubes and $\mathbf{n} := (n_0, \dots, n_m)$, where each n_j is the number of points used within cube Q_j, $j = 0, \dots, m$, and the χ_{I_j} are the indicator functions. Furthermore,

$$\begin{aligned} q_j &= \text{Volume}(Q_j) = 2^{(j+1)d}, \\ \mathbf{y}_{ij} &= -2^j\mathbf{1} + 2^{(j+1)} * \mathbf{x}_{ij}, \quad 0 \le j \le m, \ \ 1 \le i \le n_j, \\ P_j &= \{\mathbf{x}_{ij} : 1 \le i \le n_j\} \subset [0,1)^d, \quad 0 \le j \le m. \end{aligned}$$

The total amount of function evaluations is $N = \sum_{j=0}^{m} n_j$. The P_j''s are some low-discrepancy point sets, abbreviated below as QMC point sets, consisting of n_j points, respectively.

Note that the hierarchy of cubes is fixed, except the number $m+1$ of them. This number depends on assumptions on the smoothness of the weight function and will be chosen appropriately.

Also, the point sets P_j, in particular their cardinalities n_j, are unspecified at the moment.

In practical applications we have to design methods working for many functions f and weights ρ, so we assume that they range within certain sets $f \in \mathcal{F}$ and $\rho \in \mathcal{M}$, respectively. In applications we want to keep the class of integrands large. Since we want to use QMC points, the appropriate class $\mathcal{F}$ will consist of functions, integrable and of bounded variation. The detailed definition of $\mathcal{F}$ is postponed until Section 2.

Then we address the following problems:

Problem A. Is convergence of $S_{m,\mathbf{n}}(f,\rho)$ to $\int_{\mathbb{R}^d} f(\mathbf{x})\rho(\mathbf{x})\, d\mathbf{x}$ uniform over $f \in \mathcal{F}$ and $\rho \in \mathcal{M}$?

The answer is "yes" under smoothness assumptions, when $\mathcal{M}$ is compact in $L^1(\mathbb{R}^d, d\mathbf{x})$; see Theorem 1. On bounded domains, the accuracy of approximation is determined by the discrepancy of the QMC point set P. Therefore we raise the following

Problem B. Can the method $S_{m,\mathbf{n}}(f,\rho)$ provide the accuracy for approximation of $\int_{\mathbb{R}^d} f(\mathbf{x})\rho(\mathbf{x})\, d\mathbf{x}$ as well as it does on bounded domains?

Again, the answer is "yes" conditional upon the probability density functions in $\mathcal{M}$ being dominated by some power decay function as shown in Theorem 2. The algorithm used for the constructive proof of the integral approximation theorem in [9] is similar in spirit to our method (1).

Recently, A. Papageorgiou [7] studied a similar problem, the approximate computation of integrals like $\text{Int}_\rho(f)$, but under the additional assumption that f as well as ρ depends only on $\|\mathbf{x}\|$ (isotropic integrals). In this particular situation, the original integral is reduced to one on the d-dimensional unit cube by coordinate-wise inversion. The corresponding QMC quadrature is shown to converge at a rate $\sqrt{\log(n)/n}$.

In this paper the authors propose the numerical scheme (1) using low-discrepancy point sets on nested hypercubes. In Section 2 a class $\mathcal{F}$ of admissible integrands

is extracted. In Section 3 an error analysis is carried out, which yields error bounds, uniformly for classes $\mathcal{M}$ of weight functions, having a certain prescribed decay at infinity. The number of cubes and the respectively chosen number of low-discrepancy points in each of the cubes will depend on those assumptions on the decay of the weights at ∞. In Section 4 it is shown how elliptically contoured distributions fit this setup, in particular multivariate Gaussian densities. In Section 5 we explicitly discuss the two-dimensional situation, since there the Fibonacci sequence may be used to design QMC point sets.

We close our study with a potential application to measure coherent risks, as required to evaluate in mathematical finance.

2. Preliminary analysis: Functions of bounded variation on $\mathbb{R}^d$

The analysis of algorithm (1) requires preparatory material. This extends the basic approach in the applicability of QMC point sets for integration over the cube. Our presentation focuses on the introduction of the class $\mathcal{F}$ of admissible integrands. We emphasize which problems occur when scaling domains and cutting to frames, as is done in algorithm (1). This results in a class $\mathcal{F}$, which is large enough to cover many situations, but it is not the only possible class.

We refer to [6] for notation and background details concerning usage of QMC point sets, as well as to [11] for analysis of weakly differentiable functions.

We first recall the notion of $*$-discrepancy. For any given collection P of n points $(\mathbf{x}_i)_{i=1}^n$ in $[0,1)^d$ the quantity

$$D^*(P) := \sup_{a\in(0,1]^d} \left| \frac{\#\{\mathbf{x}_i, \quad \mathbf{x}_i \in [0,a),\ i=1,\dots,n\}}{n} - a_0\cdots a_{d-1} \right|$$

denotes the *$*$-discrepancy*. We know that for any n, low-discrepancy point sets P consisting of n points can be constructed (see, e.g., [6]) such that

$$D^*(P) \le C(d)\cdot n^{-1}\log^d n.$$

Such point sets are said to be of *low discrepancy*. Our analysis will be based on Koksma-Hlawka type inequalities. This is valid for a certain class of functions on $[0,1]^d$, functions of bounded variation in the sense of Hardy and Krause.

Therefore we will briefly review some calculus of weakly differentiable functions, as is required to formulate the results. If the functions are sufficiently smooth, then the respective norm is given by

$$\mathrm{Var}(f) := \sum_{\emptyset\neq I\subseteq\{1,\dots,d\}} \left\| \frac{\partial^{|I|}}{\partial x_I} f(\mathbf{x}_I, \mathbf{1}_{I^c}) \right\|_1, \tag{2}$$

which indicates that coordinates which are not covered by I are set equal to 1. (We may and do phrase this by saying that the variation is *anchored* at $\mathbf{1}$.) Integration is on coordinates from I, thus on $[0,1)^{|I|}$. For such functions g the following variant of the *Koksma-Hlawka Inequality* is valid for any point set $P\subset[0,1)^d$:

$$\left| \int_{[0,1)^d} g(\mathbf{x})\, d\mathbf{x} - \frac{1}{n}\sum_{i=1}^n g(\mathbf{x}_i) \right| \le \mathrm{Var}(g) D^*(P).$$

Remark 1. The rôle of the anchor at $\mathbf{1}$ is rarely discussed. In the derivation of the Koksma-Hlawka Inequality as in [10], which is based on partial integration, it can be seen that any anchor coordinate-wise larger than $\mathbf{1}$ can be used. However, the

variation norm $\mathrm{Var}(\cdot)$ also depends on the anchoring. It is easy to verify that it is the smallest when dropping the anchor at the "upper right" corner of $[0,1]^d$.

In our subsequent analysis we restrict ourselves to functions g which are (locally) integrable and for any $I \subset \{1,\dots,d\}$ the weak derivatives $\frac{\partial^{|I|}}{\partial x_I} g$ exist and are integrable on $[0,1]^d$. The Koksma-Hlawka Inequality extends to such functions, in particular it holds that

$$\left| \int_{[0,1]^d} g(\mathbf{y}) \chi_F(\mathbf{y}) \, d\mathbf{y} - \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{y}_i) \chi_F(\mathbf{y}_i) \right| \leq \mathrm{Var}(g\chi_F) D^*(P), \tag{3}$$

where $Q := \{\mathbf{x},\ 1/4 \leq x_j \leq 3/4,\ j = 1,\dots,d\}$ and $F := [0,1]^d \setminus Q$. $\mathrm{Var}(g\chi_F)$ is as in (2), the appropriate measure of variation on $F \subset [0,1]^d$. In this context it is important to mention the Leibniz formula, which also extends to weak derivatives. In particular (see, e.g., [11, Ex. 1.5]), we have

$$\frac{\partial^{|I|}}{\partial x_I}(g\chi_F) = \sum_{A \cup B = I} \left(\frac{\partial^{|A|}}{\partial x_A} g \right) \left(\frac{\partial^{|B|}}{\partial x_B} \chi_F \right), \tag{4}$$

where the sum extends over all disjoint unions $A \cup B = I$. Evaluating the partial derivatives $\frac{\partial^{|B|}}{\partial x_B} \chi_F$ yields the lengthy but important bound

$$\begin{aligned} \mathrm{Var}(g\chi_F) \leq \| \frac{\partial^d}{d\mathbf{x}} g \|_{1,F} + \sum_{A \neq \emptyset} \sum_{\substack{\varepsilon_j = \pm 1, \\ j \in A}} \| \frac{\partial^{|A|}}{\partial x_A} g(\mathbf{x}_{A^c}, \varepsilon_A) \|_{1,F} \\ + \sum_{\emptyset \neq I \neq \{1,\dots,d\}} \| \frac{\partial^{|I|}}{\partial x_I} g(\mathbf{x}_I, \mathbf{1}_{I^c}) \|_{1,F}, \end{aligned} \tag{5}$$

where $\| \cdot \|_{1,F}$ means restriction of the integration domain to F and ε_A is a shortcut for $1/4(2 + \varepsilon_j)_{j \in A}$.

Remark 2. Although lengthy, this formula reveals its main features, namely,

(1) that the integration is restricted to F everywhere,
(2) that anchors are dropped only at $1/4, 3/4$ and 1 multiples of $\mathbf{1}$, and
(3) that by cutting to F, we artificially increase the variation due to the jumps. But, as long as the jumps are "along" the axes, the variation remains finite.

In the construction of (1), integration is on cubes Q_j with scaled points $\mathbf{y}_i$. A simple rescaling argument reveals that for any cube Q of the form $Q := a + q[0,1]^d$ and subset $G := a + qF$ it holds that

$$\left| \int_Q g(\mathbf{y}) \chi_G(\mathbf{y}) \, d\mathbf{y} - \frac{|Q|}{n} \sum_{i=1}^{n} g(\mathbf{y}_i) \chi_G(\mathbf{y}_i) \right| \leq |Q| \, \mathrm{Var}(g\chi_G) D^*(P). \tag{6}$$

Below, this is applied to functions $f\rho\chi_{I_j}$, the respective restrictions of f and ρ onto frames I_j. Further use of the Leibniz formula yields

$$\mathrm{Var}(f\rho\chi_{I_j}) \leq 2^d (\mathrm{Var}(f\chi_{I_j}) + \|f\chi_{I_j}\|_1) C(\rho\chi_{I_j}), \tag{7}$$

where

$$C(\rho\chi_G) := \max \left\{ \sup_{\mathbf{x} \in G} \left| \frac{\partial^{|I|}}{\partial x_I} \rho(\mathbf{x}) \right|,\ I \subseteq \{1,\dots,d\} \right\}. \tag{8}$$

This leads to the following definition of the class $\mathcal{F} = \mathcal{F}(V)$ of admissible integrands: A weakly integrable function f belongs to $\mathcal{F}$, if for all m and decompositions of $\mathbb{R}^d$ into frames $I_0, \ldots, I_m$ the quantity

$$\|f\|_\infty + \sum_{j=0}^{m} \Big(\mathrm{Var}(f\chi_{I_j}) + \|f\chi_{I_j}\|_1 \Big) \le V < \infty \tag{9}$$

is bounded.

Remark 3. Since the decompositions are disjoint, $\sum_{j=0}^{m} \|f\chi_{I_j}\|_1$ actually adds up to $\|f\|_1$. This cannot be done for the first summands, since the computation of $\mathrm{Var}(f\chi_{I_j})$ involves anchors at different locations, depending on j; see Remark 2 above. This makes the definition of $\mathcal{F}$ tied to the decomposition. An easily verifiable sufficient condition, which is not tied to these decompositions, is provided by the subset $\bar{\mathcal{F}}(C, R) \subset \mathcal{F}$, defined to contain weakly differentiable functions f for which

$$\left| \frac{\partial^{|A|}}{\partial x_A} f(\mathbf{x}) \right| \le C \frac{1}{\|x\|_\infty^{|A|}}, \quad \text{for } \|x\|_\infty \ge R, \tag{10}$$

for appropriate constants C and R.

Remark 4. If a function vanishes outside the unit cube Q_0, then the quantity in (9) reduces to $\|f\|_\infty + \mathrm{Var}(f) + \|f\|_1$. Therefore, it covers the situation of functions of bounded variation in the sense of Hardy and Krause on the unit cube.

As already stressed in Remark 2, algorithm (1) requires that the artificially introduced jump sizes decay fast enough.

3. Error bounds

In this section we investigate the potential quality of schemes (1), based on QMC point set. First we shall establish that for relatively compact subsets $\mathcal{M}$ of weight functions, convergence is uniform.

In a second step we derive the speed of convergence, provided we have some specified decay of the weights.

The basic error decomposition is presented in

Proposition 1. *For any m and $\mathbf{n} = (n_0, \ldots, n_m)$ the following bound holds true for any function from $\mathcal{F}(V)$ and point sets P_j used in cube Q_j:*

$$\begin{aligned} &|\mathrm{Int}_\rho(f) - S_{m,\mathbf{n}}(f,\rho)| \\ &\qquad \le 2^d V \max_{0 \le j \le m} \Big\{ q_j D^*(P_j) C(\rho\chi_{I_j}) \Big\} + \int_{Q_m^c} |f(\mathbf{x})\rho(\mathbf{x})| \; d\mathbf{x}, \end{aligned} \tag{11}$$

where q_j is the volume of Q_j and $D^*(P_j)$ is the discrepancy.

Proof. Using the estimates derived in Section 2, we conclude

$$|\mathrm{Int}_\rho(f) - S_{m,\mathbf{n}}(f,\rho)| = \left| \int_{\mathbb{R}^d} f\rho \, d\mathbf{y} - \sum_{j=0}^{m} \frac{q_j}{n_j} \sum_{i=1}^{n_j} f(\mathbf{y}_{ij})\rho(\mathbf{y}_{ij})\chi_{I_j}(\mathbf{y}_{ij}) \right|$$

$$\le \sum_{j=0}^{m} \Big| \int_{Q_j} f\rho\chi_{I_j} \, d\mathbf{x} - \frac{q_j}{n_j} \sum_{i=1}^{n_j} f(\mathbf{y}_{ij})\rho(\mathbf{y}_{ij})\chi_{I_j}(\mathbf{y}_{ij}) \Big| + \int_{Q_m^c} |f\rho| \, d\mathbf{y}$$

$$\le \sum_{j=0}^{m} q_j D^*(P_j) \mathrm{Var}(f\rho\chi_{I_j}) + \int_{Q_m^c} |f\rho| \, d\mathbf{x} \tag{12}$$

$$\le 2^d \sum_{j=0}^{m} q_j D^*(P_j) C(\rho\chi_{I_j}) \Big(\mathrm{Var}(f\chi_{I_j}) + \|f\chi_{I_j}\|_1 \Big) + \int_{Q_m^c} |f\rho| \, d\mathbf{x} \tag{13}$$

$$= 2^d \max_{0\le j\le m} \{q_j D^*(P_j) C(\rho\chi_{I_j})\} \sum_{j=0}^{m} \Big(\mathrm{Var}(f\chi_{I_j}) + \|f\chi_{I_j}\|_1 \Big) + \int_{Q_m^c} |f\rho| \, d\mathbf{x}$$

$$\le 2^d V \max_{0\le j\le m} \{q_j D^*(P_j) C(\rho\chi_{I_j})\} + \int_{Q_m^c} |f\rho| \, d\mathbf{x}.$$

Above, we made use of (6) to derive (12) and (7) to get (13), respectively. □

3.1. **Uniformity of approximation for compact sets of weights.** Here we will discuss Problem A.

The following theorem asserts that we can approximate $\mathrm{Int}_\rho(f)$ uniformly by methods of type (1), if the weight functions belong to a set, compact in $L^1(\mathbb{R}^d, d\mathbf{x})$. For convenience we recall that as a consequence of the Weyl criterion of compactness for a certain class $\mathcal{M} \subset L^1(\mathbb{R}^d, d\mathbf{x})$ we have that for all $\varepsilon > 0$ there is a bounded set A such that $\int_{A^c} \rho(\mathbf{x}) \, d\mathbf{x} \le \varepsilon$, uniformly for $\rho \in \mathcal{M}$; see [2, Chaper 4.20].

Theorem 1. *Suppose that $\mathcal{M}$ is relatively compact in $L^1(\mathbb{R}^d, d\mathbf{x})$. If*

$$\sup_{\rho\in\mathcal{M}} \{C(\rho) + \|\rho\|_\infty\} < \infty, \tag{14}$$

then there are a sequence $m(N)$ and numbers $n_j = n_j(N)$, $j = 0, \dots, m(N)$, satisfying $N = \sum_{j=0}^{m(N)} n_j$, such that for any family of low-discrepancy point sets $\{P_j,\ j = 0, 1, 2, \dots, m(N)\}$ consisting of n_j points, we have convergence

$$\sup_{f\in\mathcal{F}} \sup_{\rho\in\mathcal{M}} |\mathrm{Int}_\rho(f) - S_{m,\mathbf{n}}(f,\rho)| \to 0, \qquad \text{as } N \to \infty.$$

Proof. We first use the basic error estimate (11) to see that the assertion is true, if

$$\max_{0\le j\le m} \Big\{ q_j D^*(P_j) C(\rho\chi_{I_j}) \Big\} \to 0 \tag{15}$$

and

$$\int_{Q_m^c} \rho(\mathbf{x}) \, d\mathbf{x} \to 0 \tag{16}$$

uniformly for $\rho \in \mathcal{M}$. By assumption (14), for any given m the left-hand side in (15) can be made as small as desired by choosing the number of points in P_j large enough. The remainder term in (16) can be made small, uniformly in ρ by choosing m large enough. This is a consequence of compactness of $\mathcal{M}$, as explained above. □

3.2. **Weights with prescribed decay at infinity.** In this section we formulate the main result about the rate of convergence for method (1) by a proper choice of the parameters, uniformly for $f \in \mathcal{F}$ and weights from a certain class $\mathcal{M}_s(R)$. This will result in a quantitative version of Theorem 1.

A weight ρ belongs to $\mathcal{M}_s(R)$, where $s > 0$ is fixed, if all derivatives $\frac{\partial^{|I|}}{\partial x_I}\rho$, $I \subset \{1, \dots, d\}$ exist, are uniformly bounded and obey for $t \geq 1$ the estimate

$$\sup_{\|\mathbf{x}\|>t} \left| \frac{\partial^{|I|}}{\partial x_I}\rho(\mathbf{x}) \right| \leq Rt^{-s}, \tag{17}$$

where $\|\cdot\|$ is the Euclidean one for definiteness. In Section 4 we will establish that certain sets of elliptically contoured distributions are covered. So far we just stick to this technical definition. It is easy to verify that for $s > d$ the classes $\mathcal{M}_s(R)$ are relatively compact in $L^1(\mathbb{R}^d, d\mathbf{x})$. In particular the following estimate will be useful.

Lemma 1. *There is $\tilde{C}(d)$ such that*

$$\sup_{\rho\in\mathcal{M}_s(R)} \int_{\|\mathbf{x}\|>t} |\rho(\mathbf{x})| \; d\mathbf{x} \leq \tilde{C}(d)R \max\left\{1, \frac{1}{s-d}\right\} \; t^{d-s}, \qquad t \geq 1. \tag{18}$$

Proof. For $t \geq 1$ and $\rho \in \mathcal{M}_s(R)$ we can bound

$$\begin{aligned}
\int_{\|\mathbf{x}\|>t} |\rho(\mathbf{x})| \; d\mathbf{x} &= \sum_{j=0}^{\infty} \int_{2^j t<\|\mathbf{x}\|\leq 2^{j+1}t} |\rho(\mathbf{x})| \; d\mathbf{x} \\
&\leq R\sum_{j=0}^{\infty} \int_{2^j t<\|\mathbf{x}\|\leq 2^{j+1}t} \left(2^j t\right)^{-s} \; d\mathbf{x} \\
&\leq R\sum_{j=0}^{\infty} \left(2^j t\right)^{-s} \left(2^{j+1}t\right)^d \int_{\|\mathbf{x}\|\leq 1} d\mathbf{x}.
\end{aligned}$$

The last integral above can be bounded by 6 independent of the dimension d, such that we finally arrive at

$$\sup_{\rho\in\mathcal{M}_s(R)} \int_{\|\mathbf{x}\|>t} |\rho(\mathbf{x})| \; d\mathbf{x} \leq 6R2^d t^{d-s} \sum_{j=0}^{\infty} \left(2^{d-s}\right)^j,$$

from which it is easy to accomplish the proof with $\tilde{C}(d) = 12 \cdot 2^d$. □

Based on this estimate, we can identify m for which the remainder $\sup_{\rho\in\mathcal{M}} \int_{Q_m^c} \rho(\mathbf{x}) \, d\mathbf{x}$ outside cubes of length 2^{m+1} has prescribed size.

Corollary 1. *Given N large enough, we let $m := m(N) = \lceil \frac{1}{s-d} \log_2 N \rceil$. Then*

$$\int_{Q_m^c} |\rho(\mathbf{x})| \; d\mu(\mathbf{x}) \leq \tilde{C}(d)R \max\left\{1, \frac{1}{s-d}\right\} \; d^{(d-s)/2} \frac{1}{N}. \tag{19}$$

This follows just by observing that $Q_m \subset \left\{\mathbf{x}, \; \|\mathbf{x}\| \leq \sqrt{d}2^m\right\}$ and inserting the respective value of m into the estimate (18).

We turn to the main result.

Theorem 2. *Let* $m := m(N) = \lceil \frac{1}{s-d} \log_2 N \rceil$ *and choose*

$$n_j = \lceil N \frac{2^{-j(s-d)}}{\sum_{j=0}^{m} 2^{-j(s-d)}} \rceil, \quad j = 0, \dots, m. \tag{20}$$

The overall number of points used by (1) *is bounded by* CN*, for some constant* C*, depending only on* s *and* d*.*

For any low-discrepancy point sets $\{P_j,\ j = 0, 1, 2, \dots, m\}$ *consisting of* n_j *points, we have*

$$\sup_{f \in \mathcal{F}} \sup_{\rho \in \mathcal{M}_s(R)} |\mathrm{Int}_\rho(f) - S_{m,\mathbf{n}}(f, \rho)| \le C(V, R, s, d)\ N^{-1} \log^d N. \tag{21}$$

Proof. The overall number of points will not equal N, due to the ceiling for the n_j. But it can be bounded by $N + m \le C(s, d)N$. Furthermore, by the above choice of m, Corollary 1 asserts that we can bound the remainder by

$$\int_{Q_m^c} |f\rho|\ d\mathbf{x} \le C \frac{1}{N},$$

for some constant $C = C(V, R, s, d)$.

It remains to bound $\max_{0 \le j \le m} \{C(\rho \chi_{I_j}) q_j D^*(P_j)\}$. First, the denominator in (20) is bounded from above by a constant depending on s, d. By the definition of $\mathcal{M}_s(R)$ we have $C(\rho \chi_{I_j}) \le R 2^{-js}$. By the choice of m and since $q_j = 2^{(j+1)d}$, we arrive at

$$C(\rho \chi_{I_j}) q_j D^*(P_j) \le C(d) 2^{(j+1)d} R \cdot 2^{-js} \frac{1}{n_j} \log^d n_j \le C(R, s, d) \frac{1}{N} \log^d N.$$

Hence we have

$$|\mathrm{Int}_\rho(f) - S_{m,\mathbf{n}}(f, \rho)| \le V \left\{ C(s, d, R) \frac{1}{N} \log^d N + C(V, R, s, d) \frac{1}{N} \right\}.$$

This completes the proof. □

Remark 5. As already mentioned in the Introduction, method (1) is of optimal order by the above choice of parameters, since the rate in (21) cannot be beaten even on finite domains; see Remark 4.

Remark 6. As can be seen from Figure 2 (in Section 5), the hierarchical QMC point sets are stacked adaptively to the radial behaviour of the weight, such that the error bound order is optimized to $N^{-1} \log^d N$. It can be shown readily that this order cannot be achieved, when only one cube is used instead of the hierarchy structure.

4. Elliptically contoured distributions as weights

As an important class of weight functions we introduce the following $\mathcal{M}(s, B, \Lambda_0, \Lambda_1)$, determined by the following restrictions: The weights $\rho(\mathbf{x})$ are elliptically contoured probability density functions

$$\rho(\mathbf{x}) = |\Sigma|^{-\frac{1}{2}} g\left(\frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{x} \right),$$

where Σ is positive definite with eigenvalues bounded as

$$\infty > \Lambda_1^d \ge |\Sigma| \ge \Lambda_0^d > 0, \tag{22}$$

and $g \in C^d(\mathbb{R}_+)$ is bounded, decreasing and, apart from a neighbourhood of 0, takes the power decaying form:

$$g(y) = By^{-\frac{s}{2}}, \quad y > \varepsilon_0 > 0.$$

Remark 7. These restrictions are quite natural in the following sense. The distribution with ECD pdf ρ has covariance matrix proportional to Σ. So if there is no lower bound for the eigenvalues of Σ, then the weight function could be close to singular. In that situation the problem should be preconditioned before any algorithm is applicable or meaningful. On the other hand, if there is no upper bound for the eigenvalues of Σ, the huge variance of the distribution would make the weight function meaningless or just impossible to bound the cut-off error so as to practice the QMC scheme in a bounded domain.

Again, there is another natural implicit restriction: $s > d$. This is because after the spherical transformation the radial pdf is proportional to r^{d-s-1}, which is integrable iff $s > d$.

In this section we aim at showing that $\mathcal{M}(s, B, \Lambda_0, \Lambda_1) \subset \mathcal{M}_s(R)$ for R large enough, such that the results from Section 3 hold true. The required smoothness is easily derived from the smoothness of g. So it remains to show that the decay is as in (17). This is done in

Theorem 3. *There is a constant A, depending only on the parameters $(\Lambda_0, \Lambda_1, d, s)$, such that uniformly for $I \subset \{1, 2, \ldots, d\}$ we have*

$$\sup_{\|\mathbf{x}\| > t} ||\frac{\partial^{|I|}}{\partial \mathbf{x}_I} \rho(\mathbf{x})||_\infty \leq At^{-s}, \qquad t \geq 1,$$

To this end we need to address some properties of the elliptically contoured distributions. For simplicity we introduce the notations:

$$\begin{aligned}
\Sigma^{-1} &= (S_{ij})_{d \times d}, \\
L^{(0)}_{ij} &= S_{ij}, \qquad i, j = 1, 2, \ldots, d, \\
L^{(1)}_{i} &= \sum_{j=0}^{d} S_{ij} x_j, \qquad i = 1, 2, \ldots, d, \\
L^{(k)}_{i_1, \ldots, i_k} &= \prod_{h=1}^{k} L^{(1)}_{i_h}, \qquad \{i_1, \ldots, i_k\} \subset \{1, 2, \ldots, d\}.
\end{aligned}$$

Lemma 2. *For any $0 \leq k \leq d$,*

$$\frac{\partial^k \rho}{\partial x_1 \cdots \partial x_k} = |\Sigma|^{-\frac{1}{2}} \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \sum_{\mathbb{P}^k_{2i}} g^{(k-i)} L^{(0)}_{(l_1, l_2)} \cdots L^{(0)}_{(l_{2i-1}, l_{2i})} L^{(d-2i)}_{\mathcal{L}_k \setminus \{l_1, \ldots, l_{2i}\}},$$

where $\mathbb{P}^k_{2i}$ is all the possible partitions of $\mathcal{L}_k = \{1, 2, \ldots, k\}$ into i pairs $\{(l_1, l_2), \ldots, (l_{2i-1}, l_{2i})\}$ and a remainder subset.

Proof. The formula is obtained by elementary induction and the basic fact that the partitions of $\mathcal{L}_{k+1}$ into i pairs and a remainder subset can be characterized by whether the particular element $k + 1$ is in a pair set or in the remainder set. Thus the partitions can be related to $\mathcal{L}_k$ recursively. □

Proof of Theorem 3. First we notice that

$$\begin{aligned} |L_{ij}^{(0)}| &= |S_{ij}| \qquad (\text{the } ij(\text{th}) \text{ element of } \Sigma^{-1}) \\ &\le \Lambda_0^{-1}; \\ |L_i^{(1)}| &= |\sum_{j=0}^{d} S_{ij} x_j| \le (\sum_{j=0}^{d} S_{ij}^2)^{\frac{1}{2}} ||\mathbf{x}|| \\ &\le (d\Lambda_0^{-2})^{\frac{1}{2}} ||\mathbf{x}|| = \sqrt{d}\Lambda_0^{-1} ||\mathbf{x}||; \\ |L_{\{1,2,\dots,k\}}^{(k)}| &\le d^{\frac{k}{2}} \Lambda_0^{-k} ||\mathbf{x}||^k. \end{aligned}$$

Another fact is that, apart from a neighbourhood of 0, for $0 \le k \le d$,

$$\begin{aligned} g(y) &= B(y)^{-\frac{s}{2}}, \\ g^{(k)}(y) &= B(-1/2)^k s(s+2)\cdots[s+2(k-1)] y^{-\frac{s}{2}-k}, \\ g\left(\frac{1}{2}r^2\right) &= B 2^{\frac{s}{2}} r^{-s}, \\ g^{(k)}\left(\frac{||\mathbf{x}||^2}{2\Lambda_1}\right) &= (-1/2)^k B 2^{\frac{s}{2}} s(s+2)\cdots[s+2(k-1)] ||\mathbf{x}||^{-s-2k} \Lambda_1^{\frac{s}{2}+k}. \end{aligned}$$

So the term among $\{||\mathbf{x}||^k g^{(k)}(||\mathbf{x}||^2/(2\Lambda_1)) : \ k = 0, 1, 2, \dots, d\}$ with the slowest decaying power is g itself, namely, $||\mathbf{x}||^{-s}$. Thus from Lemma 2 we see that

$$\begin{aligned} \left|\frac{\partial^k \rho}{\partial x_1 \cdots \partial x_k}\right| &\le |\Sigma|^{-\frac{1}{2}} \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \sum_{\mathbb{P}_{2i}^k} |g^{(k-i)}| |L_{(l_1,l_2)}^{(0)} \cdots L_{(l_{2i-1},l_{2i})}^{(0)} L_{\mathcal{L}_k \setminus \{l_1,\dots,l_{2i}\}}^{(d-2i)}| \\ &\le \Lambda_0^{-\frac{d}{2}} |\mathbb{P}_{2i}^k| \lfloor \frac{k+1}{2} \rfloor \max_{0 \le i \le \lfloor \frac{k}{s} \rfloor} \left| g^{(k-i)}\left(\frac{||\mathbf{x}||^2}{2\Lambda_1}\right)\right| d^{\frac{k-2i}{2}} \Lambda_0^{-(k-i)} ||\mathbf{x}||^{(k-2i)} \\ &\le A ||\mathbf{x}||^{-s}. \end{aligned}$$

Here A depends upon only the parameters $(\Lambda_0, \Lambda_1, d, s)$. □

4.1. **Multivariate normal distributions.** Probably most important is the case of multivariate normal distributions. Let us introduce $\mathcal{N}(\Lambda_0, \Lambda_1)$, the class of all multivariate normal distributions with covariances Σ obeying (22). Plainly, this class belongs to $\mathcal{M}_s(R)$ for every $s > d$ and appropriate R. Thus the convergence rate presented in Theorem 2 can be achieved. But, due to the rapid decay of the weight, this goal can be achieved with fewer cubes in the hierarchy. We formulate this in

Corollary 2. *Let* $m := \lfloor \frac{1}{2} \log\log N \rfloor$ *and* $n_j := \lceil N \frac{e^{-2^{2j-1}}}{\sum_{j=0}^{m} e^{-2^{2j-1}}} \rceil, \quad j = 0, \dots, m.$ *Then there is* C *such that*

$$\sup_{f \in \mathcal{F}} \sup_{\rho \in \mathcal{N}(\Lambda_0, \Lambda_1)} |\mathrm{Int}_\rho(f) - S_{m,\mathbf{n}}(f, \rho)| \le C \ N^{-1} \log^d N.$$

Remark 8. When restricted to normal weight functions, we have fewer boxes $m = \frac{1}{2}\log\log N + O(1)$ as compared with $\frac{1}{s-d}\log N + O(1)$ for the general situations. Also the choices of n_j for each hypercube are evaluated as the portion of the partial sum

of exponential terms rather than in power terms as shown in the proof of Theorem 2. It should also be mentioned that for the normal weight integral we may even choose to use only one hypercube Q_m to apply the QMC algorithm. The error bound would then have the convergence order of $N^{-1}\log^{d+\frac{d}{2}} N$, an interesting comparison with the hierarchical algorithm suggested by Theorem 2; see also Remark 6.

5. $d = 2$: Fibonacci numbers

In two dimensions we can specify the algorithm using the Fibonacci sequence $(F(j))_{j=1}^{\infty}$, generated by $F(j+2) := F(j+1) + F(j),\ j = 1, \dots,$ with initial conditions $F(1) = F(2) = 1$, hence $(1, 1, 2, 3, 5, 8, 13, \dots)$.

Indeed, a look at the determination of the numbers n_j of points used in the cubes Q_j reveals that the quotients behave like

$$\frac{n_j}{n_{j+1}} \sim 2^{s-d}, \tag{23}$$

where base 2 results from the prescribed growth of the widths of the cubes by steps 2. If this is replaced by some $w > 1$, the relation (23) has to be replaced by $\frac{n_j}{n_{j+1}} \sim w^{s-d}$. In two spatial dimensions we can employ this to design our algorithm using Fibonacci points. Namely, if we let $\gamma := \frac{1}{2}(1+\sqrt{5})$ and choose $w := \gamma^{1/(s-2)}$, then the number of points to be chosen in cube Q_j is to satisfy $n_j/n_{j+1} \sim \gamma$. This is convenient, since successive Fibonacci numbers $F(j),\ F(j+1)$ satisfy $F(j+1)/F(j) \to \gamma$ as $j \to \infty$. Convergence is rapid, and a good approximation is already obtained for $j \geq 10$.

Moreover, how to construct low-discrepancy points in two dimensions with exactly $F(j)$ points is well known. As shown in Hua and Wang [3], given $j > 1$, the following point set $\{(x_i, y_i),\ i = 0, \dots, F(j)\}$, given, for $i = 1, 2, \dots, F(j)$, by

$$\begin{aligned} x(i) &= (2 * i - 1)/(2 * F(j)), \\ y(i) &= [2 * (i * F(j-1) \bmod F(j)) - 1]/(2 * F(j)), \end{aligned}$$

has asymptotically the optimal discrepancy of order $\frac{1}{F(j)} \log F(j)$.

Summarizing,

(1) we first fix a certain initial volume V^2;
(2) we choose m according to Theorem 2 and let the cubes $Q_j,\ j = 0, \dots, m$, be chosen with widths $w_j := \gamma^{j/(s-2)} V,\ j = 0, \dots, m$;
(3) we then discard an initial segment of length l from the Fibonacci sequence and let $n_j := F(l + m - j),\ j = 0, \dots, m$, i.e., in reverse order (if l is large enough ($l \geq 10$), then $n_j/n_{j+1} \approx \gamma$ with high accuracy);
(4) we use algorithm (1) with the above parameters and low-discrepancy points based on the Fibonacci numbers.

We exhibit this approach in the following examples.

We consider the ECD family with probability density function

$$\rho(x, y) := \frac{1}{|\Sigma|^{\frac{1}{2}}} g\left(\frac{1}{2}(x, y)\Sigma^{-1}(x, y)'\right), \qquad \Sigma > 0,$$

where

$$g\left(\frac{1}{2}r^2\right) := \frac{s-2}{4\pi}\left(1 + \frac{1}{2}r^2\right)^{-\frac{s}{2}}, \qquad s > 2.$$

Figure 1 shows the situation with $s = 4$, $\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 4.0 & 1.9 \\ 1.9 & 1.0 \end{bmatrix}$, respectively. Particularly we evaluate the following integral as a demonstration:

$$I = \int_{\mathbb{R}^2} \left(1 + (x,y)\Sigma^{-1}(x,y)'\right)^{-1} \frac{1}{|\Sigma|^{\frac{1}{2}}} g\left(\frac{1}{2}(x,y)\Sigma^{-1}(x,y)'\right) dx\, dy,$$

where $s = 4$.

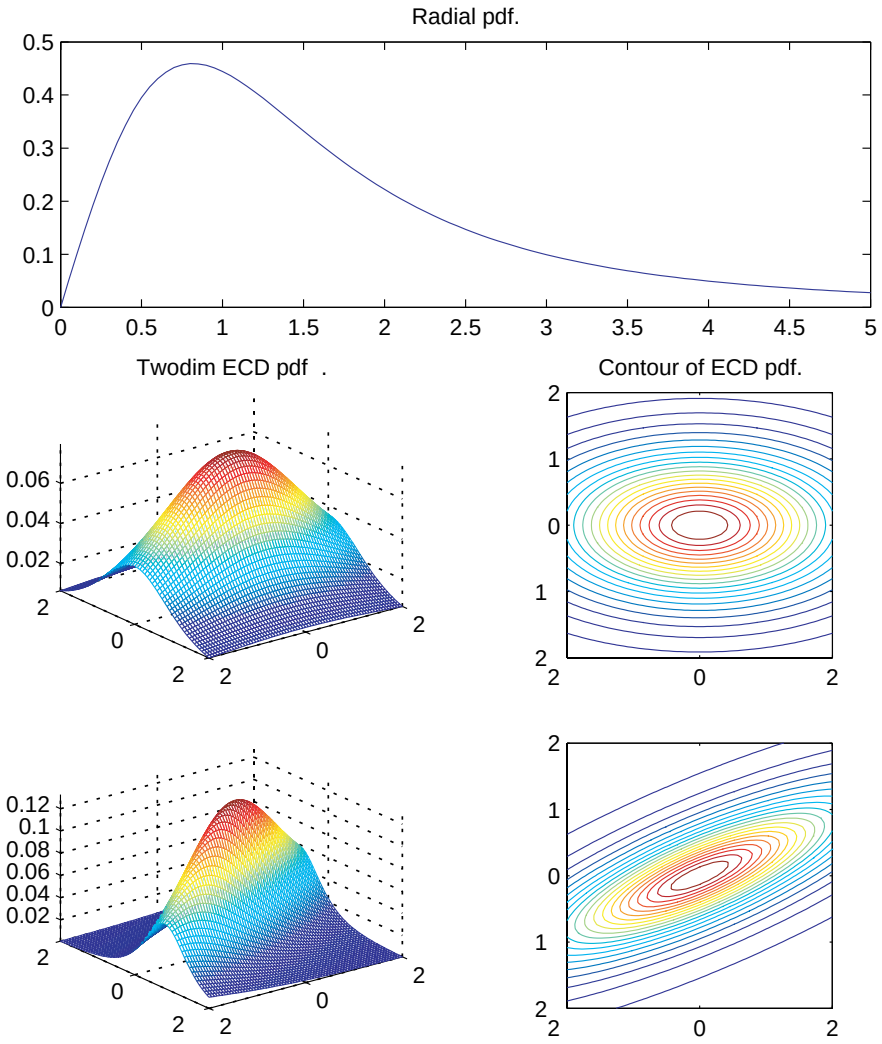

FIGURE 1. A family of ECD densities.

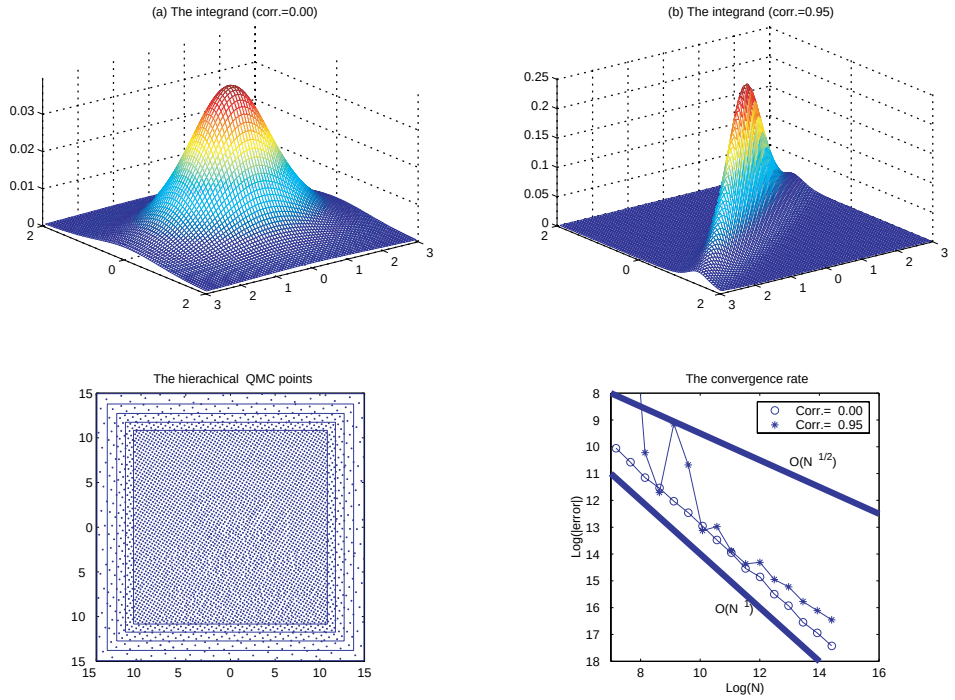

FIGURE 2. Numerical example of the hierarchical QMC scheme.

The exact value of I is $2\log 2 - 1$. By taking $V = 20$, $l = 10$ and $m = 6, 7, \ldots, 21$, respectively, we obtained the empirical convergence rates shown in Figure 2. The correlation effect is obvious when the total number of QMC points is small but would not hinder the optimality of our algorithm in general.

6. Evaluation of coherent risks

There is some "by-product", which actually motivated the present study. Let us consider for nonnegative f the quantities

$$L_{\mathcal{M}}(f) := \sup\{-\mathrm{Int}_\rho(f) : \quad \rho \in \mathcal{M}\}. \tag{24}$$

If we analogously let

$$L_{m,\mathbf{n}}(f) := \sup\{-S_{m,\mathbf{n}}(f,\rho) : \quad \rho \in \mathcal{M}\}, \tag{25}$$

then the triangle inequality implies

$$|L_{\mathcal{M}}(f) - L_{m,\mathbf{n}}(f)| \le \sup\{|\mathrm{Int}_\rho(f) - S_{m,\mathbf{n}}(f,\rho)|, \quad \rho \in \mathcal{M}\}. \tag{26}$$

In this context Theorem 1 translates to

Corollary 3. *For relatively compact $\mathcal{M}$ there are a sequence $m(N)$ and numbers $n_j = n_j(N)$, $j = 0, \ldots, m(N)$, satisfying $N = \sum_{j=0}^{m(N)} n_j$, such that for any family of QMC point sets $\{P_j,\ j = 0, 1, 2, \ldots, m(N)\}$ consisting of n_j points, we have*

$$|L_{\mathcal{M}}(f) - L_{m,\mathbf{n}}(f)| \to 0 \quad \text{as } N \to \infty.$$

We emphasize that in contrast to the evaluation of $L_{\mathcal{M}}$, the evaluation of $L_{m,\mathbf{n}}$ is actually an optimization over a finite set of points. Indeed, the discretization $L_{\mathcal{M}}$ to $L_{m,\mathbf{n}}$ can be seen as moving from continuous weights ρ to discrete mixtures $\sum_{i\in J} c_i\rho(\mathbf{y}_i)$. Since $L_{m,\mathbf{n}}$ is an optimization problem with linear target function, its maximal value is attained at some extreme point of $\{\rho(\mathbf{y}_i)\}\,,\ i\in J$. This shows that the computation of $L_{m,\mathbf{n}}$ reduces to a linear program.

The quantity $L_{\mathcal{M}}(f)$ forms a *coherent risk measure* for f, if f is assumed to be a nonnegative uniformly bounded function on $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$. More generally, coherent risk measure can be characterized by convex, $\sigma(\mathbf{ba},L^\infty)$-closed subsets $\mathbb{M}$ of bounded finitely additive measures on $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$; see, e.g., [1, Thm. 2.2]. There are various widely accepted examples for coherent risk measures, discussed in [1] and we also refer to [5], among them the case of point distributions, where $\mathcal{M}\subset\{\delta_\omega,\ \omega\in\Omega\}$ can be identified with a convex polyhedron in $\mathbb{R}^d$, such that the computation of (24) reduces to a linear program.

For these "accepted" risk measures there is always a different representation, which makes its approximate computation easier than solving the optimization problem. There is however a scenario where this setup appears naturally: Suppose that we have to evaluate a portfolio, say f, but we are not sure about the distribution ρ of the underlying risk, such that we want to know how stable any computed value is with respect to the prior. Then $L_{\mathcal{M}}(f)$ may be understood as a confidence bound.

Within this context we can establish that for subsets $\mathcal{M}$ compact in $L^1(\mathbb{R}^d,d\mathbf{x})$, coherent risk measures can be approximated with arbitrary accuracy by quantities $L_{m,\mathbf{n}}(f)$, which leads to a finite dimensional optimization procedure; see Corollary 3. For further applications to mathematical finance, we refer to A. Papageorgiou [8].

Acknowledgment

The discussion in Section 2 was initiated by G. Wasilkowski, Kentucky. The answer given there is probably not the only one possible and there is room for extensions and modifications.

References

1. F. Delbaen. Coherent risk measures on general probability spaces. `http://www.math.ethz.ch/~delbaen/`, 1998.
2. R. E. Edwards. *Functional Analysis. Theory and Applications*. Holt, Rinehart & Winston, New York, Chicago, 1965.
3. L.K. Hua and Y. Wang *Applications of Number Theory to Numerical Analysis*. Springer-Verlag, Berlin, New York, 1981. MR **83g:**10034
4. Norbert Hofmann and Peter Mathé. On quasi–Monte Carlo simulation of stochastic differential equations. *Math. Comp.*, 66(218):573–589, 1997. MR **97f:**65005
5. Stefan Jaschke and Uwe Küchler. Coherent risk measures and good-deal bounds. *Finance Stoch.*, 5(2):181–200, 2001. MR **2003a:**91066
6. H. Niederreiter. *Random Number Generation and Quasi–Monte Carlo Methods*, volume 63 of *CBMS–NSF Region. Conf. Series in Appl. Math.*, SIAM, Philadelphia, 1992. MR **93h:**65008
7. A. Papageorgiou. Fast convergence of quasi–Monte Carlo for a class of isotropic integrals. *Math. Comp.*, 70(233):297–306, 2001. MR **2001f:**65008
8. ———. `http://www.cs.columbia.edu/~ap/html/finder.html`.
9. L. Plaskota, K. Ritter and G. W. Wasilkowski. Average case complexity of weighted integration and approximation over $\mathbb{R}^d$ with isotropic weight. In *Monte Carlo and quasi–Monte Carlo*

Methods 2000 (K.-T. Fang, F. J. Hickernell, and H. Niederreiter, eds.), pp. 446–459, Springer-Verlag, Berlin, 2002.

10. I. M. Sobol′. An exact bound for the error of multivariate integration formulas for functions of classes $\tilde{W}_1$ and $\tilde{H}_1$. *Ž. Vyčisl. Mat. i Mat. Fiz.*, 1:208–216, 1961. MR **24:**B2546
11. William P. Ziemer. *Weakly differentiable functions*. Springer-Verlag, New York, 1989. MR **91e:**46046

Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, D-10117 Berlin, Germany
E-mail address: mathe@wias-berlin.de

Department of Mathematics, Hong Kong Baptist University, Kowloon, Hong Kong
E-mail address: gwei@math.hkbu.edu.hk