L_2 DISCREPANCY AND MULTIVARIATE INTEGRATION

ERICH NOVAK AND HENRYK WOŹNIAKOWSKI

Dedicated to Klaus F. Roth on the occasion of his 80th birthday

Abstract. We study the L_2 discrepancy of measurable subsets $B(t)$ of $\mathbb{R}^d$ for $t \in D$ with $D \subseteq \mathbb{R}^{\tau(d)}$ for some integer $\tau(d)$. We assume that t is distributed according to the density ρ, $\int_D \rho(t)\,\mathrm{d}t = 1$. For specific cases of $B(t)$, we obtain various discrepancies including probably the most celebrated discrepancy anchored at zero studied by Roth, or its analogue anchored at α, the quadrant discrepancy anchored at α, the unanchored discrepancy, and the periodic ball discrepancy. We show that the L_2 discrepancy for sets $B(t)$ corresponds to multivariate integration for the Hilbert space with some reproducing kernel K_d related to B and ρ. In particular, the L_2 discrepancy of arbitrary n points in the d dimensional case is the same as the worst case error of a linear algorithm that uses these n points and approximates the integral of d variate functions from the unit ball of the space $H(K_d)$.

We survey a number of bounds on the weighted L_2 discrepancy anchored at 0 with the emphasis on the dependence on n and d. We present necessary and sufficient conditions on the weights to guarantee that the weighted L_2 discrepancy does not depend exponentially on d.

1. Introduction

The notion of *discrepancy* goes back to the work of Weyl [66] in 1916 and van der Corput [13, 14] in the 1930s. Discrepancy is a quantitative measure of the uniformity of the distribution of points in the d dimensional Euclidean space. Today we have various notions of discrepancy, and there are literally thousands of papers studying different aspects of discrepancy. Research on discrepancy is very intensive, and the reader is referred to the recent books [4, 6, 19, 34, 37, 49, 54]. Various notions of discrepancy are widely used and studied in many areas such as number theory, approximation, stochastic analysis, combinatorics, ergodic theory and numerical analysis. The notions of discrepancy are related to Sobolev spaces, Wiener measure, VC dimension and Ramsey theory; see [4, 6, 19, 20, 32, 34, 37, 38, 52, 53, 67].

To keep this paper relatively short, we limit ourselves to L_2 discrepancy leaving the case of L_p discrepancy for a general $p \in [1, \infty]$ untouched. The case $p = \infty$ corresponds to the star discrepancy and is usually the most challenging.

It is known that many standard L_2 discrepancies are related to multivariate integration over specific reproducing kernel Hilbert spaces. We show that this relation holds for a more general class of B-discrepancies. We define the B-discrepancy as the L_2 discrepancy for measurable subsets $B(t)$ of $\mathbb{R}^d$. Here t belongs to a set $D \subseteq \mathbb{R}^{\tau(d)}$ and is distributed according to some density ρ, $\int_D \rho(t)\,\mathrm{d}t = 1$. Under natural assumptions on $B(t)$, see (2) and (26), we prove that the B-discrepancy of arbitrary n points is the same as the worst case error of a linear algorithm using the same points for multivariate integration defined for the Hilbert space $H(K_d)$ with the reproducing kernel K_d given by the formula

$$K_d(x, y) = \int_D 1_{B(t)}(x) 1_{B(t)}(y) \rho(t)\,\mathrm{d}t, \quad x, y \in D_d := \bigcup_{t \in D} B(t). \tag{1}$$

For specific choices of $B(t)$ we obtain the L_2 discrepancy anchored at $\alpha \in [0,1]^d$ which for $\alpha = 0$ was studied by Roth [47, 48] and many others, the quadrant discrepancy anchored at $\alpha \in [0,1]^d$ studied for the L_∞ norm by Hickernell, Sloan and Wasilkowski [26], the unanchored discrepancy proposed by Morokoff and Caflisch in [36] and studied in [39], and finally the ball and periodic ball discrepancy studied by Beck, Chen, Montgomery and Travaglini, see [3, 5, 10, 35, 56].

In particular, the L_2 discrepancy anchored at 0 is related to multivariate integration for the Sobolev space anchored at 1 with the reproducing kernel given by

$$K_d(x,y) = \prod_{k=1}^{d} (1 - \max\{x_k, y_k\}), \quad x, y \in [0,1]^d.$$

We are not sure who first realized this relation but this result can be already easily deduced from Hlawka's identity [29] from 1961, see also Zaremba's paper [69] from 1968.

It is also possible to show that the L_2 discrepancy for sets $B(t)$ is equal to the *average case error* of multivariate integration for some normed space G_d equipped with a zero-mean probability measure μ_d. We should take the space G_d such that the linear functional $f(x)$ is bounded for any x, and the measure μ_d such that its covariance function is K_d given by (1). For all specific discrepancies mentioned above, we may take G_d as the space of continuous functions with the maximum norm, and μ_d as the zero-mean Gaussian measure with the covariance function K_d; details can be found in [41]. Such a relation was first presented in [67].

One can also ask the converse question whether multivariate integration over reproducing kernel Hilbert spaces is always related to L_2 discrepancy for some sets $B(t)$. The answer is now *no*. This simply follows from the fact that not every reproducing kernel has the form (1). Note that if K_d is given by (1) then $K_d(x,y) \in [0,1]$. There are kernels which can take arbitrary values including negative values as well. For example, the Korobov kernel takes negative values, see *e.g.* [49].

In this paper we also survey bounds on the L_2 discrepancy anchored at 0 and, in particular, their dependence on the dimension d. The lower and upper bounds of Roth and Frolov [47, 48, 21] on the minimal discrepancy anchored at 0 of n points are sharp as a function of n, but also depend on d and the exact form of the dependence on d is *not* known. Even the asymptotic constants, *i.e.* when $d \geq 2$ is fixed and n tends to infinity, are not known. The minimal discrepancy anchored at 0 is at most equal to $3^{-d/2}$ and this bound is sharp for $n = 0$. The case $n = 0$ corresponds to the *initial* discrepancy anchored at 0 and is exponentially small for large d. In terms of the corresponding integration problem this means that the boundary conditions $f(x) = 0$, if one component of x is one for functions f from the unit ball of the Sobolev space anchored at 1, imply that their integrals are at most $3^{-d/2}$. This may indicate that this L_2 discrepancy and multivariate integration are not properly normalized for large d. We can remove the boundary conditions and switch to the *weighted* discrepancy anchored at 0, and to multivariate integration for the *weighted* Sobolev space anchored at 1. Again the weighted discrepancy anchored at 0 of arbitrary n points is the same as the worst case error of a linear algorithm using n function values at the same n points for approximating the integrals for the unit ball of the weighted Sobolev space anchored at 1.

The weights are used to moderate the influence of all groups of variables. We give a brief introduction in Section 4. The choice of weights is a delicate problem. We believe that the weights should be chosen such that the initial weighted discrepancy anchored at 0 is roughly d^q for some $q \geq 0$. For specific weights, we may model functions which are sums of functions of at most ω variables with ω independent of d; this corresponds to the *finite-order weights*, see [18]. Or, we may model functions which depend on the successive variables in a diminishing way; this corresponds to the *product weights*, see [51]. We discuss these in Section 7. The weights are especially needed for large d which occurs in many applications including financial mathematics, physics, chemistry and statistics. In such applications, d is often in the hundreds or thousands. Another type of applications is path integration, where $d = \infty$. Then a finite, but usually very large, d is obtained by approximation of a path integral.

We present a number of estimates on the L_2 weighted discrepancy anchored at 0, and deduce from them *tractability* which is the subject of Section 7. More precisely, we study the minimal number of points $n = n_\gamma(\varepsilon, d)$ for which the weighted L_2 discrepancy anchored at 0 in the d dimensional case is at most ε, which corresponds to the *absolute error criterion*, or at most ε times the initial discrepancy, which corresponds to the *normalized error criterion*. The minimal n means that we choose points t_j optimally. The coefficients in the discrepancy formula can be also chosen optimally or we may fix them to be n^{-1} as it is done for widely used QMC (Quasi Monte Carlo) algorithms. Tractability means that $n_\gamma(\varepsilon, d)$ is *not* exponential in $\varepsilon^{-1} + d$. There are different ways of measuring the lack of exponential dependence. We discuss

- *weak* tractability when
$$\lim_{\varepsilon^{-1}+d\to\infty} (\varepsilon^{-1} + d)^{-1} \log n_\gamma(\varepsilon, d) = 0;$$
- *polynomial* and *strong polynomial* tractability when $n_\gamma(\varepsilon, d)$ is bounded polynomially in ε^{-1} and d, or only polynomially in ε^{-1},
- T-tractability and *strong* T-tractability when $n_\gamma(\varepsilon, d)$ is bounded by a multiple of some power of $T(\varepsilon^{-1}, d)$ or $T(\varepsilon^{-1}, 1)$. Here T is a non-decreasing function of both arguments and T is *not exponential*, *i.e.*
$$\lim_{\varepsilon^{-1}+d\to\infty} (\varepsilon^{-1} + d)^{-1} \log T(\varepsilon^{-1}, d) = 0.$$

In Section 7, we present a number of tractability results. In particular, we discuss the relationship between weights and tractability, and conditions on the weights under which weak, polynomial and strong tractability can be achieved. We also cite conditions on T-tractability and strong T-tractability from [23]. Tractability for other L_2 discrepancies is only briefly mentioned and the reader is referred to [41] for details.

The current paper is related to our paper [40], where a survey of results received up to roughly the year 2000 may be found for L_p discrepancy and multivariate integration including especially the case of the star discrepancy. For the reader's convenience and for completeness we repeat some parts of [40] and present a number of estimates on L_2 discrepancy already given in [40] with updates if there was progress since 2000. A detailed account of the concepts and results presented here may be found in [42].

Acknowledgements. The second author was partially supported by the National Science Foundation under Grant DMS-0608727 and by the Humboldt Research Award at the University of Jena. The authors thank William Chen, Ronald Cools, Harald Niederreiter, Friedrich Pillichshammer, Volodya Temlyakov, Greg Wasilkowski and Art Werschulz for their valuable comments concerning this paper.

2. B-discrepancy

We first define a general L_2 discrepancy which covers many classical examples of L_2 discrepancy and later we relate it to multivariate integration.

Let $D \subseteq \mathbb{R}^{\tau(d)}$ be a measurable set and let $\rho : D \to \mathbb{R}$ be a non-negative measurable function such that $\int_D \rho(t)\,\mathrm{d}t = 1$. Here, $\tau : \mathbb{N} \to \mathbb{N}$ is a given function, so $\tau(d)$ is a given positive integer.

We assume that for any $t \in D$ we have a measurable set $B(t) \subseteq \mathbb{R}^d$, and let $\mathrm{vol}(B(t))$ denote its Lebesgue measure (volume). We also assume that $\mathrm{vol}(B(\cdot))$ is a measurable function, and

$$\int_D (\mathrm{vol}(B(t)))^2 \rho(t)\,\mathrm{d}t < \infty. \tag{2}$$

For given points $t_1, t_2, \dots, t_n \in \mathbb{R}^d$ and coefficients $a_1, a_2, \dots, a_n \in \mathbb{R}$, we approximate the volume of $B(t)$ by a weighted sum of the points t_j which are in the set $B(t)$, so that

$$\mathrm{disc}(t) := \mathrm{vol}(B(t)) - \sum_{j=1}^n a_j 1_{B(t)}(t_j).$$

The L_2 *B-discrepancy* of points t_j and coefficients a_j, or in short *B-discrepancy*, is the weighted L_2 norm of the function $\mathrm{disc}(\cdot)$; more precisely, it is

$$\mathrm{disc}_2^B(\{t_j\},\{a_j\}) = \left(\int_D \left(\mathrm{vol}(B(t)) - \sum_{j=1}^n a_j 1_{B(t)}(t_j) \right)^2 \rho(t)\,\mathrm{d}t \right)^{1/2}.$$

By direct integration we then have

$$\begin{aligned}(\mathrm{disc}_2^B(\{t_j\},\{a_j\}))^2 = &\int_D (\mathrm{vol}(B(t)))^2 \rho(t)\,\mathrm{d}t \\ &- 2\sum_{j=1}^n a_j \int_D \mathrm{vol}(B(t)) 1_{B(t)}(t_j)\rho(t)\,\mathrm{d}t \\ &+ \sum_{i,j=1}^n a_i a_j \int_D 1_{B(t)}(t_i) 1_{B(t)}(t_j) \rho(t)\,\mathrm{d}t. \end{aligned} \tag{3}$$

A popular choice of coefficients a_j is $a_j = n^{-1}$ which corresponds, as we shall see, to *quasi Monte Carlo* (*QMC*) algorithms for multivariate integration.

A major problem of L_2 discrepancy is to find t_j and a_j for which $\mathrm{disc}_2(\{t_j\},\{a_j\})$ is minimized. Let

$$\overline{\mathrm{disc}}_2^B(n,d) = \inf_{t_1,\dots,t_n \in \mathbb{R}^d} \mathrm{disc}_2^B(\{t_j\},\{n^{-1}\})$$

and

$$\mathrm{disc}_2^B(n,d) = \inf_{\substack{t_1,\dots,t_n \in \mathbb{R}^d[0,1)^d \\ a_1,\dots,a_n \in \mathbb{R}}} \mathrm{disc}_2^B(\{t_j\},\{a_j\})$$

denote the minimal L_2 discrepancy when we use n points in dimension d. For $\overline{\mathrm{disc}}_2^B(n,d)$ we choose optimal t_j for $a_j = n^{-1}$ whereas for $\mathrm{disc}_2(n,d)$ we also choose optimal a_j.

For $n = 0$ we do not use any points t_j or coefficients a_j, and obtain the initial B-discrepancy

$$\overline{\mathrm{disc}}_2^B(0,d) = \mathrm{disc}_2^B(0,d) = \left(\int_D (\mathrm{vol}(B(t)))^2 \rho(t)\,\mathrm{d}t \right)^{1/2}.$$

We now present some examples of B-discrepancy.

Discrepancy Anchored at the Origin. Take $D = [0,1]^d$ with $\tau(d) = d$ and $\rho(t) = 1$. Then $B(t) = [0,t)$ corresponds to the L_2 *discrepancy anchored at* 0. This is probably the most popular choice of L_2 discrepancy studied in many papers including the fundamental contributions of Roth; see [47, 48]. In this case, we drop the superscript B and denote $\mathrm{disc}_2^B = \mathrm{disc}_2$. Now (3) becomes

$$(\mathrm{disc}_2(\{t_j\},\{a_j\}))^2 = \frac{1}{3^d} - \frac{1}{2^{d-1}}\sum_{j=1}^n a_j \prod_{k=1}^d (1-t_{j,k}^2) + \sum_{i,j=1}^n a_i a_j \prod_{k=1}^d (1-\max\{t_{i,k},t_{j,k}\}). \tag{4}$$

This formula was first given by Warnock [57]. Hence, $\mathrm{disc}_2^2(\{t_i\},\{a_i\})$ can be computed using $O(dn^2)$ arithmetic operations. Faster algorithms for computing discrepancy for relatively small d have been found by Heinrich; see [20, 24]. The initial discrepancy is now

$$\overline{\mathrm{disc}}_2(0,d) = \mathrm{disc}_2(0,d) = \left(\int_{[0,1]^d} t_1^2\dots t_d^2\,\mathrm{d}t\right)^{1/2} = 3^{-d/2}, \tag{5}$$

which is exponentially small in d. This suggests perhaps that for large d, the L_2 discrepancy is not properly normalized. We shall see later how we cope with this problem.

Discrepancy Anchored at α. As before, we take $D=[0,1]^d$, with $\tau(d)=d$ and $\rho(t)=1$. For $\alpha=(\alpha_1,\dots,\alpha_d)\in[0,1]^d$, let

$$B(t) = [\min\{\alpha_1,t_1\},\max\{\alpha_1,t_1\})\times\dots\times[\min\{\alpha_d,t_d\},\max\{\alpha_d,t_d\}).$$

In particular, if $\alpha_k\le t_k$ for all $k\in[d]:=\{1,\dots,d\}$, we have $B(t)=[\alpha,t)$, whereas if $\alpha_k\ge t_k$ for all $k\in[d]$, we have $B(t)=[t,\alpha)$. This corresponds to the L_2 *discrepancy anchored at* α, and is denoted by $\mathrm{disc}_2^B=\mathrm{disc}_2^\alpha$. The volume of $B(t)$ is now given by

$$\mathrm{vol}(B(t)) = \prod_{k=1}^d (\max\{\alpha_k,t_k\}-\min\{\alpha_k,t_k\}),$$

and the initial discrepancy is

$$\left(\int_{[0,1]^d} (\mathrm{vol}(B(t)))^2\,\mathrm{d}t\right)^{1/2} = \prod_{k=1}^d \left(\frac{1}{3}-\alpha_k(1-\alpha_k)\right)^{1/2} \in [12^{-d/2},3^{-d/2}].$$

Hence, it is exponentially small in d for all α. The formula (3) now becomes

$$\begin{aligned}(\mathrm{disc}_2^\alpha(\{t_j\},\{a_j\}))^2 &= \prod_{k=1}^d\left(\frac{1}{3}-\alpha_k(1-\alpha_k)\right)\\ &- 2\sum_{j=1}^n a_j\prod_{k=1}^d \frac{t_{j,k}(2\alpha_k-t_{j,k})1_{[0,\alpha_k)}(t_{j,k}) + (1-t_{j,k})(1+t_{j,k}-2\alpha_k)1_{[\alpha_k,1]}(t_{j,k})}{2}\\ &+ \sum_{i,j=1}^n a_i a_j \Xi(\alpha,t_i,t_j),\end{aligned}$$

where $\Xi(\alpha,t_i,t_j)$ denotes the product

$$\prod_{k=1}^d (\min\{t_{i,k},t_{j,k}\}1_{[0,\alpha_k)^2}(t_{i,k},t_{j,k}) + (1-\max\{t_{i,k},t_{j,k}\})1_{[\alpha_k,1]^2}(t_{i,k},t_{j,k})).$$

This can be computed using $O(dn^2)$ arithmetic operations. Obviously, for $\alpha=0$ this notion coincides with the L_2 discrepancy mentioned before.

Quadrant Discrepancy Anchored at α. Again we take $D=[0,1]^d$, with $\tau(d)=d$ and $\rho(t)=1$. For $\alpha=(\alpha_1,\dots,\alpha_d)\in[0,1]^d$, let

$$B(t)=[w_1(t),z_1(t))\times\ldots\times[w_d(t),z_d(t)),$$

where $[w_k(t),z_k(t))=[0,t_k)$ if $t_k<\alpha_k$, and $[w_k(t),z_k(t))=[t_k,1)$ if $t_k\ge\alpha_k$. In other words, the set of points $t\in[0,1]^d$ is partitioned into 2^d quadrants according to whether $t_k<\alpha_k$ or $t_k\ge\alpha_k$. The set $B(t)$ denotes the box with one corner at t and the opposite corner defined by the unique vertex of $[0,1]^d$ that lies in the same quadrant as t. For the special case $\alpha=1$, we have $B(t)=[0,t)$ for $t\in[0,1)^d$, as for the L_2 discrepancy anchored at 0. The general situation now corresponds to L_2 *same-quadrant discrepancy with anchor at* α, or in short, the L_2 *quadrant discrepancy at* α. We denote $\mathrm{disc}_2^B=\mathrm{disc}_2^{\alpha,\mathrm{quad}}$. For $\alpha=(1/2,\dots,1/2)$, this type of discrepancy was studied by Hickernell [25] who calls it the *centred discrepancy.* For a general α, the quadrant discrepancy was studied in the L_∞ norm by Hickernell, Sloan and Wasilkowski [26], and we presented its L_2 analogue above. The volume of $B(t)$ is

$$\mathrm{vol}(B(t))=\prod_{k=1}^d\Big(t_k1_{[0,\alpha_k)}(t_k)+(1-t_k)1_{[\alpha_k,1]}(t_k)\Big),$$

and the initial discrepancy is given by

$$\left(\int_{[0,1]^d}(\mathrm{vol}(B(t)))^2\,\mathrm{d}t\right)^{1/2}=\prod_{k=1}^d\left(\frac{1}{3}-\alpha_k(1-\alpha_k)\right)^{1/2}\in[12^{-d/2},3^{-d/2}].$$

Again, it is exponentially small in d. The formula (3) now becomes

$$\begin{aligned}(\mathrm{disc}_2^{\alpha,\mathrm{quad}}(\{t_j\},\{a_j\}))^2&=\prod_{k=1}^d\left(\frac{1}{3}-\alpha_k(1-\alpha_k)\right)\\ &-2\sum_{j=1}^n a_j\prod_{k=1}^d\frac{(\alpha_k^2-t_{j,k}^2)1_{[0,\alpha_k)}(t_{j,k})+(t_{j,k}-\alpha_k)(2-t_{j,k}-a_k)1_{[\alpha_k,1]}(t_{j,k})}{2}\\ &+\sum_{i,j=1}^n a_ia_j\prod_{k=1}^d\frac{|t_{i,k}-\alpha_k|+|t_{j,k}-\alpha_k|-|t_{i,j}-t_{j,k}|}{2}.\end{aligned}$$

Extreme or Unanchored Discrepancy. In this case, we have $\tau(d)=2d$ and $\rho(t)=1$, with $D=\{(x,y)\in[0,1]^{2d}:x\le y\}$. Then for $t=(x,y)$ and $B(t)=[x,y)$ we have the L_2 *extreme or unanchored discrepancy*, denoted by $\mathrm{disc}_2^B=\mathrm{disc}_2^{\mathrm{ex}}$. This type of L_2 discrepancy was introduced by Morokoff and Caflisch in [36]. These authors prefer to use the unanchored discrepancy since it does not prefer a particular vertex, like the L_2 discrepancy anchored at the origin or at α. Now (3) becomes

$$\begin{aligned}&(\mathrm{disc}_2^{\mathrm{ex}}(\{t_j\},\{a_j\}))^2\\ &=\frac{1}{12^d}-\frac{1}{2^{d-1}}\sum_{j=1}^n a_j\prod_{k=1}^d t_{j,k}(1-t_{j,k})+\sum_{i,j=1}^n a_ia_j\prod_{k=1}^d(\min\{t_{i,k},t_{j,k}\}-t_{i,k}t_{j,k}).\end{aligned}$$

For $n=0$, we obtain the initial L_2 unanchored discrepancy $\mathrm{disc}_2^{\mathrm{ex}}(0,d)=12^{-d/2}$ which is exponentially small in d, as all discrepancies before.

Ball Discrepancy. We now take $\tau(d)=d+1$, and define $D=\{[c,r]:c\in\mathbb{R}^d,\ r\ge 0\}$. The weight function ρ may be defined as $\rho(c,r)=1$ for $[c,r]\in[0,1]^{d+1}$ and zero otherwise. We may also consider the case for which ρ is the density function of the Gaussian measure on $\mathbb{R}^d \times \mathbb{R}_+$,

$$\rho(c,r) = 2(2\pi)^{-(d+1)/2} \exp\left(-\frac{r^2}{2} - \frac{1}{2}\sum_{k=1}^{d} c_k^2\right).$$

Then for $t = (c,r)$, let $B(t) = \{x \in \mathbb{R}^d : \|x-c\|_p \le r\}$ be the ball at centre c and radius r in the l_p norm, *i.e.*

$$\|x-c\|_p = \left(\sum_{k=1}^{d} |x_k - c_k|^p\right)^{1/p}, \quad p \in [1,\infty),$$

and

$$\|x-c\|_\infty = \max_{j\in[d]} |x_j - c_j|.$$

This situation corresponds to the L_2 *ball discrepancy in the* l_p *norm.* We shall only consider the case $p=\infty$ in Section 6. Here as an aside, we also follow Chen and Travaglini [12] and define a *periodic ball discrepancy in the* ℓ_p *case*, denoted by $\mathrm{disc}_2^B = \mathrm{disc}_2^{\mathrm{pball}}$, as follows. For $x, y \in [0,1]^d$, let

$$\|x-y\|_p^* = \left(\sum_{k=1}^{d} |x_k - y_k|_*^p\right)^{1/p},$$

where $|x_k - y_k|_* = \min\{|x_k - y_k|, 1 - |x_k - y_k|\}$. Observe that

$$\|x-y\|_p^* \le \frac{d^{1/p}}{2}, \quad x, y \in [0,1]^d.$$

Now we let $\tau(d) = d+1$ and define

$$D = \left\{(c,r) : c \in [0,1]^d,\ 0 \le r \le \frac{d^{1/p}}{2}\right\}.$$

For $t=(c,r)$, we consider the set $B(t) = \{x \in [0,1]^d : \|x-c\|_p^* \le r\}$. Formulas for the periodic ball discrepancy in the case $p=2$ can be found in [12] in terms of the Fourier coefficients of the characteristic function of the ball of radius r. At the end of Section 3, we shall briefly discuss estimates for the discrepancy in this case.

3. Bounds for the L_2 Discrepancy

We briefly review bounds on the minimal L_2 discrepancy. These bounds are similar for the L_2 discrepancy anchored at α, the L_2 quadrant discrepancy anchored at α and the L_2 unanchored discrepancy, see [39, 40]. That is why we restrict ourselves only to the L_2 discrepancy anchored at 0 and to the L_2 periodic ball discrepancy.

We begin with the minimal L_2 discrepancy anchored at 0. It is easy to find the exact values of $\overline{\mathrm{disc}}_2(n,d)$ and $\mathrm{disc}_2(n,d)$ for $d=1$ which are of exact order n^{-1}. For $d \ge 1$, the exact value of $\overline{\mathrm{disc}}_2(n,d)$ is only known for $n=1$, see [44]. For a fixed d and $n \ge 2$ we only know bounds. More precisely, there exist positive numbers c_d and C_d such that

$$c_d \frac{(\log n)^{(d-1)/2}}{n} \le \mathrm{disc}_2(n,d) \le \overline{\mathrm{disc}}_2(n,d) \le C_d \frac{(\log n)^{(d-1)/2}}{n}. \tag{6}$$

The lower bound is a celebrated result of Roth [47] proved in 1954 for $a_j = n^{-1}$. For arbitrary a_j, using essentially the same proof technique, the lower bound was proved by Chen [8, 9]. The upper bound was proven by Roth [48] in 1980 and independently by Frolov [21] again for $a_j = n^{-1}$. The original proofs of upper bounds were not fully constructive. Today we have constructive proofs due to Chen and Skriganov [11]. The same upper bounds can be also achieved by using randomized digital nets in prime base as shown by Chen [7] and elaborated in a recent paper by Cristea, Dick and Pillichshammer [16], see also Dick and Pillichshammer [17].

The essence of (6) is that modulo a logarithmic factor, the L_2 discrepancy behaves asymptotically in n like n^{-1} independently of d. The power of the logarithmic factor is $(d-1)/2$ and as long as d is not too large this factor is negligible. On the other hand, if d is large, say $d = 360$ as in some financial applications, the factor $(\log n)^{(d-1)/2}$ is very essential. Indeed, the function $n^{-1}(\log n)^{(d-1)/2}$ is increasing for $n \le \exp((d-1)/2)$. The latter number for $d = 360$ is $\exp(179.5) \approx 9 \cdot 10^{77}$. Obviously, it is impossible to use n so large, and therefore for large d, the good asymptotic behavior of $\mathrm{disc}_2(n,d)$ cannot be really utilized for practical purposes.

For large d, the numbers c_d and C_d from (6) are also very important. We do not know much about them. However, we know that the asymptotic constant

$$A_d = \limsup_{n\to\infty} \overline{\mathrm{disc}}_2(n,d) \frac{n}{(\log n)^{(d-1)/2}}$$

is super-exponentially small in d.

For large d and a relatively small n, we need other estimates on $\mathrm{disc}_2(n,d)$. By a simple averaging argument with respect to t_j and for $a_j = n^{-1}$, we have

$$\overline{\mathrm{disc}}_2(n,d) \le \left(\int_{[0,1]^{nd}} (\mathrm{disc}_2(\{t_j\},\{n^{-1}\}))^2 \, \mathrm{d}t_1 \dots \mathrm{d}t_n \right)^{1/2} \le \frac{2^{-d/2}}{n^{1/2}}. \qquad (7)$$

The last estimate looks very promising since we have an exponentially small dependence on d through the factor $2^{-d/2}$. However, we should keep in mind that even the initial L_2 discrepancy is $3^{-d/2}$ which is much smaller than $2^{-d/2}$ for large d. From the last estimate we can easily conclude by applying Chebyshev's inequality that for any number $c > 1$, the set of sample points

$$A_c = \{\{t_j\} \in [0,1]^{dn} \ : \ \mathrm{disc}_2(\{t_j\},\{n^{-1}\}) \le c\, 2^{-d/2} n^{-1/2}\}$$

has Lebesgue measure at least $1 - c^{-2}$. Hence, for $c = 10$ we have a set of points $\{t_j\}$ of measure at least 0.99 for which the L_2 discrepancy is at most $10 \cdot 2^{-d/2} n^{-1/2}$. Surprisingly enough, we still do *not* know how to construct such points. Of course, such points can be found computationally. Indeed, it is enough to take points $t_1, \dots, t_n$ randomly, as independent and uniformly distributed points over $[0,1]^d$, and compute their discrepancy with $a_j = n^{-1}$. If their discrepancy is at most $10 \cdot 2^{-d/2} n^{-1/2}$ we are done. If not, we repeat random selection of points $t_1, \dots, t_n$. Then after a few such selections we will get the desired points since the failure of k trials is c^{-2k}, or 10^{-2k} for $c = 10$, which is exponentially small in k.

The bound (6) justifies the definition of *low discrepancy* sequences, and points which we do not cover here, for the coefficients a_j equal n^{-1}. Namely, the sequence $\{t_j\}$ is a low discrepancy sequence if there is a positive number C_d such that

$$\mathrm{disc}_2(\{t_j\},\{n^{-1}\}) \le C_d \frac{(\log n)^d}{n} \quad \text{for all } n \ge 2. \qquad (8)$$

That is, the L_2 discrepancy of low discrepancy sequences enjoys almost the same asymptotic as the minimal L_2 discrepancy with the only difference being in the power of the logarithmic factor. The search for low discrepancy sequences has been a very active research area, and many beautiful and deep constructions have been obtained. Such sequences usually bear the name of their authors. Today we know low discrepancy sequences (and points) of Faure, Halton, Hammersley, Niederreiter, Sobol and Tezuka, as well as (t,m,s) points and (t,m) nets, and lattice or shifted lattice points as their counterparts for the periodic case; see [4, 6, 19, 34, 37, 49, 54].

There are also points and sequences satisfying (8) with more general coefficients a_j than n^{-1}. An example is provided by hyperbolic points, see [52, 58], although in this case we have $(\log n)^{3d/2}$ instead of $(\log n)^d$ in (8). Explicit bounds on the L_2 discrepancy for hyperbolic points can be found in [58]. In particular, hyperbolic points $t_1, \ldots, t_n$ and coefficients $a_1, \ldots, a_n$ were constructed such that $\mathrm{disc}_2(\{t_j\}, \{a_j\}) \le \varepsilon$ with

$$n \le \min\left\{ \frac{3.304}{\varepsilon}\left(1.77959 + 2.714\frac{\log \varepsilon^{-1} - 1.12167}{d-1}\right)^{3(d-1)/2}, 7.26\left(\frac{1}{\varepsilon}\right)^{2.454}\right\}.$$

Observe an intriguing dependence on d in the first term of the minimum. For a fixed large d, and ε tending to zero, we have

$$n = O\left(\left(\frac{2.714}{d-1}\right)^{3(d-1)/2} \frac{(\log \varepsilon^{-1})^{3(d-1)/2}}{\varepsilon}\right).$$

On the other hand, we also have a polynomial bound on n in ε^{-1} for *all* d. Again it looks more surprising than it is since the initial L_2 discrepancy is exponentially small in d. It is not known if the exponent 2.454 for the second term of the minimum is sharp. Clearly, it has to be at least one but it is very likely that it may be lower.

The last estimate can be rewritten as

$$\mathrm{disc}_2(n,d) \le \frac{2.244}{n^{0.408}} \quad \text{for all } n \text{ and } d,$$

and is obtained by hyperbolic points. This bound suggests that we should try to find the smallest number (or the infimum) of positive p for which there exists a positive C such that

$$\mathrm{disc}_2(n,d) \le C n^{-1/p} \quad \text{for all } n \text{ and } d. \tag{9}$$

We stress that the last estimate holds for all n and d; hence neither C nor p depends on n and d. Such a value of p is denoted by p^* and called the *exponent* p^* of the L_2 discrepancy; see [59].

The bound $p^* \ge 1$ is obvious, since for $d = 1$, we have

$$\mathrm{disc}_2(n,1) = \Theta(\overline{\mathrm{disc}}_2(n,1)) = \Theta(n^{-1}).$$

For $a_j = n^{-1}$, it is proved by Matoušek [33] that $p \ge 1.0669$ in (9). This means that the case of arbitrary d is harder than the univariate case and the presence of the logarithmic factors in (6) cannot be entirely neglected. The upper bound

$$p^* \le 1.4779$$

is established in [59], but its proof is *non-constructive.* The best constructive bound currently known is $p = 2.454$ from the estimate for hyperbolic points; see [58]. It was shown by Plaskota [45] that using hyperbolic points, or more generally nested sparse grids points, leads to $p \ge 2.1933$. To obtain $p < 2.1933$, we must therefore use other constructions than nested sparse grids points.

There are two challenging problems concerning the exponent of L_2 discrepancy. The first is to find p^*, and the second is to construct points t_j for which (9) holds with $p < 2$.

We now address the problem of the exponentially small initial L_2 discrepancy. One way to omit this problem is to switch to the normalized case. By the normalized L_2 discrepancy we mean

$$\frac{\mathrm{disc}_2(\{t_j\},\{a_j\})}{\mathrm{disc}_2(0,d)}.$$

In other words, we normalize by the initial value of the L_2 discrepancy, which is $3^{-d/2}$. We now define

$$\overline{n}(\varepsilon, d) = \min\{n : \overline{\mathrm{disc}}_2(n,d) \le \varepsilon\, \mathrm{disc}_2(0,d)\} \tag{10}$$

and

$$n(\varepsilon,d) = \min\{n : \mathrm{disc}_2(n,d) \le \varepsilon\, \mathrm{disc}_2(0,d)\} \tag{11}$$

as the minimal number of points necessary to reduce the initial discrepancy by a factor ε respectively with the coefficients $a_j = n^{-1}$ and with optimally chosen a_j. We ask whether $\overline{n}(\varepsilon,d)$ and $n(\varepsilon,d)$ behave polynomially in ε^{-1} and d or at least not exponentially in ε^{-1} and d. We stress that polynomial bounds on the absolute value of the L_2 discrepancy, which we presented so far, are useless for the normalized case since we now have to compare the minimal L_2 discrepancy to $\varepsilon 3^{-d/2}$ instead of ε.

The problem how $\overline{n}(\varepsilon,d)$ and $n(\varepsilon,d)$ depend on d has been solved and we now report its solution. First of all, notice that it follows directly from (5) and (7) that

$$\overline{n}(\varepsilon,d) \le \left(\frac{3}{2}\right)^d \varepsilon^{-2}. \tag{12}$$

It was shown in [68], see also [51], that

$$\overline{n}(\varepsilon,d) \ge \left(\frac{9}{8}\right)^d (1-\varepsilon^2). \tag{13}$$

The bound (13) is also valid if all coefficients a_j are non-negative. Hence, we have exponential dependence on d.

For arbitrary a_j, it was shown in [39] that for any positive $\varepsilon_0 < 1$ there exists a positive c such that

$$c\, 1.0628^d \le n(\varepsilon,d) \le \left(\frac{3}{2}\right)^d \varepsilon^{-2} \tag{14}$$

for all d and $\varepsilon \in (0,\varepsilon_0)$. Hence, $n(\varepsilon,d)$ goes to infinity exponentially fast in d. The upper bounds in (12) and in (14) coincide. We do not know whether arbitrary a_j are better than positive a_j or the same as n^{-1}.

Finally, we briefly turn to the minimal L_2 periodic ball discrepancy $\mathrm{disc}_2^{\mathrm{pball}}(n,d)$. The lower bound is due to Beck [3] and Montgomery [35] and states that for an arbitrary d there is a positive c_d such that

$$\mathrm{disc}_2^{\mathrm{pball}}(n,d) \ge c_d n^{-1/2-1/2d}$$

for all n. This bound is essentially sharp as shown by Beck and Chen [5]; see also Chen [10] and Travaglini [56]. Note that for large d, the exponent of n^{-1} is close to $1/2$ which is the worst possible exponent for all B-discrepancy for which the sets $B(t)$ are subsets D_d with $\mathrm{vol}(D_d) < \infty$; see Section 5.

4. Weighted L_2 Discrepancy

As we shall see later, the L_2 discrepancy anchored at 0 is related to multivariate integration for functions satisfying some boundary conditions which are probably not very common in practical computation. To remove these boundary conditions we need to consider a little more general L_2 discrepancy. Furthermore, for large d, the integrands may depend differently on groups of variables. To address this property, we need to consider *weighted* L_2 discrepancy.

By $\mathfrak{u}$ we denote an arbitrary subset of $[d] = \{1,\dots,d\}$. We are given a sequence

$$\gamma = \{\gamma_{d,\mathfrak{u}}\}_{d=1,2,\dots;\mathfrak{u}\subseteq[d]}$$

of non-negative weights. For simplicity, we assume that $\gamma_{d,\mathfrak{u}} \in [0,1]$.

For a vector $t \in [0,1]^d$, we denote by $t_{\mathfrak{u}}$ the vector from $[0,1]^{|\mathfrak{u}|}$, where $|\mathfrak{u}|$ is the cardinality of $\mathfrak{u}$, with the components of t whose indices are in $\mathfrak{u}$. For example, for $d = 7$ and $\mathfrak{u} = \{2,4,5,6\}$ we have $t_{\mathfrak{u}} = (t_2,t_4,t_5,t_6)$. Then

$$\mathrm{d}t_{\mathfrak{u}} = \prod_{k\in\mathfrak{u}} \mathrm{d}t_k.$$

By $(t_\mathfrak{u}, 1)$ we mean the vector from $[0,1]^d$ with the same components as t for indices in $\mathfrak{u}$ and with the rest of components being replaced by 1. For our example, we have $(t_\mathfrak{u}, 1) = [1, t_2, 1, t_4, t_5, t_6, 1]$. Recall that for given points $t_1, \dots, t_n \in [0,1)^d$ and real coefficients $a_1, \dots, a_n$, we have

$$\mathrm{disc}_2((t_\mathfrak{u}, 1)) = \prod_{k\in\mathfrak{u}} t_k - \sum_{j=1}^n a_j 1_{[0,t_\mathfrak{u})}((t_j)_\mathfrak{u}).$$

The *weighted* L_2 discrepancy anchored at 0, or simply the L_2 weighted discrepancy, is then defined as

$$\mathrm{disc}_{2,\gamma}(\{t_j\}, \{a_j\}) = \left(\sum_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} \int_{[0,1]^{|\mathfrak{u}|}} (\mathrm{disc}_2((t_\mathfrak{u}, 1)))^2 \, \mathrm{d}t_\mathfrak{u} \right)^{1/2}. \tag{15}$$

If $\gamma_{d,\mathfrak{u}} = 0$ for all $\mathfrak{u}$ with $|\mathfrak{u}| < d$ and $\gamma_{d,[d]} = 1$, then the L_2 weighted discrepancy reduces to the L_2 discrepancy. It is easy to obtain an explicit formula for the L_2 weighted discrepancy. Using Warnock's formula (4), we see immediately that $(\mathrm{disc}_{2,\gamma}(\{t_j\}, \{a_j\}))^2$ is given by

$$\sum_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} \left(\frac{1}{3^{|\mathfrak{u}|}} - \frac{1}{2^{|\mathfrak{u}|-1}} \sum_{j=1}^n a_j \prod_{k\in\mathfrak{u}} (1 - t_{j,k}^2) + \sum_{i,j=1}^n a_i a_j \prod_{k\in\mathfrak{u}} (1 - \max\{t_{i,k}, t_{j,k}\}) \right).$$

The standard (unweighted) case corresponds to $\gamma = \{1\}$, *i.e.* $\gamma_{d,\mathfrak{u}} = 1$ for all $\mathfrak{u} \subseteq [d]$. In this case, one can show that $(\mathrm{disc}_{2,\{1\}}(\{t_j\}, \{a_j\}))^2$ is given by

$$\left(\frac{4}{3}\right)^d - 2\sum_{j=1}^n a_j \prod_{k=1}^d \frac{3 - t_{j,k}^2}{2} + \sum_{i,j=1}^n a_i a_j \prod_{k=1}^d (2 - \max\{t_{i,k}, t_{j,k}\}).$$

As before, for an arbitrary sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$, let

$$\overline{\mathrm{disc}}_{2,\gamma}(n,d) = \inf_{t_1,\dots,t_n\in[0,1)^d} \mathrm{disc}_{2,\gamma}(\{t_j\}, \{n^{-1}\}) \tag{16}$$

and

$$\mathrm{disc}_{2,\gamma}(n,d) = \inf_{\substack{t_1,\dots,t_n\in[0,1)^d \\ a_1,\dots,a_n\in\mathbb{R}}} \mathrm{disc}_{2,\gamma}(\{t_j\}, \{a_j\}) \tag{17}$$

be the minimal weighted L_2 discrepancies. For $n = 0$, we obtain

$$(\overline{\mathrm{disc}}_{2,\gamma}(0,d))^2 = (\mathrm{disc}_{2,\gamma}(0,d))^2 = \sum_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} 3^{-|\mathfrak{u}|}.$$

Observe that for the unweighted case $\gamma_{d,\mathfrak{u}} = 1$, we have $\mathrm{disc}_2(0,d) = (4/3)^{d/2}$, which is exponentially large in d. Hence, the initial L_2 discrepancy is exponentially small in d whereas the unweighted L_2 discrepancy is exponentially *large* in d. Both cases seem to be ill-normalized. We believe that the choice of the weight sequence γ should be such that the initial weighted L_2 discrepancy is roughly d^q for some $q \geq 0$.

How small is the minimal weighted discrepancy? We can average the square of the weighted L_2 discrepancy for uniformly and independently distributed t_j over

$[0,1]^d$ and coefficients $a_j = n^{-1}$. We obtain

$$\begin{aligned}
&\int_{[0,1]^{nd}} (\mathrm{disc}_{2,\gamma}(\{t_j\},\{n^{-1}\}))^2 \,\mathrm{d}t_1 \dots \mathrm{d}t_n \\
&\quad = \sum_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} \left(\frac{1}{3^{|\mathfrak{u}|}} - \frac{1}{2^{|\mathfrak{u}|-1}} \left(\frac{2}{3}\right)^{|\mathfrak{u}|} + \frac{1}{n}\left(\frac{1}{2}\right)^{|\mathfrak{u}|} + \frac{n-1}{n}\left(\frac{1}{3}\right)^{|\mathfrak{u}|} \right) \\
&\quad = \frac{1}{n} \sum_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} \left(\left(\frac{1}{2}\right)^{|\mathfrak{u}|} - \left(\frac{1}{3}\right)^{|\mathfrak{u}|} \right).
\end{aligned}$$

By the mean value theorem we conclude that

$$\mathrm{disc}_{2,\gamma}(n,d) \le \overline{\mathrm{disc}}_{2,\gamma}(n,d) \le \frac{1}{n^{1/2}} \left(\sum_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} (2^{-|\mathfrak{u}|} - 3^{-|\mathfrak{u}|}) \right)^{1/2}. \tag{18}$$

Applying Chebyshev's inequality we also conclude that the set of sample points

$$A_c = \left\{ \{t_j\} : \mathrm{disc}_2(\{t_j\},\{n^{-1}\}) \le c\, n^{-1/2} \left(\sum_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} (2^{-|\mathfrak{u}|} - 3^{-|\mathfrak{u}|}) \right)^{1/2} \right\}$$

has Lebesgue measure at least $1 - c^{-2}$. For the unweighted case $\gamma_{d,\mathfrak{u}} = 1$, we have

$$\mathrm{disc}_{2,\gamma}(n,d) \le \overline{\mathrm{disc}}_{2,\gamma}(n,d) \le n^{-1/2} \left(\left(\frac{3}{2}\right)^d - \left(\frac{4}{3}\right)^d \right)^{1/2} \le n^{-1/2} \left(\frac{3}{2}\right)^{d/2}.$$

Lower bounds on the weighted L_2 discrepancy can be found in [39]. The consequences of the upper and lower bounds on the weighted L_2 discrepancy will be discussed later after we have presented relations between the weighted L_2 discrepancy and multivariate integration for some Sobolev spaces.

As for the normalized L_2 discrepancy, we let

$$\overline{n}_\gamma(\varepsilon,d) = \min\{n : \overline{\mathrm{disc}}_{2,\gamma}(n,d) \le \varepsilon\ \mathrm{disc}_{2,\gamma}(0,d)\}$$

and

$$n_\gamma(\varepsilon,d) = \min\{n : \mathrm{disc}_{2,\gamma}(n,d) \le \varepsilon\ \mathrm{disc}_{2,\gamma}(0,d)\}$$

be the minimal number of points necessary to reduce the initial weighted discrepancy by a factor ε respectively with the coefficients $a_j = n^{-1}$ and with optimally chosen a_j. For the unweighted case, $\gamma = \{1\}$, it was shown in [39] that for any positive $\varepsilon_0 < 1$ there exists a positive c such that

$$c\,1.0463^d \le n_{\{1\}}(\varepsilon,d) \le \overline{n}_{\{1\}}(\varepsilon,d) \le \left(\frac{9}{8}\right)^d \varepsilon^{-2} \tag{19}$$

for all d and $\varepsilon \in (0,\varepsilon_0)$.

Hence, for the normalized L_2 discrepancy as well as for the normalized weighted L_2 discrepancy with $\gamma_{d,\mathfrak{u}} = 1$, we have an exponential dependence on d, and the corresponding $n(\varepsilon,d)$ and $n_{\{1\}}(\varepsilon,d)$ go exponentially fast with d to infinity. It is now natural to ask what necessary and sufficient conditions on the weight sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ guarantee no exponential dependence on d, and what we have to assume about γ to guarantee, say, polynomial dependence on d, or no dependence on d at all. We will study these questions later.

5. Multivariate Integration

We consider multivariate integration for real functions defined on a measurable set $D_d \subset \mathbb{R}^d$ and which belong to a Hilbert space with a reproducing kernel $K_d : D_d \times D_d \to \mathbb{R}$. This space is denoted by $H(K_d)$ and its inner product by $\langle \cdot, \cdot \rangle_{H(K_d)}$. The basic information about such spaces can be found in [1]. Here, we only mention that $K_d(\cdot, x) \in H(K_d)$ for all $x \in D_d$, that $(K_d(x_i, x_j))_{i,j=1,\ldots,m}$ is a $m \times m$ symmetric positive semi-definite matrix and this holds for any choice of m and $x_i, x_j \in D_d$. Furthermore, and this is probably the most important property, for any function $f \in H(K_d)$ and any $x \in D_d$, we have

$$f(x) = \langle f, K_d(\cdot, x) \rangle_{H(K_d)}.$$

The space $H(K_d)$ is the completion of linear combinations of functions of the form

$$\sum_{j=1}^{m} a_j K_d(\cdot, x_j)$$

for any m, real a_j and $x_j \in D_d$.

We illustrate the reproducing kernel Hilbert spaces for two examples with $D_d = [0,1]^d$. For the first example and $d = 1$, we take a number $\beta \in [0,1]$ and let

$$K_1^{\beta}(x, y) = \frac{|x - \beta| + |y - \beta| - |x - y|}{2}, \quad x, y \in [0,1].$$

Note that for $\beta = 0$, we have $K_1^0(x, y) = (x + y - |x - y|)/2 = \min\{x, y\}$, whereas for $\beta = 1$, we have $K_1^1(x, y) = (1 - x + 1 - y - |x - y|)/2 = 1 - \max\{x, y\}$. For an arbitrary β, the kernel K_1^{β} has the property that it vanishes for $x \le \beta \le y$ and $y \le \beta \le x$.

The space $H(K_1^{\beta})$ consists of absolutely continuous functions vanishing at β and whose first derivatives are in $L_2([0,1])$. In other words,

$$H(K_1^{\beta}) = \{f : [0,1] \to \mathbb{R} : f(\beta) = 0,\ f \text{ is absolutely continuous},\ f' \in L_2([0,1])\}$$

with the inner product for $f, g \in H(K_1^{\beta})$ given by

$$\langle f, g \rangle_{H(K_1^{\beta})} = \int_0^1 f'(x) g'(x) \, \mathrm{d}x.$$

For $d \ge 1$ and a vector $\beta \in [0,1]^d$, we define $H(K_d^{\beta})$ to be the d fold tensor product of $H(K_1^{\beta_j})$, *i.e.*

$$H(K_d^{\beta}) = H(K_1^{\beta_1}) \otimes \ldots \otimes H(K_1^{\beta_d}),$$

with the reproducing kernel given by

$$K_d^{\beta}(x, y) = \prod_{k=1}^{d} K_1^{\beta_k}(x_k, y_k), \quad x, y \in [0,1]^d.$$

The space $H(K_d^{\beta})$ consists of functions such that $f(x) = 0$ if there exists an index $j \in [d]$ such that $x_j = \beta_j$, and which are one time differentiable with respect to all variables, and the resulting partial derivatives are in $L_2([0,1]^d)$. The inner product for $f, g \in H(K_d^{\beta})$ is given by

$$\langle f, g \rangle_{H_{K_d^{\beta}}} = \int_{[0,1]^d} \frac{\partial^d}{\partial x_1 \ldots \partial x_d} f(x) \, \frac{\partial^d}{\partial x_1 \ldots \partial x_d} g(x) \, \mathrm{d}x.$$

The space $H(K_d^{\beta})$ is called the *Sobolev space anchored at* β.

As a second example of a reproducing kernel Hilbert space, take an arbitrary weight sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ with $\gamma_{d,\mathfrak{u}} \geq 0$. Define the reproducing kernel as

$$K_{d,\gamma}^{\beta}(x,y) = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} K_{\mathfrak{u}}^{\beta}(x,y),$$

with

$$K_{\mathfrak{u}}^{\beta}(x,y) = \prod_{k \in \mathfrak{u}} K_1^{\beta}(x_k,y_k) = \prod_{k \in \mathfrak{u}} \frac{|x_k - \beta_k| + |y_k - \beta_k| - |x_k - y_k|}{2}$$

for $x, y \in [0,1]^d$. For the unweighted case $\gamma_{d,\mathfrak{u}} = \{1\}$, we have

$$\begin{aligned} K_{d,\{1\}}^{\beta}(x,y) &= \prod_{k=1}^{d} (1 + K_1^{\beta}(x_k,y_k)) \\ &= \prod_{k=1}^{d} \left(1 + \frac{|x_k - \beta_k| + |y_k - \beta_k| - |x_k - y_k|}{2}\right). \end{aligned}$$

The Hilbert space $H(K_{d,\gamma}^{\beta})$ is the sum of tensor product Hilbert space $H(K_{\mathfrak{u}}^{\beta})$ for all $\mathfrak{u}$ for which $\gamma_{d,\mathfrak{u}}$ is positive. For all positive $\gamma_{d,\mathfrak{u}}$, the inner product for $f, g \in H(K_{d,\gamma})$ is given by

$$\langle f, g \rangle_{H(K_{d,\gamma}^{\beta})} = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \int_{[0,1]^{|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|}}{\partial x_{\mathfrak{u}}} f(x_{\mathfrak{u}}, \beta) \int_{[0,1]^{|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|}}{\partial x_{\mathfrak{u}}} g(x_{\mathfrak{u}}, \beta) \, \mathrm{d}x_{\mathfrak{u}},$$

with the notation

$$\partial x_{\mathfrak{u}} = \prod_{k \in \mathfrak{u}} \partial x_k \quad \text{and} \quad \mathrm{d}x_{\mathfrak{u}} = \prod_{k \in \mathfrak{u}} \mathrm{d}x_k.$$

Here, $(x_{\mathfrak{u}}, \beta)$ denotes the vector with d components, with the j-th component equal to x_j if $j \in \mathfrak{u}$ and equal to β_j if $j \notin \mathfrak{u}$. In particular, for $\mathfrak{u} = \emptyset$ we have $(x_{\emptyset}, \beta) = \beta$, whereas for $\mathfrak{u} = [d]$ we have $(x_{[d]}, \beta) = x$. For $\mathfrak{u} = \emptyset$, we have $K_{\emptyset}^{\beta} = 1$ and $H(K_{\emptyset}^{\beta}) = \mathrm{span}(1)$. The term in the inner product corresponding to $\mathfrak{u} = \emptyset$ is equal to $\gamma_{d,\emptyset}^{-1} f(\beta) g(\beta)$.

We have a unique decomposition of functions f from $H(K_{d,\gamma}^{\beta})$, in the form

$$f = \sum_{\mathfrak{u} \subseteq [d]} f_{\mathfrak{u}}, \quad f_{\mathfrak{u}} \in H(K_{\mathfrak{u}}^{\beta}),$$

and the terms $f_{\mathfrak{u}}$ are mutually orthogonal so that

$$\|f\|_{H(K_{d,\gamma})}^2 = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \|f_{\mathfrak{u}}\|_{H(K_{\mathfrak{u}}^{\beta})}^2.$$

If some $\gamma_{d,\mathfrak{u}} = 0$, then we assume that the corresponding term $f_{\mathfrak{u}} = 0$ and interpret $0/0$ as 0. Hence, the inner product is the sum of terms for positive $\gamma_{d,\mathfrak{u}}$ with all $f_{\mathfrak{u}} = 0$ if $\gamma_{d,\mathfrak{u}} = 0$.

Observe that $f_{\mathfrak{u}}$ depends only on variables in $\mathfrak{u}$. In particular, $f_{\emptyset}(x) = f(\beta)$ and $f_{\{j\}}(x) = f(\beta_1, \dots, \beta_{j-1}, x_j, \beta_{j+1}, \dots, \beta_d) - f(\beta)$. It is shown in [31] that for any $\mathfrak{u} \subseteq [d]$, we have

$$f_{\mathfrak{u}}(x) = \sum_{\mathfrak{v} \subseteq \mathfrak{u}} (-1)^{|\mathfrak{u}| - |\mathfrak{v}|} f(x_{\mathfrak{v}}, \beta).$$

In general, when some $\gamma_{d,\mathfrak{u}} = 0$, the Hilbert space $H(K_{d,\gamma}^{\beta})$ is not a tensor product space. However, if $\gamma_{d,\mathfrak{u}} = 1$, or more generally if

$$\gamma_{d,\mathfrak{u}} = \prod_{k \in \mathfrak{u}} \gamma_{d,k}$$

for some $\gamma_{d,k} \in [0,1]$, then

$$K_{d,\gamma}^{\beta}(x,y) = \prod_{k=1}^{d} (1 + \gamma_{d,k} K_1^{\beta}(x_k, y_k))$$
$$= \prod_{k=1}^{d} \left(1 + \frac{\gamma_{d,k}(|x_k - \beta_k| + |y_k - \beta_k| - |x_k - y_k|)}{2}\right)$$

is of the product form and this implies that $H(K_{d,\gamma}^{\beta})$ is a tensor product space.

The space $H(K_{d,\gamma}^{\beta})$ is called the *weighted Sobolev space anchored at* β.

We now define multivariate integration for functions from a general reproducing kernel Hilbert space $H(K_d)$. We need to assume that the space $H(K_d)$ consists of integrable functions. To guarantee that multivariate integration is a bounded linear functional, we need to assume that the function

$$h_d(x) = \int_{D_d} K_d(y,x)\,\mathrm{d}y, \quad x \in D_d, \tag{20}$$

belongs to $H(K_d)$. For $f \in H(K_d)$, we define multivariate integration as approximation of

$$I_d(f) = \int_{D_d} f(x)\,\mathrm{d}x. \tag{21}$$

Since $f(x) = \langle f, K_d(\cdot,x)\rangle_{H(K_d)}$, we can rewrite $I_d(f)$ as

$$I_d(f) = \left\langle f, \int_{D_d} K_d(\cdot,x)\,\mathrm{d}x \right\rangle_{H(K_d)} = \langle f, h_d\rangle_{H(K_d)}.$$

Multivariate integration corresponds to the inner product with the generator h_d. Clearly,

$$\|I_d\| := \sup_{\|f\|_{H(K_d)} \le 1} |I_d(f)| = \|h_d\|_{H(K_d)} = \left(\int_{[0,1]^{2d}} K_d(x,y)\,\mathrm{d}x\mathrm{d}y\right)^{1/2}.$$

We approximate $I_d(f)$ by computing function values $f(x_j)$ at some sample points x_j. In general, these points can be chosen *adaptively*, *i.e.* the choice of x_j may depend on the already computed function values $f(x_i)$ for $i = 1,\dots,j-1$. Furthermore, knowing $f(x_j)$ for, say, $j = 1,\dots,n$, we may take $\phi(f(x_1),\dots,f(x_n))$ as an approximation of $I_d(f)$ for some, in general, nonlinear function ϕ. It turns out that adaptation as well as nonlinear choices of ϕ do not help, as proven by Bakhvalov (adaptation) and by Smolyak (nonlinear ϕ); see the original paper of Bakhvalov [2] which presents both results. These results can be also found in, for example, [55]. Hence, without loss of generality, we may restrict ourselves to *linear and non-adaptive* approximations of the form

$$Q_{n,d}(f) = \sum_{j=1}^{n} a_j f(t_j) \tag{22}$$

for some real a_j and, *a priori*, non-adaptively given t_j from $[0,1]^d$. Usually, $Q_{n,d}$ is called a *linear* algorithm. If we let $a_j = n^{-1}$, then

$$Q_{n,d}(f) = \frac{1}{n}\sum_{j=1}^{n} f(t_j) \tag{23}$$

is called a QMC (quasi-Monte Carlo) algorithm and these formulas are often used in numerical computational practice as approximations of multivariate integrals especially when d is large.

The worst case error of $Q_{n,d}$ is defined as the largest error between $I_d(f)$ and $Q_{n,d}(f)$ over the unit ball of $H(K_d)$; more precisely,

$$e^{\mathrm{wor}}(Q_{n,d}; H(K_d)) = \sup_{\substack{f\in H(K_d)\\ \|f\|_{H(K_d)}\le 1}} |I_d(f) - Q_{n,d}(f)|.$$

Due to linearity of $I_d - Q_{n,d}$, the worst case error is obviously the same as the norm $\|I_d - Q_{n,d}\|$. Furthermore, for any $f \in H(K_d)$ of arbitrary norm, we have

$$|I_d(f) - Q_{n,d}(f)| \le e^{\mathrm{wor}}(Q_{n,d}; H(K_d))\|f\|_{H(K_d)}.$$

At first glance, it may seem surprising but there is an explicit formula for the worst case error $e^{\mathrm{wor}}(Q_{n,d}; H(K_d))$. Indeed, we have

$$Q_{n,d}(f) = \left\langle f, \sum_{j=1}^n a_j K_d(\cdot, t_j)\right\rangle_{H(K_d)},$$

which yields

$$I_d(f) - Q_{n,d}(f) = \langle f, h_{d,n}\rangle_{H(K_d)}, \quad h_{d,n} = h_d - \sum_{j=1}^n a_j K_d(\cdot, t_j).$$

From this we easily conclude that

$$e^{\mathrm{wor}}(Q_{n,d}; H(K_d)) = \|I_d - Q_{n,d}\| = \|h_{d,n}\|_{H(K_d)}.$$

Using properties of the reproducing kernel K_d, we have

$$\|h_{d,n}\|^2_{H(K_d)} = \|h_d\|^2_{H(K_d)} - 2\sum_{j=1}^n a_j h_d(t_j) + \sum_{i,j=1}^n a_i a_j K_d(t_i, t_j).$$

We want to choose coefficients a_j and sample points x_j such that the worst case error of $Q_{n,d}$ is minimized. Let

$$\overline{e^{\mathrm{wor}}}(n, H(K_d)) = \inf\{e^{\mathrm{wor}}(Q_{n,d}; H(K_d)) \,:\, Q_{n,d} \text{ with arbitrary } x_j\},$$

with $a_j = 1/n$, and

$$e^{\mathrm{wor}}(n, H(K_d)) = \inf\{e^{\mathrm{wor}}(Q_{n,d}; H(K_d)) \,:\, Q_{n,d} \text{ with arbitrary } x_j \text{ and } a_j\}.$$

There are many papers on the behaviour of $\overline{e^{\mathrm{wor}}}(n, H(K_d))$ and $e^{\mathrm{wor}}(n, H(K_d))$ for various spaces $H(K_d)$. The special emphasis is on finding sharp estimates of these quantities in terms of n and d. That is, we would like to know how fast they go to zero as n approaches infinity, and what their dependence on d is. In particular, we want to know if they depend polynomially on d or at least non-exponentially on d. We report on such estimates in Section 7 on tractability.

6. Relations between Multivariate Integration and Various Notions of L_2 Discrepancy

In this section we show that B-discrepancy is closely related to multivariate integration defined over a reproducing kernel Hilbert space $H(K_d)$ with the reproducing kernel dependent on B.

Recall that B-discrepancy is defined for measurable sets $B(t) \subseteq \mathbb{R}^d$ for $t \in D \subseteq \mathbb{R}^{\tau(d)}$ for which (2) holds. Recall that

$$D_d = \bigcup_{t\in D} B(t).$$

Note that for the L_2 discrepancy anchored at α, quadrant discrepancy anchored at α or unanchored discrepancy, we have $D_d = [0,1]^d$, whereas for the L_2 ball discrepancy we may have $D_d = \mathbb{R}^d$. Define multivariate integration for functions from $H(K_d)$ with $K_d : D_d \times D_d \to \mathbb{R}$, with kernel K_d yet to be specified, by (21).

This can be written as $I_d(f) = \langle f, h_d\rangle_{H(K_d)}$, with h_d given by (20). Then the worst case error of a linear algorithm (22) is

$$e^{\rm wor}(Q_{n,d}) = \sup_{\substack{f\in H(K_d)\\ \|f\|_{H(K_d)}\le 1}} |I_d(f) - Q_{n,d}(f)| = \|h_{d,n}\|_{H(K_d)},$$

with

$$\|h_{d,n}\|^2_{H(K_d)} = \int_{D_d^2} K_d(x,y)\,\mathrm{d}x\mathrm{d}y - 2\sum_{j=1}^n a_j \int_{D_d} K_d(x,t_j)\,\mathrm{d}x + \sum_{i,j=1}^n a_i a_j K_d(t_i,t_j). \tag{24}$$

We want to make this worst case error equal to the B-discrepancy for the same points t_j and coefficients a_j. If we compare the worst case formula (24) with the formula (3) for the B-discrepancy we see that the candidate for the reproducing kernel is

$$K_d(x,y) = \int_D 1_{B(t)}(x) 1_{B(t)}(y)\rho(t)\,\mathrm{d}t, \quad x,y\in D_d. \tag{25}$$

Observe that $K_d(x,y)$ is well defined and $K_d(x,y)\in[0,1]$. It is easy to check that K_d is a reproducing kernel. Indeed, it is symmetric with $K_d(x,y) = K_d(y,x)$. Consider the $m\times m$ matrix $M = (K_d(x_i,x_j))_{i,j=1,\dots,m}$ for arbitrary points $x_j\in D_d$. Then M is symmetric and it is also positive semi-definite since

$$(Ma,a) = \sum_{i,j=1}^m K_d(x_i,x_j)a_i a_j = \int_D \left(\sum_{j=1}^m a_j 1_{B(t)}(t_j)\right)^2 \rho(t)\,\mathrm{d}t \ge 0.$$

In order to make sure that multivariate integration for $H(K_d)$ is well defined we need to assume that the function

$$h_d(x) = \int_{D_d} K_d(y,x)\,\mathrm{d}y = \int_D \mathrm{vol}(B(t)) 1_{B(t)}(x)\rho(t)\,\mathrm{d}t \tag{26}$$

belongs to $H(K_d)$.

This choice of K_d will make the third terms in (24) and in (3) equal. We obviously need to check that the first and second terms coincide. For the first term of (24), we have

$$\begin{aligned}\int_{D_d^2} K_d(x,y)\,\mathrm{d}x\mathrm{d}y &= \int_D \left(\int_{D_d} 1_{B(t)}(x)\,\mathrm{d}x\right)\left(\int_{D_d} 1_{B(t)}(y)\,\mathrm{d}y\right)\rho(t)\,\mathrm{d}t\\ &= \int_D (\mathrm{vol}(B(t)))^2\rho(t)\,\mathrm{d}t,\end{aligned}$$

which agrees with the first term in (3). For the second term of (24), we have

$$\begin{aligned}\int_{D_d} K_d(x,t_j)\,\mathrm{d}x &= \int_D \left(\int_{D_d} 1_{B(t)}(x)\,\mathrm{d}x\right) 1_{B(t)}(t_j)\rho(t)\,\mathrm{d}t\\ &= \int_D \mathrm{vol}(B(t)) 1_{B(t)}(t_j)\rho(t)\,\mathrm{d}t,\end{aligned}$$

which agrees with the second term in (3).

As before, we can obtain the minimal B-discrepancy $\mathrm{disc}_2^B(n,d)$ and the minimal worst case errors $e^{\rm wor}(n,H(K_d))$ of multivariate integration by taking optimal sample points t_j and optimal coefficients a_j. Clearly $\mathrm{disc}_2^B(n,d) = e^{\rm wor}(n,H(K_d))$.

It is easy to show that the minimal B-discrepancy, or equivalently the minimal multivariate integration errors, is at most of order $n^{-1/2}$ if we consider sets $B(t)$

such that $B(t) \subseteq B_d$ for all $t \in D$ with $B_d \subset \mathbb{R}^d$ and $\mathrm{vol}(B_d) < \infty$. Indeed, take an algorithm

$$Q_{n,d}(f) = \frac{\mathrm{vol}(B_d)}{n} \sum_{j=1}^{n} f(t_j)$$

for some sample points $t_j \in B_d$. Then the square of the worst case error is given by (24) with $a_j = \mathrm{vol}(B_d)/n$, and takes the form

$$\int_{D_d^2} K_d(x,y)\,\mathrm{d}x\mathrm{d}y - \frac{2\,\mathrm{vol}(B_d)}{n} \sum_{j=1}^{n} \int_{D_d} K_d(x,t_j)\,\mathrm{d}x + \frac{\mathrm{vol}^2(B_d)}{n^2} \sum_{i,j=1}^{n} K_d(t_i,t_j).$$

Observe that if $x \notin B_d$ or if $y \notin B_d$, then $K_d(x,y) = 0$. Therefore the above can be rewritten as

$$\int_{B_d^2} K_d(x,y)\,\mathrm{d}x\mathrm{d}y - \frac{2\,\mathrm{vol}(B_d)}{n} \sum_{j=1}^{n} \int_{B_d} K_d(x,t_j)\,\mathrm{d}x + \frac{\mathrm{vol}^2(B_d)}{n^2} \sum_{i,j=1}^{n} K_d(t_i,t_j).$$

Denote this by $f(t_1,\dots,t_n)$. We now compute the average value of f assuming that t_j are independent and uniformly distributed over B_d. Using the standard proof technique which is also used for the study of Monte Carlo algorithms, we obtain

$$\begin{aligned}
&\frac{1}{(\mathrm{vol}(B_d))^n} \int_{B_d^n} f(t_1,\dots,t_d)\,\mathrm{d}t_1 \dots \mathrm{d}t_d \\
&= \frac{1}{n}\left(\mathrm{vol}(B_d) \int_{B_d} K_d(x,x)\,\mathrm{d}x - \int_{B_d^2} K_d(x,y)\,\mathrm{d}x\mathrm{d}y\right) \\
&= \frac{1}{n}\left(\mathrm{vol}(B_d) \int_{D} \mathrm{vol}(B(t))\rho(t)\,\mathrm{d}t - \int_{D} (\mathrm{vol}(B(t)))^2\rho(t)\,\mathrm{d}t\right) \\
&\le \frac{(\mathrm{vol}(B_d))^2}{n}.
\end{aligned}$$

By the mean value theorem, we conclude that there exists at least one choice of the sample points t_j for which the worst case error of $Q_{n,d}$ is at most the square root of the last value. Hence

$$e^{\mathrm{wor}}(n, H(K_d)) \le \frac{1}{\sqrt{n}}\left(\mathrm{vol}(B_d) \int_{B_d} K_d(x,x)\,\mathrm{d}x\right)^{1/2} \le \frac{\mathrm{vol}(B_d)}{\sqrt{n}}.$$

We summarize the analysis of this section in the following theorem.

Theorem.

(i) *The worst case error of $Q_{n,d}$ in the space $H(K_d)$ with the reproducing kernel K_d given by (25) is the same as the B-discrepancy.*

(ii) *We also have*

$$\mathrm{disc}_2^B(n,d) = e^{\mathrm{wor}}(n, H(K_d)).$$

(iii) *If $B(t) \subseteq B_d$ for all $t \in D$ and $\mathrm{vol}(B_d) < \infty$, then*

$$\mathrm{disc}_2^B(n,d) \le n^{-1/2}\,\mathrm{vol}(B_d).$$

It seems interesting to check what kind of reproducing kernels we obtain for various notions of L_2 discrepancy. For the L_2 discrepancy anchored at the origin, we have $D = [0,1]^d$, $\rho(t) = 1$ and $B(t) = [0,t)$. Therefore

$$K_d(x,y) = \prod_{k=1}^{d} \left(1 - \max(x_k, y_k)\right).$$

This corresponds to the Sobolev space anchored at 1.

Consider next the L_2 discrepancy anchored at α. We now have $D = [0,1]^d$ and $\rho(t) = 1$. For $d = 1$, we have $B(t) = [t, \alpha)$ for $t \le \alpha$, and $B(t) = [\alpha, t)$ for $t > \alpha$. A little calculation then gives

$$K_1(x,y) = \min\{x,y\} 1_{[0,\alpha)^2}((x,y)) + (1 - \max\{x,y\}) 1_{[\alpha,1)^2}((x,y)).$$

For $\alpha = 0$, we obtain the previous case, whereas for $\alpha = 1$, we have $K_1(x,y) = \min\{x,y\}$ for $(x,y) \in [0,1)$ and $K(x,y) = 1$ if $\max\{x,y\} = 1$. This corresponds to the Sobolev space anchored at 0; formally for functions defined over $[0,1)$ and with zero value at 1.

Consider now $\alpha \in (0,1)$. Then $H(K_1)$ is the space of functions f defined over $[0,1]$ such that f vanishes at 0 and 1. Furthermore, f restricted to $[0,\alpha)$ is absolutely continuous with $f' \in L_2([0,\alpha))$, and f restricted to $[\alpha,1]$ is absolutely continuous with $f' \in L_2([\alpha,1])$. However, the function f may be *discontinuous* at α. The inner product for $f, g \in H(K_1)$ is

$$\langle f,g\rangle_{H(K_1)} = \int_0^\alpha f'(x)g'(x)\,\mathrm{d}x + \int_\alpha^1 f'(x)g'(x)\,\mathrm{d}x = \int_0^1 f'(x)g'(x)\,\mathrm{d}x.$$

Despite many similarities to the subspace of the Sobolev space, the property that f may be discontinuous at α makes this space different from the Sobolev space.

For $d \ge 1$, we use the tensor product property and obtain

$$\begin{aligned} &K_d(x,y) \\ &\quad= \prod_{k=1}^d \left(\min\{x_k,y_k\} 1_{[0,\alpha_k)^2}((x_k,y_k)) + (1-\max\{x_k,y_k\}) 1_{[\alpha_k,1)^2}((x_k,y_k))\right). \end{aligned}$$

We now turn to the L_2 quadrant discrepancy anchored at α. Again we have $D = [0,1]^d$ and $\rho(t) = 1$. For $d = 1$, we have $B(t) = [0,t)$ for $t < \alpha$, and $B(t) = [t,1)$ for $t \ge \alpha$. A little calculation then gives

$$K_1(x,y) = \frac{|x-\alpha| + |y-\alpha| - |x-y|}{2}.$$

This and the tensor product property of $Q(t)$ yields

$$K_d(x,y) = \prod_{k=1}^d \frac{|x_k-\alpha_k| + |y_k-\alpha_k| - |x_k-y_k|}{2}.$$

Hence, $H(K_d)$ is the Sobolev space anchored at α.

We discuss next the unanchored discrepancy. We have $\tau(d) = 2d$, $D = \{(x,y) \in [0,1]^{2d} : x \le y\}$, $\rho(t) = 1$ and $B(t) = [t_1, t_2)$ for $t = (t_1,t_2)$ with $t_1, t_2 \in [0,1]^d$ and $t_1 \le t_2$. For $d = 1$, a little calculation gives

$$K_1(x,y) = \min\{x,y\} - xy.$$

For $d \ge 1$, using the tensor product property we obtain

$$K_d(x,y) = \prod_{k=1}^d (\min\{x_k,y_k\} - x_k y_k).$$

The space $H(K_d)$ consists of periodic functions which satisfy boundary conditions $f(x) = 0$ if there exists $x_k \in \{0,1\}$ for some $k \in [d]$, and which are one time differentiable with respect to all variables. The inner product for $f, g \in H(K_d)$ is now given by

$$\langle f,g\rangle_{H(K_d)} = \int_{[0,1]^d} \frac{\partial^d}{\partial x_1 \ldots \partial x_d} f(x)\, \frac{\partial^d}{\partial x_1 \ldots \partial x_d} g(x)\,\mathrm{d}x.$$

Observe that the space $H(K_d)$ is a subspace of the Sobolev space $H(K_d^0)$ anchored at 0 or at 1. This result was originally obtained in [39].

We finally turn to the ball discrepancy in the l_∞ case. We now have

$$D = \mathbb{R}^d \times \mathbb{R}_+ = \{[c, r] : c \in \mathbb{R}^d,\ r \geq 0\},$$

and we take $\rho(c, r) = 1$ for $t = [c, r] \in [0, 1]^{d+1}$ and zero otherwise. The sets $B(t)$ are taken as the balls[1]

$$B(t) = \left\{ x \in \mathbb{R}^d : \max_{k \in [d]} |x_k - c_k| \leq r \right\}.$$

Observe that $x \in B(t)$ is equivalent to $x_k - r \leq c_k \leq x_k + r$ for all $k \in [d]$. Hence, $x, y \in B(t)$ and $c \in [0, 1]^d$ yield

$$c_j \in [\max\{0, x_k - r, y_k - r\}, \min\{1, x_k + r, y_k + r\}].$$

This easily implies[2]

$$K_d(x, y) = \int_0^1 \prod_{k=1}^{d} (\min\{1, x_k + r, y_k + r\} - \max\{0, x_k - r, y_k - r\})_+ \, \mathrm{d}r.$$

From this formula we conclude that $K_d(x, y) = 0$ if there exists k such that $x_k \geq 2$ or $y_k \geq 2$. Similarly, $K_d(x, y) = 0$ if there exists k such that $x_k \leq -1$ or $y_k \leq -1$. This means that the space $H(K_d)$ consists of functions that vanish outside $(-1, 2)^d$.

For the *periodic ball discrepancy in the ℓ_∞ case*, we obtain $x \in B(t)$ iff

$$x_j - r \leq c_j \leq x_j + r \quad \text{or} \quad c_j \leq x_j + r - 1 \quad \text{or} \quad c_j \geq 1 - r + x_j \quad \text{for all } j. \tag{27}$$

For given $x_j, y_j \in [0, 1]$ and $r \in [0, 1/2]$, let

$$\ell(x_j, y_j, r) \ = \ \int_0^1 1_{|x_j - c_j|_* \leq r}(x_j) \, 1_{|y_j - c_j|_* \leq r}(y_j) \, \mathrm{d}c_j.$$

Then (27) yields that $\ell(x_j, y_j, r)$ depends only on $\alpha = |x_j - y_j|_*$ and r, i.e., $\ell(x_j, y_j, r) = \ell(\alpha, r)$, and $\ell(\alpha, r) = 0$ if $r \leq \alpha/2$, $\ell(\alpha, r) = 2r - \alpha$ if $\alpha/2 \leq r \leq 1/2 - \alpha/2$, and $\ell(\alpha, r) = -1 + 4r$ if $1/2 - \alpha/2 \leq r \leq 1/2$. Hence, we obtain the reproducing kernel

$$K_d(x, y) = \int_0^{1/2} \prod_{j=1}^{d} \ell(|x_j - y_j|_*, r) \, \tilde{\rho}(r) \, \mathrm{d}r.$$

Observe that $K_d(x, y)$ only depends on the $|x_j - y_j|_*$ and is, in particular, of the form $K_d(x, y) = k_d(x - y)$. In the case $\tilde{\rho} = 2 \cdot 1_{[0,1/2]}$ and $d = 1$ we obtain the kernel $K_1(x, y) = \frac{1}{2} - |x - y|_* + |x - y|_*^2$.

7. Tractability

Tractability of multivariate problems is an active research area in information-based complexity; see the forthcoming book [41]. It would be too much to cover completely the subject of tractability for multivariate integration, or equivalently, for L_2 discrepancy. Instead, we only mention a few recent results mostly, for L_2 (weighted) discrepancy anchored at 0. We hope it will be enough to raise the interest of the readers for this area. We now define a few notions of tractability. We will do it for multivariate integration but due to the intimate relations between multivariate integration and the L_2 discrepancy it will be obvious that the same holds for the L_2 discrepancy.

Recall that $\overline{e^{\mathrm{wor}}}(n, H(K_d))$, and $e^{\mathrm{wor}}(n, H(K_d))$, denote the minimal worst case errors for multivariate integration in the reproducing kernel Hilbert space $H(K_d)$ for optimally chosen sample points and coefficients $a_j = n^{-1}$, and for optimally chosen coefficients a_j, respectively. For simplicity we write $e(n, d)$ to denote either

[1] Note that these "balls" are actually aligned rectangular boxes.

[2] For real number z, we write $z_+ = \max\{z, 0\}$.

of them. For $n = 0$, the two are the same, and $e(0, d) = \|I_d\|$ denotes the initial error.

For the absolute error criterion, we want to find the smallest n for which $e(n, d)$ is at most ε. For the normalized error criterion, we want to find the smallest n for which $e(n, d)$ is at most $\varepsilon e(0, d)$; in other words, we want to reduce the initial error by a factor ε. Let $\mathrm{CRI}_d = 1$ if we consider the absolute error, and $\mathrm{CRI}_d = e(0, d)$ if we consider the normalized error. Let

$$n(\varepsilon, d) = \min\{n : e(n, d) \le \varepsilon \, \mathrm{CRI}_d\}$$

denote the minimal number of sample points which is necessary to solve the problem to within ε. Since each of $e(n, d)$ and CRI_d may take two different values, we have four different cases of $n(\varepsilon, d)$.

Tractability means that $n(\varepsilon, d)$ does *not* depend exponentially on ε and d. There are obviously many different ways to measure the lack of exponential behaviour but we restrict ourselves to only three cases. By the multivariate problem INT $= \{I_d\}_{d=1,2,\ldots}$, we mean multivariate integration I_d defined on the reproducing kernel Hilbert space $H(K_d)$ for varying $d = 1, 2, \ldots$.

We say that INT is *weakly tractable* iff

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\log n(\varepsilon, d)}{\varepsilon^{-1} + d} = 0.$$

Hence $n(\varepsilon, d)$ is much smaller than $a^{\varepsilon^{-1}+d}$ for large $\varepsilon^{-1} + d$, and this holds for any $a > 1$. Weak tractability implies that $n(\varepsilon, d)$ may go to infinity but slower than exponentially in $\varepsilon^{-1} + d$.

We say that INT is *polynomially tractable* iff there are three non-negative numbers C, p, q such that

$$n(\varepsilon, d) \le C\varepsilon^{-p} d^q, \quad \varepsilon \in (0, 1), \quad d = 1, 2, 3, \ldots.$$

Polynomial tractability means that $n(\varepsilon, d)$ may grow not faster than polynomially in ε^{-1} and d. If $q = 0$ in the bound above, so that

$$n(\varepsilon, d) \le C\varepsilon^{-p}, \quad \varepsilon \in (0, 1), \quad d = 1, 2, 3, \ldots,$$

then we say that INT is *strongly polynomially tractable*, and the infimum of p satisfying the last bound is called the *exponent of strong polynomial tractability.*

We also admit different, always non-exponential, behavior of $n(\varepsilon, d)$. As in [22], let $T : [1, \infty) \times [1, \infty) \to [1, \infty)$ be a non-decreasing function of the two arguments such that

$$\lim_{x+y\to\infty} \frac{\log T(x, y)}{x + y} = 0.$$

We say that INT is *T-tractable* iff there are two non-negative numbers C, t such that

$$n(\varepsilon, d) \le C(T(\varepsilon^{-1}, d))^t, \quad \varepsilon \in (0, 1), \quad d = 1, 2, 3, \ldots,$$

and that INT is *strongly T-tractable* iff there are two non-negative number C, t such that

$$n(\varepsilon, d) \le C(T(\varepsilon^{-1}, 1))^t, \quad \varepsilon \in (0, 1), \quad d = 1, 2, 3, \ldots.$$

The infimum of t satisfying the last estimate is called the *exponent of strong T-tractability.*

For $T(x, y) = xy$, polynomial tractability and T-tractability are the same. Interesting choices of T include $T(x, y) = \exp((1 + \log x)(1 + \log y))$, and $T(x, y) = \exp((x + y)^a)$ for any $a \in (0, 1)$. These two examples deal with T tending to infinity faster than a polynomial of any degree.

Polynomial tractability or T-tractability implies weak tractability. The lack of weak tractability implies the lack of polynomial and T-tractability. The lack of weak tractability is called *intractability*.

To distinguish the case when $e(n,d) = \overline{e^{\rm wor}}(n, H(K_d))$, *i.e.* when we use QMC algorithms for approximating the multivariate integrands, instead of tractability, we will use the term *QMC-tractability*. Obviously, QMC-tractability implies tractability. To review our (elaborated) notation, note that we have weak, polynomial, strong polynomial, T and strong T-tractability when we use arbitrary coefficients a_j, and all these concepts for QMC-tractability if we use $a_j = n^{-1}$. Furthermore, all these concepts are defined for the absolute or normalized error criterion.

We are ready to present a number of tractability results for multivariate integration. We first briefly discuss the L_2 discrepancy anchored at 0 which, as we now know, corresponds to multivariate integration for the Sobolev space anchored at 1.

We first consider the absolute error criterion. Note that we can now use the Theorem with $B = [0,1]^d$, and conclude that INT is strongly polynomially QMC-tractable with the exponent at most 2. From (9) we conclude that INT is strongly polynomially tractable with the exponent at most 1.4779. We stress that both the exponents of strong tractability are not known and it is also not known if they are different. By Matoušek's result [33], we know that the exponent of strong polynomial QMC-tractability must be at least 1.0669.

We now consider the normalized error criterion still for the L_2 discrepancy anchored at 0. Since the initial error $3^{-d/2}$ is exponentially small in d, the tractability results are quite different. The bound (14) shown in [39] means that INT is now intractable.

We switch to the L_2 weighted discrepancy defined in Section 4 for a non-zero weight sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ with $\gamma_{d,\mathfrak{u}} \in [0,1]$. It is clear that this discrepancy corresponds to multivariate integration for the weighted Sobolev space anchored at 1. For the absolute error criterion, let

$$f_\gamma(d) = \sum_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} (2^{-|\mathfrak{u}|} - 3^{-|\mathfrak{u}|}),$$

whereas for the normalized error criterion, let

$$f_\gamma(d) = \frac{\displaystyle\sum_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} (2^{-|\mathfrak{u}|} - 3^{-|\mathfrak{u}|})}{\displaystyle\sum_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} 3^{-|\mathfrak{u}|}}.$$

For the unweighted case $\gamma = \{1\}$, the initial error of $(4/3)^{d/2} \ge 1$ indicates that multivariate integration for the absolute error is now much more difficult than for the normalized error. This and (19) yield intractability of INT in both the absolute and normalized error criteria.

For both the error criteria, we must therefore consider decaying weights to obtain tractability. Due to the definition of f_γ depending on the error criteria, we can consider simultaneously the absolute and normalized error criteria. From (18), we conclude that

$$n(\varepsilon, d) \le \left\lceil \frac{f_\gamma(d)}{\varepsilon^2} \right\rceil.$$

Hence

$$\lim_{d\to\infty} \frac{\log f_\gamma(d)}{d} = 0$$

implies weak QMC-tractability of INT, and

$$\limsup_{d\to\infty} \frac{\log f_\gamma(d)}{\log d} < \infty$$

implies polynomial QMC-tractability of INT as well as T-QMC-tractability if we choose $T(x,y) = \exp((1+\log x)(1+\log y))$. For $T(x,y) = \exp((x+y)^a)$ with $a \in (0,1)$, we obtain T-QMC-tractability if

$$\limsup_{d\to\infty} \frac{\log f_\gamma(d)}{d^a} < \infty.$$

Observe also that

$$\sup_d f_\gamma(d) < \infty$$

implies strong polynomial QMC-tractability with the exponent of strong tractability at most 2.

We now consider special weights.

The weights are called *finite-order* weights if there is some integer ω independent of d such that $\gamma_{d,\mathfrak{u}} = 0$ whenever $|\mathfrak{u}| > \omega$; see [18]. Then we may have $O(d^\omega)$ non-zero weights, and this implies $f_\gamma(d) = O(d^\omega)$ for the absolute error criterion, with the factor in the O-notation independent of d and γ. For the normalized error, we have

$$f_\gamma(d) \le \frac{\displaystyle\sum_{\substack{\mathfrak{u}\subseteq[d]\\ |\mathfrak{u}|\le\omega}} \gamma_{d,\mathfrak{u}} 2^{-|\mathfrak{u}|}}{\displaystyle\sum_{\substack{\mathfrak{u}\subseteq[d]\\ |\mathfrak{u}|\le\omega}} \gamma_{d,\mathfrak{u}} 3^{-|\mathfrak{u}|}} = \frac{\displaystyle\sum_{\substack{\mathfrak{u}\subseteq[d]\\ |\mathfrak{u}|\le\omega}} \left(\frac{3}{2}\right)^{|\mathfrak{u}|} \gamma_{d,\mathfrak{u}} 3^{-|\mathfrak{u}|}}{\displaystyle\sum_{\substack{\mathfrak{u}\subseteq[d]\\ |\mathfrak{u}|\le\omega}} \gamma_{d,\mathfrak{u}} 3^{-|\mathfrak{u}|}} \le \left(\frac{3}{2}\right)^\omega.$$

This means that we have polynomial QMC-tractability for the absolute error criterion, and strong polynomial QMC-tractability for the normalized error criterion.

The weights are called *finite-diameter* weights if there is some integer ω independent of d such that $\gamma_{d,\mathfrak{u}} = 0$ whenever $\mathrm{diam}(\mathfrak{u}) \ge \omega$, where

$$\mathrm{diam}(\mathfrak{u}) = \max_{k,m\in\mathfrak{u}} |k-m|,$$

as defined by Creutzig [15]. Finite-diameter weights are a special case of finite-order weights but now we can have only $O(d)$ non-zero weights. Hence $f_\gamma(d) = O(d)$ for the absolute error and $f_\gamma(d) = O(1)$ for the normalized error. Again, we have polynomial QMC-tractability for the absolute error criterion, and strong polynomial QMC-tractability for the normalized error criterion.

For finite-order weights, we know the bounds on the worst case errors of the QMC algorithms using the Niederreiter, Halton or Sobol sample points, see [50]. This implies

$$n(\varepsilon,d) \le d^\tau \frac{(Cd\log d)^\omega}{\varepsilon} (\log\varepsilon^{-1} + \log(Cd\log d))^\omega,$$

where $\tau = \omega$ for the absolute error and $\tau = 0$ for the normalized error, and C is an absolute constant greater than 1, independent of ε^{-1} and d.

Note that modulo logarithms we have the best dependence on ε^{-1} since for $d = 1$, we have $n(\varepsilon,1) = \Omega(\varepsilon^{-1})$. The last bound is especially interesting since the construction of the sample points do *not* depend on the finite-order weights. Still we have only polynomial dependence on d.

We may also use a shifted lattice rule

$$Q_{n,d}(f) = \frac{1}{n}\sum_{j=0}^{n-1} f\left(\left\{\frac{j}{n}z + \Delta\right\}\right),$$

where the generator vector $z \in \{1, \dots, n-1\}^d$ can be computed by the component-by-component (CBC) algorithm with cost $O(dn \log n)$ and $\Delta \in [0,1)^d$; see [43]. Then there exists a vector Δ such that for $n \le C_\alpha \varepsilon^{-2/\alpha} d^{\omega(1-1/\alpha)}$, the worst case error of $Q_{n,d}$ is at most ε for the normalized error criterion. Here $\alpha \in [1,2)$ and C_α is a positive number depending only on α; see [50]. This implies that for the normalized error criterion, we have

$$n(\varepsilon, d) \le C_\alpha \varepsilon^{-2/\alpha} d^{\omega(1-1/\alpha)}.$$

Note that for $\alpha = 1$ we have strong polynomial QMC-tractability, whereas for α close to 2 we have the best possible dependence on ε^{-1} and polynomial dependence on d. However, in this case, the choice of z and Δ depends on the finite-order weights.

We add that finite-order weights imply polynomial or even strong polynomial tractability for many other multivariate linear and selected non-linear problems in the worst case and average case settings. The reader is referred to papers [27, 61, 62, 63, 64, 65] as well to the forthcoming book [41].

We now consider product weights which were the first type of weights studied for multivariate integration and other multivariate problems; see [51]. The weights are called *product weights* if

$$\gamma_{d,\mathfrak{u}} = \prod_{k \in \mathfrak{u}} \gamma_{d,k},$$

where $0 \le \gamma_{d,d} \le \gamma_{d,d-1} \le \dots \le \gamma_{d,1} \le 1$. The essence of product weights is that $\gamma_{d,k}$ moderates the importance of the k-th variable and the groups of $\mathfrak{u}$ variables are moderated by the product of weights of variables from $\mathfrak{u}$. The successive variables are ordered according to their importance, with the first variable being the most important and so on.

For product weights, the initial discrepancy is

$$\prod_{k=1}^{d} \left(1 + \frac{1}{3}\gamma_{d,k}\right)^{1/2}.$$

It is even uniformly bounded in d if

$$\sup_d \sum_{k=1}^{d} \gamma_{d,k} < \infty,$$

and is polynomial in d if

$$\sum_{k=1}^{d} \gamma_{d,k} = O(d^q)$$

for some q. We also have for the absolute error,

$$f_\gamma(d) = \prod_{k=1}^{d} \left(1 + \frac{1}{2}\gamma_{d,k}\right),$$

and for the normalized error,

$$f_\gamma(d) = \prod_{k=1}^{d} \frac{1 + \frac{1}{2}\gamma_{d,k}}{1 + \frac{1}{3}\gamma_{d,k}} \in \left[\prod_{k=1}^{d} \left(1 + \frac{1}{8}\gamma_{d,k}\right), \prod_{k=1}^{d} \left(1 + \frac{1}{6}\gamma_{d,k}\right)\right].$$

This shows that the absolute error criterion is harder than the normalized error criterion.

For the absolute and normalized error criterion, we obtain strong polynomial QMC-tractability if

$$\limsup_{d\to\infty} \sum_{k=1}^{d} \gamma_{d,k} < \infty,$$

and polynomial QMC-tractability if

$$\limsup_{d\to\infty} \frac{1}{\log d} \sum_{k=1}^{d} \gamma_{d,k} < \infty;$$

see [51]. These conditions are also necessary for both the absolute and normalized error criteria for strong polynomial QMC-tractability and polynomial QMC-tractability, for $\gamma_{d,k}$ independent of d; see again [51]. The same conditions are also necessary for strong tractability and polynomial tractability, as proved in [39] for $\gamma_{d,k}$ independent of d, and in [23] for general $\gamma_{d,k}$.

From [23], we have for both the absolute and normalized error criteria that

- weak tractability holds iff

$$\lim_{d\to\infty} \frac{1}{d} \sum_{k=1}^{d} \gamma_{d,k} = 0;$$

- T-tractability holds iff

$$\limsup_{d\to\infty} \frac{1}{\log(1+T(1,d))} \sum_{k=1}^{d} \gamma_{d,k} < \infty, \quad \text{and} \quad \limsup_{\varepsilon^{-1}\to\infty} \frac{\log \varepsilon^{-1}}{\log(1+T(\varepsilon^{-1},1))} < \infty;$$

- strong T-tractability holds iff

$$\limsup_{d\to\infty} \sum_{k=1}^{d} \gamma_{d,k} < \infty, \quad \text{and} \quad \limsup_{\varepsilon^{-1}\to\infty} \frac{\log \varepsilon^{-1}}{\log(1+T(\varepsilon^{-1},1))} < \infty.$$

There are also results relating the exponent of strong tractability to how fast product weights go to zero. The reader is referred to the papers [16, 17, 28, 30, 46, 60]. Details can also be found in [41].

So far we discussed the L_2 discrepancy anchored at 0. Similar results hold for the L_2 discrepancy anchored at α, for the L_2 quadrant discrepancy, and for the unanchored discrepancy. The main technical tool for lower bounds is the property that the corresponding reproducing kernels are decomposable or have finite rank decomposable parts which allows us to use the results from [39]. Details are given in [41].

We finish this section by briefly addressing tractability for the B-discrepancy. For simplicity we consider only the absolute error criterion. As before, let $B(t) \subset B_d$ for all $t \in D$, where $B_d \subset \mathbb{R}^d$ and $\mathrm{vol}(B_d) < \infty$. From the Theorem, it is obvious that we have strong polynomial QMC-tractability with the exponent at most 2 if $\mathrm{vol}(B_d)$ is uniformly bounded in d, and polynomial QMC-tractability if $\mathrm{vol}(B_d)$ is polynomially bounded in d. We leave to the reader the problem of analyzing tractability for the normalized error criterion as well as to generalize the B-discrepancy to the weighted case and study its tractability.

References

[1] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68 (1950), 337–404.

[2] N.S. Bakhvalov. On the optimality of linear methods for operator approximation in convex classes of functions. *Comput. Math. Math. Phys.*, 11 (1971), 244–249.

[3] J. Beck. Irregularities of distribution I. *Acta Math.*, 159 (1987), 1–49.

[4] J. Beck, W.W.L. Chen. *Irregularities of Distribution* (Cambridge Tracts in Mathematics 89, Cambridge University Press, 1987).

[5] J. Beck, W.W.L. Chen. Note on irregularities of distribution II. *Proc. London Math. Soc.*, 61 (1990), 251–272.

[6] J. Beck, V.T. Sós. Discrepancy theory. *Handbook of Combinatorics*, Volume II (R.L. Graham, M. Grötschel, L. Lovász, eds.), pp. 1405–1446 (Elsevier, 1995).

[7] W.W.L. Chen. On irregularities of distribution II. *Q. J. Math.* 34 (1983), 257–279.

[8] W.W.L. Chen. On irregularities of distribution and approximate evaluation of certain functions' *Q. J. Math.*, 36 (1985), 173–182.

[9] W.W.L. Chen. On irregularities of distribution and approximate evaluation of certain functions II. *Analytic Number Theory and Diophantine Problems* (A.C. Adolphson, J.B. Conrey, A. Ghosh, R.I. Yager, eds.) pp. 75–86 ((Progress in Mathematics 70, Birkhäuser, 1987).

[10] W.W.L. Chen. Fourier techniques in the theory of irregularities of point distribution. *Fourier Analysis and Convexity* (L. Brandolini, L. Colzani, A. Iosevich, G. Travaglini, eds.), pp. 59–82 (Applied and Numerical Harmonic Analysis, Birkhäuser, 2004).

[11] W.W.L. Chen, M.M. Skriganov. Explicit constructions in the classical mean squares problem in irregularities of point distribution. *J. Reine Angew. Math.*, 545 (2002), 67–95.

[12] W.W.L. Chen, G. Travaglini. Deterministic and probabilistic discrepancies (submitted).

[13] J.G. van der Corput. Verteilungsfunktionen I. *Proc. Kon. Ned. Akad. v. Wetensch.*, 38 (1935), 813–821.

[14] J.G. van der Corput. Verteilungsfunktionen II. *Proc. Kon. Ned. Akad. v. Wetensch.*, 38 (1935), 1058–1066.

[15] J. Creutzig. Finite-diameter weights (private communication).

[16] L.L. Cristea, J. Dick, F. Pillichshammer. On the mean square weighted L_2 discrepancy of randomized digital nets in prime base. *J. Complexity*, 22 (2006), 605–629.

[17] J. Dick, F. Pillichshammer. On the mean square weighted L_2 discrepancy of randomized digital (t, m, s)-nets over Z_2. *Acta Arith.*, 117 (2005), 371–403.

[18] J. Dick, I.H. Sloan, X. Wang, H. Woźniakowski. Liberating the weights. *J. Complexity*, 20 (2004), 593–623.

[19] M. Drmota, R.F. Tichy. *Sequences, Discrepancies and Applications* (Lecture Notes in Mathematics 1651, Springer-Verlag, 1997).

[20] K. Frank, S. Heinrich. Computing discrepancies of Smolyak quadrature rules. *J. Complexity*, 12 (1996), 287–314.

[21] K.K. Frolov. Upper bounds on the discrepancy in metric L_p, $2 \le p < \infty$. *Dokl. Akad. Nauk SSSR*, 252 (1980), 805–807; English translation in *Soviet Math. Dokl.*, 21 (1980), 840–842.

[22] M. Gnewuch, H. Woźniakowski. Generalized tractability for multivariate problems I: Linear tensor product problems and linear information. *J. Complexity*, 23 (2007), 262–295.

[23] M. Gnewuch, H. Woźniakowski. Generalized tractability for linear functionals. *Proceedings of MCQMC 2006* (to appear).

[24] S. Heinrich. Efficient algorithms for computing the L_2-discrepancy. *Math. Comp.*, 65 (1995), 1621–1633.

[25] F.J. Hickernell. A generalized discrepancy and quadrature error bound. *Math. Comp.*, 67 (1998), 299–322.

[26] F.J. Hickernell, I.H. Sloan, G.W. Wasilkowski. On tractability of weighted integration over bounded and unbounded regions in $\mathbb{R}^s$. *Math. Comp.*, 73 (2004), 1885–1902.

[27] F.J. Hickernell, G.W. Wasilkowski, H. Woźniakowski. Tractability of linear multivariate problems in the average case setting. *Proceedings of MCQMC 2006* (to appear).

[28] F.J. Hickernell, H. Woźniakowski. Integration and approximation in arbitrary dimensions. *Adv. Comput. Math.*, 12 (2000), 25–58.

[29] E. Hlawka. Über die Diskrepanz mehrdimensionaler Folgen mod 1. *Math. Z.*, 77 (1961), 273–284.

[30] F.Y. Kuo. Component-by-component constructions achieve the optimal rate of convergence for multivariate integration in weighted Korobov and Sobolev spaces. *J. Complexity*, 19 (2003), 301–320.

[31] F.Y. Kuo, I.H. Sloan, G.W. Wasilkowski, H. Woźniakowski. Decomposition of functions with finite order (in preparation).

[32] G. Larcher. Digital point sets: analysis and application. *Random and Quasi-Random Point Sets* (P. Hellekalek, G. Larcher, eds), pp. 167–222 (Lecture Notes in Statistics 138, Springer-Verlag, 1998).

[33] J. Matoušek. The exponent of discrepancy is at least 1.0669. *J. Complexity*, 14 (1998), 448–453.

[34] J. Matoušek. *Geometric Discrepancy* (Algorithms and Combinatorics 18, Springer-Verlag, 1999).

[35] H.L. Montgomery. *Ten Lectures on the Interface between Analytic Number Theory and Harmonic Analysis* (CBMS Regional Conference Series in Mathematics 84, American Mathematical Society, 1994).

[36] W.J. Morokoff, R.E. Caflisch. Quasi-random sequences and their discrepancies. *SIAM J. Sci. Comput.*, 15 (1994), 1251–1279.

[37] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods* (CBMS-NSF Regional Conference Series in Applied Mathematics 63, Society for Industrial and Applied Mathematics, 1992).

[38] H. Niederreiter, C. Xing. Nets, (t, s)-sequences, and algebraic geometry. *Random and Quasi-Random Point Sets* (P. Hellekalek, G. Larcher, eds), pp. 267–302 (Lecture Notes in Statistics 138, Springer-Verlag, 1998).

[39] E. Novak, H. Woźniakowski. Intractability results for integration and discrepancy. *J. Complexity*, 17 (2001), 388–441.

[40] E. Novak, H. Woźniakowski. When are integration and discrepancy tractable? *Foundations of Computational Mathematics* (R.A. DeVore, A. Iserles, E. Süli, eds.), pp. 211–266 (London Mathematical Society Lecture Note Series 284, Cambridge University Press, 2001).

[41] E. Novak, H. Woźniakowski. *Tractability of Multivariate Problems* (in progress).

[42] E. Novak, H. Woźniakowski. Relations between L_2 discrepancies and multivariate integration (in progress).

[43] D. Nuyens, R. Cools. Fast algorithms for component-by-component construction of rank-1 lattice rules in shift invariant reproducing kernel Hilbert spaces. *Math. Comp.*, 75 (2006), 903–920.

[44] T. Pillards, B. Vandewoestyne, R. Cools. Minimizing the L_2 and L_∞ star discrepancies of a single point in the unit hypercube. *J. Comput. Appl. Math.*, 197 (2006), 282–285.

[45] L. Plaskota. The exponent of discrepancy of sparse grids is at least 2.1933. *Adv. Comput. Math.*, 12 (2000), 3–24.

[46] L. Plaskota, G.W. Wasilkowski. The exact exponent of sparse grid quadratures in the weighted case. *J. Complexity*, 17 (2001), 840–848.

[47] K.F. Roth. On irregularities of distribution. *Mathematika*, 1 (1954), 73–79.

[48] K.F. Roth. On irregularities of distribution IV. *Acta Arith.*, 37 (1980), 67–75.

[49] I.H. Sloan, S. Joe. *Lattice Methods for Multiple Integration* (Oxford University Press, 1994).

[50] I.H. Sloan, X. Wang, H. Woźniakowski. Finite-order weights imply tractability of multivariate integration. *J. Complexity*, 20 (2004), 46–74.

[51] I.H. Sloan, H. Woźniakowski. When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *J. Complexity*, 14 (1998), 1–33.

[52] V.N. Temlyakov. *Approximation of Periodic Functions* (Nova Science Publishers, 1993).

[53] V.N. Temlyakov. Cubature formulas, discrepancy, and nonlinear approximation. *J. Complexity*, 19 (2003), 352–391.

[54] S. Tezuka. *Uniform Random Numbers: Theory and Practice* (Kluwer Academic Publishers, 1995).

[55] J.F. Traub, G.W. Wasilkowski, H. Woźniakowski. *Information-Based Complexity* (Academic Press, 1988).

[56] G. Travaglini. Average decay of the Fourier transform. *Fourier Analysis and Convexity* (L. Brandolini, L. Colzani, A. Iosevich, G. Travaglini, eds.), pp. 245–268 (Applied and Numerical Harmonic Analysis, Birkhäuser, 2004).

[57] T.T. Warnock. Computational investigations of low discrepancy point sets. *Applications of Number Theory to Numerical Analysis* (S.K. Zaremba, ed.), pp. 319–343 (Academic Press, 1972).

[58] G.W. Wasilkowski, H. Woźniakowski. Explicit cost bounds of algorithms for multivariate tensor product problems. *J. Complexity*, 11 (1995), 1–56.

[59] G.W. Wasilkowski, H. Woźniakowski. The exponent of discrepancy is at most 1.4778 .... *Math. Comp.*, 66 (1997), 1125–1132.

[60] G.W. Wasilkowski, H. Woźniakowski. Weighted tensor-product algorithms for linear multivariate problems. *J. Complexity*, 15 (1999), 402–447.

[61] G.W. Wasilkowski, H. Woźniakowski. Finite-order weights imply tractability of linear multivariate problems. *J. Approx. Theory*, 130 (2004), 57–77.

[62] G.W. Wasilkowski, H. Woźniakowski. Polynomial-time algorithms for multivariate linear problems with finite-order weights: worst case setting. *Found. Comput. Math.*, 5 (2005), 451–491.

[63] G.W. Wasilkowski, H. Woźniakowski. Polynomial-time algorithms for multivariate linear problems with finite-order weights: average case setting. *Found. Comput. Math.* (to appear).

[64] A.G. Werschulz, H. Woźniakowski. Tractability of quasilinear problems I: general results. *J. Approx. Theory*, 145 (2007), 266–285.

[65] A.G. Werschulz, H. Woźniakowski. Tractability of quasilinear problems II: second-order elliptic problems. *Math. Comp.*, 258 (2007), 745–776.

[66] H. Weyl. Über die Gleichverteilung von Zahlen mod Eins. *Math. Ann.*, 77 (1916), 313–352.

[67] H. Woźniakowski. Average case complexity of multivariate integration. *Bull. Amer. Math. Soc. (N. S.)*, 24 (1991), 185–191.

[68] H. Woźniakowski. Efficiency of quasi-Monte Carlo algorithms for high dimensional integrals. *Monte Carlo and Quasi-Monte Carlo Methods 1998* (H. Niederreiter, J. Spanier, eds.), pp. 114–136 (Springer-Verlag, 2000).

[69] S.K. Zaremba. Some applications of multidimensional integration by parts. *Ann. Polon. Math.*, 21 (1968), 85–96.

Mathematisches Institut, Universität Jena, Ernst-Abbe-Platz 2, D-07740 Jena, Germany

E-mail address: novak@mathematik.uni-jena.de

Department of Computer Science, Columbia University, New York, NY 10027, USA, *and* Institute of Applied Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland

E-mail address: henryk@cs.columbia.edu