1

When are Integration and Discrepancy tractable?

Erich Novak

Mathematisches Institut
Universität Erlangen-Nürnberg
Bismarckstraße 1 1/2
91054 Erlangen, Germany, and
Department of Computer Science
KU Leuven
Celestijnenlaan 200A, B-3001 Heverlee, Belgium
Email: novak@mi.uni-erlangen.de

Henryk Woźniakowski

Department of Computer Science
Columbia University
New York, NY 10027, USA, and
Institute of Applied Mathematics
University of Warsaw
ul. Banacha 2, 02-097 Warsaw, Poland
Email: henryk@cs.columbia.edu

1 Introduction

There has been an increasing interest in studying computational aspects of high dimensional problems. Such problems are defined on spaces of functions of d variables and occur in many applications, with d that can be hundreds or even thousands. Examples include:

- *High dimensional integrals* or *path integrals* with respect to the *Wiener measure*. These are important for many applications, in particular, in physics and in finance. High dimensional integrals also occur when we want to compute certain parameters of stochastic processes. Moreover, path integrals arise as solutions of partial differential equations given, for example, by the *Feynman–Kac formula*. See [25, 40, 66, 82, 85, 91].
- *Global optimization* where we need to compute the (global) minimum of a function of d variables. This occurs in many applications, for example, in *pattern recognition* and in *image processing*, see [97], or in the modelling and prediction of the geometry of proteins, see [45]. Simulated annealing strategies and genetic algorithms are often used, as well as smoothing techniques and other stochastic algorithms, see

[10] and [74]. Some error bounds for deterministic and stochastic algorithms can be found in [42, 43, 44, 48, 53].
- The *Schrödinger equation* for $m > 1$ particles in $\mathbb{R}^3$ is a $d = 3m$-dimensional problem.

The solutions of many high dimensional problems are difficult to obtain or even to approximate. Typically, the cost of solving these problems increases exponentially with the dimension d. This is true for numerical integration, optimal recovery, global optimization, and for differential and integral equations over most classical function spaces. However, this is *not* true for systems of ordinary differential equations and for convex optimization, see [85] for a survey. For example, the minimal error for integration of C^r-functions in one variable is of order n^{-r}, where n is the number of function samples used by the algorithm. Using product rules, it is easy to obtain an error of order $n^{-r/d}$ in dimension $d > 1$. It is known that the exponent r/d is best possible. Hence, if we want to guarantee an error ε we must take n of order $\varepsilon^{-d/r}$. For fixed regularity r, this is an exponential function of d. For large d, we cannot afford to sample the function so many times. This exponential dependence is called *intractability* or the *curse of dimensionality*; the latter notion goes back to Bellman, see [7].

It is natural to ask which high dimensional problems are *tractable* and which are *intractable*. Tractability means that there is an algorithm that approximates the solution with error ε using $n = n(\varepsilon, d)$ samples of the function, where $n(\varepsilon, d)$ is polynomially bounded in ε^{-1} and d. If $n(\varepsilon, d)$ has a bound independent of d and polynomially dependent on ε^{-1} then we have *strong tractability*. The number $n(\varepsilon, d)$ of samples is usually directly related to the cost of an algorithm, and therefore tractability means that we can solve the problem to within ε in cost polynomially dependent on ε^{-1} and d. The error ε of an algorithm can be defined in various settings including the worst case, average case, randomised or probabilistic settings.

This definition is similar† to the notion of tractability for discrete problems using the Turing machine as a model of computation. To prove the lack of the existence of polynomial-time algorithms for a number of discrete problems is at the core of theoretical computer science. This is the essence of the famous conjecture P≠NP. According to Steve Smale,

† For discrete problems, d is replaced by the number of input data, and often we want to solve the problem exactly, i.e., with $\varepsilon = 0$. For many discrete problems, intractability is conjectured (often modulo the famous open problem that P≠NP), whereas for high dimensional continuous problems intractability is often *proven*.

see [75], the question of whether P$\neq$NP holds is one of the three most important problems in mathematics‡.

However, we use the *real number model with oracles* † in our study, see [8, 9, 39, 49, 84, 86]. In this model we assume that we can exactly compute arithmetic operations over reals, comparisons of real numbers and function values as oracles. This model is typically used for numerical and scientific computations since it is an abstraction of floating point arithmetic in fixed precision. More about this model of computation can be found in [60, 61, 63, 94, 101].

The subject of tractability and strong tractability for high dimensional problems over the real number model with oracles was introduced in [99] and [100]. Today this topic is being thoroughly studied in information-based complexity and a number of characterisation results have been recently obtained, see [11, 29, 32, 51, 52, 59, 62, 72, 73, 90, 91, 93, 99, 100, 102].

The notion of strong tractability seems very demanding and one may suspect that only trivial problems are strongly tractable. Indeed, many problems defined over classical function spaces are not strongly tractable or even not tractable. Sometimes, the same problems become strongly tractable over function spaces with suitable weights that moderate the behaviour with respect to successive components. This subject will be explained in detail later.

To simplify the exposition we restrict ourselves primarily only to the very important problem of high dimensional integration in the worst case setting. For some classes of functions integration is related to various notions of *discrepancy*, which is widely used and studied in many areas such as number theory, approximation, stochastic analysis, combinatorics, ergodic theory and numerical analysis. The notion of discrepancy is related to Sobolev spaces, Wiener measure, VC dimension and Ramsey theory, see [5, 6, 19, 20, 35, 38, 46, 47, 83, 98].

‡ Smale writes: "In fact, included are what I believe to be the three greatest open problems of mathematics: the Riemann Hypothesis, the Poincaré Conjecture, and 'Does P = NP?' "

† Some problems can be analysed in the Turing machine model as well as in the real number model. Which model is more appropriate depends on the application. A problem can be tractable in one model and intractable in the other, because the assumptions concerning cost and size of a problem are quite different in both models. It is well known that the problem of linear optimization is tractable with respect to the Turing model. It is still unknown, and an important open problem, whether the same is true with respect to the real number model, see [87].
We stress that a conjecture of the type P$\neq$NP has been defined also over the real (or the complex) numbers, see [8] and [9].

This paper is a survey and proofs are only sketched or completely skipped. We will present a few classical results from the theory of numerical quadrature. Most results, however, are quite recent since most classical error estimates do not address the dependence on d. They contain unknown factors that depend on dimension and often increase exponentially with d. If we want to establish tractability, we must know how the error depends on d. Therefore we need to revisit all the classical problems and seek sharp estimates also in terms of d.

Tractability of multivariate integration obviously depends on the class of functions. We will consider a number of classes and present three types of results:

- Classes of functions for which multivariate integration is intractable and suffers the curse of dimensionality. This will be done by establishing exponential lower bounds in d on the number of function samples.
- Classes of functions for which multivariate integration is tractable or strongly tractable. This will be done by proving the *existence* of polynomial-time algorithms by non-constructive arguments.
- Classes of functions for which multivariate integration is tractable or strongly tractable. This will be done by *constructing* polynomial-time algorithms.

For problems belonging to the first case we *prove* intractability. This is in contrast to discrete complexity theory, where it is only *believed*, not proved, that certain problems are difficult (in the sense of exponential lower bounds).

An important area for future research is to move problems from the second to the third case. That is, to provide *constructive* algorithms for problems where we currently have only existence. Obviously, we are especially interested in the case when the construction is simple and does not involve extensive precomputations. Classical algorithms are often based on grids, and the computing time is exponential in the dimension for such algorithms. Only algorithms based on non-grids may have the computing time bounded by a polynomial in the dimension.

2 Multivariate integration and tractability

In this paper we survey results for the following multivariate integration problem. Let F_d be a normed space of integrable functions $f : D_d \to \mathbb{R}$ where $D_d \subset \mathbb{R}^d$. Let $\varrho_d : D_d \to \mathbb{R}_+$ be a weight function such that

$\int_{D_d} \varrho_d(t)\,dt = 1$. For $f \in F_d$ we want to approximate the multivariate integral

$$I_d(f) = \int_{D_d} f(t)\,\varrho_d(t)\,dt. \tag{2.1}$$

Typical examples of such problems are *uniform* integration with $D_d = [0,1]^d$ and $\varrho_d(t) = 1$, and *Gaussian* integration with $D_d = \mathbb{R}^d$ and

$$\varrho_d(t) = (2\pi)^{-d/2}(\sigma_1 \cdots \sigma_d)^{-1} \exp(-t_1^2/(2\sigma_1^2) - \cdots - t_d^2/(2\sigma_d^2))$$

for given positive σ_i.

2.1 Algorithms and Errors

We approximate $I_d(f)$ by algorithms that use finitely many function values. We consider a number of classes of algorithms:

- The class QMC of *quasi-Monte Carlo* algorithms, which is widely used for financial problems. The QMC algorithms are of the form

  $$Q_{n,d}(f) = \frac{1}{n}\sum_{i=1}^{n} f(t_i). \tag{2.2}$$

  The sample points t_i are deterministic, belong to D_d, and may depend on n and d as well as on the weight ϱ_d and the space F_d. The sample points t_i are *nonadaptive*. That is, they are given a priori and do not depend on the integrand f. This is a desirable property for parallel or distributed computations, since the function values $f(t_i)$ can be computed simultaneously by different processors or computers. The weights of QMC algorithms are by definition all equal to $1/n$. This means that QMC algorithms integrate constant functions exactly.
- The class POS of algorithms for which the weights $1/n$ of QMC algorithms are replaced by *positive or non-negative* weights a_i. The POS algorithms are of the form

  $$Q_{n,d}(f) = \sum_{i=1}^{n} a_i f(t_i), \qquad a_i \geq 0. \tag{2.3}$$

  The coefficients a_i are, as the sample points t_i, deterministic and may depend on n, d, ϱ and F_d. Algorithms in the class POS having small error for constant functions are numerically stable, since $\sum_{i=1}^{n} a_i$ is close to 1.

- The class LIN is the class of *linear* algorithms of the form (2.3) with arbitrary real weights a_i. In this case, some of a_i's may be negative. The sample points t_i for this class satisfy the same conditions as for the classes QMC and LIN, i.e., they are nonadaptive. As we shall see the class LIN can sometimes be much more powerful than the class POS.
- The class ARB is the class of *nonlinear* algorithms that use *adaptive* sample points. That is, the sample point t_1 is selected based on our knowledge of the space F_d, and each sample point t_i for $i \geq 2$ may be now dependent on the previously computed function values $f(t_1), \dots, f(t_{i-1})$. Thus, ARB algorithms are of the form

$$Q_{n,d}(f) \,=\, \varphi\left(f(t_1), f(t_2), \dots, f(t_n)\right), \tag{2.4}$$

  where φ is an arbitrary (possible non-linear) mapping from $\mathbb{R}^n$ to $\mathbb{R}$, see e.g., [84] for more details.

The class ARB is the most general of the four. Since we do not restrict in any way how the adaptive choice of the sample points t_i is performed and how we combine the computed function values, one may be afraid that good error properties of the class ARB cannot be computationally realized. As we shall see, the class ARB, despite its generality, is *not* more powerful than the class LIN in many cases.

We are ready to define the error of an algorithm $Q_{n,d}$ for any of these classes. In this paper we restrict ourselves to the worst case error, which is defined by the worst case performance of $Q_{n,d}$ over the unit ball of F_d,

$$e(Q_{n,d}) \,=\, \sup_{f \in F_d,\, \|f\|_d \leq 1} \left|I_d(f) \,-\, Q_{n,d}(f)\right|. \tag{2.5}$$

For $n = 0$ we do not sample the function and we† set $Q_{0,d} = 0$. Then

$$e(Q_{0,d}) \,=\, \sup_{f \in F_d,\, \|f\|_d \leq 1} |I_d(f)| \,=\, \|I_d\|$$

is the *initial error*. This is the a priori error in multivariate integration without sampling the function. We call $e(Q_{n,d})$ the *(absolute) error* of $Q_{n,d}$ and $e(Q_{n,d})/e(Q_{0,d})$ its *normalised error*.

For each class of algorithms we want to find $Q_{n,d}$ having minimal error. Let A be one of the class QMC, POS, LIN or ARB, and let

$$e(n, F_d, \mathrm{A}) \,=\, \inf\left\{\, e(Q_{n,d}) \;:\; Q_{n,d} \,\in\, \mathrm{A} \,\right\} \tag{2.6}$$

† For $n = 0$ we could take $Q_{0,d} = c$ for some number c. It is easy to see that $c = 0$ minimises the error of $Q_{0,d}$.

be the minimal error of algorithms from the class A when we use n function values. Clearly,

$$e(0, F_d, \mathrm{A}) = e(Q_{0,d}) = \|I_d\|, \quad \mathrm{A} \in \{\mathrm{QMC,POS,LIN,ARB}\}.$$

Since the classes are nested we also have

$$e(n, F_d, \mathrm{ARB}) \le e(n, F_d, \mathrm{LIN}) \le e(n, F_d, \mathrm{POS}) \le e(n, F_d, \mathrm{QMC}).$$

Surprisingly enough, we have

$$e(n, F_d, \mathrm{LIN}) = e(n, F_d, \mathrm{ARB}). \tag{2.7}$$

That is, the worst case error is minimised by linear algorithms using non-adaptive sample points. This important result is due to Bakhvalov and Smolyak, see [3].

We first discuss the result of Bakhvalov, who proved that adaption does not help when we approximate linear functionals over balanced and convex sets†. Note that these assumptions obviously hold for multivariate integration (which is a linear functional) and for the unit ball of F_d (which is balanced and convex). It is also known that adaption does not help very much when we approximate linear operators over balanced and convex sets. In this case, adaption may help at most by a factor of 2, as proven in [23] and [86]; the factor 2 cannot be replaced by 1, see [34]. All the assumptions in the theorem about adaption are essential. That is, if the operator is not linear, the set is not balanced, or the set is not convex, then adaption may help; moreover adaption may be significantly better than non-adaption. The reader interested in this issue is referred to the survey [50].

The result of Smolyak is on the power of linear algorithms. He proved that if we approximate a linear functional over a balanced and convex set, then an arbitrary non-linear algorithm must have a worst case error at least as large as that of an appropriately chosen linear algorithm. His theorem has many different generalisations. This subject is beyond the scope of this paper and the reader is referred to [84, 85].

2.2 Tractability

In view of (2.7), we may restrict ourselves to the three classes QMC, POS and LIN. We now formally define tractability and strong tractability for the class $\mathrm{A} \in \{\mathrm{QMC,POS,LIN}\}$. We would like to reduce the initial

† A set B is balanced if $f \in B$ implies that $-f \in B$, and convex if $f_1, f_2 \in B$ implies that $tf_1 + (1-t)f_2 \in B$ for all $t \in [0,1]$.

error by a factor ε, where $\varepsilon \in (0,1)$. We are looking for the smallest $n = n(\varepsilon, F_d, A)$ for which there exists an algorithm from the class A such that $e(Q_{n,d}) \le \varepsilon\, e(Q_{0,d})$. That is,

$$n(\varepsilon, F_d, A) \,=\, \min\,\{\, n \,:\, e(n, F_d, A) \,\le\, \varepsilon\, e(Q_{0,d}) \,\}\,. \tag{2.8}$$

We say that integration for a sequence $\{F_d\}$ of spaces is *tractable* (with respect to the normalised error) in the class A iff there exist nonnegative C, q and p such that

$$n(\varepsilon, F_d, A) \,\le\, C\, d^{\,q}\, \varepsilon^{-p} \qquad \forall\, d = 1, 2, \dots\,, \forall\, \varepsilon \in (0,1). \tag{2.9}$$

Tractability means that we can reduce the initial error† by a factor ε by using a number of function values which is polynomial in d and ε^{-1}. The smallest q and p (or the infima of q and p) in (2.9) are called the *d-exponent* and the *ε-exponent* of tractability in the class A.

If $q = 0$ in (2.9) then we say that integration for a sequence $\{F_d\}$ of spaces is *strongly tractable* in the class A. In this case, the number of samples is independent of d and depends polynomially on ε^{-1}. The smallest (or the infimum of) p in (2.9) is now called the *strong exponent* of tractability in the class A.

We stress that the minimal number $n(\varepsilon, F_d, \mathrm{A})$ of function samples is directly related to the complexity, which is the minimal cost of computing an approximation with error ε. Indeed, this is obvious for the class QMC. In this case, we need only to precompute the sample points $t_1, t_2, \dots, t_n$ with $n = n(\varepsilon, F_d, A)$. For any $f \in F_d$, we compute n function values at t_i, and then perform $n-1$ additions and one division. For the classes POS and LIN we also need to precompute the weights a_i; for each $f \in F_d$, we will need to perform additionally at most n multiplications, instead of one division. The cost $\mathbf{c}$ of one function value is usually much greater than the cost of one arithmetic operation. Therefore, the cost of algorithms from the classes QMC, POS and LIN is roughly equal to $\mathbf{c}\, n$. This means that the complexity is roughly equal to $\mathbf{c}\, n(\varepsilon, F_d, \mathrm{A})$.

2.3 Hilbert Spaces

In general, it is hard to find a formula for the worst case error $e(Q_{n,d})$. One important exception is the case when the space F_d is a reproducing

† It is also of interest to study tractability for the absolute error instead of the normalised error. Then in (2.8) we replace $e(n, F_d, A) \le \varepsilon\, e(Q_{0,d})$ by $e(n, F_d, A) \le \varepsilon$. In general, tractability results for the absolute and the normalised error are different. We shall see such examples later.

kernel Hilbert space. The theory of such spaces can be found in, e.g., [1] and [88]. Here, we recall that a Hilbert space F_d is a reproducing kernel Hilbert space iff the linear functional $L(f) = f(t)$, $\forall f \in F_d$, is continuous for any $t \in D_d$. If so, then there exists a *reproducing kernel* $K_d : D_d \times D_d \to \mathbb{R}$ such that $K_d(\cdot, t) \in F_d$ for any $t \in D_d$ and

$$f(t) \;=\; \langle f, K_d(\cdot,t)\rangle \qquad \forall\, f \in F_d,\; \forall\, t \in D_d.$$

Here, $\langle \cdot, \cdot \rangle$ denotes the inner product of F_d, and $\|f\| = \langle f, f\rangle^{1/2}$. Let

$$h_d(x) \;=\; \int_{D_d} K_d(x,t)\varrho_d(t)\,dt \qquad \forall\, x \in D_d. \tag{2.10}$$

We assume that $h_d \in F_d$. Then the linear functional I_d given by (2.1) is continuous and

$$I_d(f) = \langle f, h_d\rangle\,.$$

The initial error is now given by

$$e(Q_{0,d}) \;=\; \|I_d\| \;=\; \|h_d\| \;=\; \left(\int_{D_d \times D_d} K_d(x,t)\varrho_d(x)\varrho_d(t)\,dt\,dx\right)^{1/2}. \tag{2.11}$$

It is also easy to check the well-known fact that the worst case error of $Q_{n,d}(f) = \sum_{i=1}^n a_i f(t_i)$ is of the form

$$e^2(Q_{n,d}) \;=\; e^2(Q_{0,d}) - 2\sum_{i=1}^n a_i \int_{D_d} K_d(t,t_i)\varrho_d(t)\,dt + \sum_{i,j=1}^n a_i a_j K_d(t_i,t_j). \tag{2.12}$$

2.4 ***The Spaces*** $C^r([0,1]^d)$

We end this section by an example of a sequence $\{F_d\}$ of spaces for which uniform integration is intractable. This example is due Bakhvalov, see [2]. Let $F_d = C^r([0,1]^d)$ be the space of r times continuously differentiable functions with the norm

$$\|f\| = \max_{|\alpha| \le r} \|D^\alpha f\|_\infty.$$

Here $\alpha = [\alpha_1, \dots, \alpha_d]$ is a multi-index of non-negative integers and $|\alpha| = \sum_{i=1}^d \alpha_i$. The operator D^α stands for partial differentiation,

$$D^\alpha f\,(x) \;=\; \frac{\partial^{|\alpha|}\, f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}\,(x) \quad \text{with } x = [x_1, \dots, x_d].$$

It is relatively easy to show that uniform integration

$$I_d(f) = \int_{[0,1]^d} f(x)\,dx, \quad f \in F_d,$$

is intractable. It suffices to show that there are positive numbers $c_{r,d}$ and $C_{r,d}$ such that

$$c_{r,d} \cdot \varepsilon^{-d/r} \le n(\varepsilon, F_d, \mathrm{LIN}) \le C_{r,d} \cdot \varepsilon^{-d/r} \quad \forall\, \varepsilon \in (0,1). \tag{2.13}$$

We sketch the idea of the proof. We consider algorithms of the form

$$Q_{n,d}(f) = \sum_{i=1}^{n} a_i\, f(t_i).$$

The upper bound on $n(\varepsilon, F_d, \mathrm{LIN})$ can be proved with product formulas, for example, we may take products of Gaussian rules or products of the Clenshaw–Curtis formulas. Then for $n = O(\varepsilon^{-d/r})$, the error is at most ε.

For the lower bound, we consider an arbitrary $Q_{n,d}$ with $2n = \ell^d$, where $\ell \in \mathbb{N}$ and n is of order $\varepsilon^{-d/r}$. We take the partition of $[0,1]^d$ into ℓ^d cubes having edge size ℓ^{-1}. Then there are at least n small cubes without a sample point, hence $Q_{n,d}(f)$ cannot depend on the behaviour of $f \in F_d$ over these unsampled cubes. We then can construct two functions $f_1, f_2 \in F_d$ such that they only differ on the unsampled cubes and for which

$$I_d(f_1) - I_d(f_2) \ge 2\,\varepsilon$$

Since $Q_{n,d}(f_1) = Q_{n,d}(f_2)$ we obtain

$$e(Q_{n,d}) \ge \varepsilon$$

and (2.13) is proven, see e.g. [48] for the details.

We comment on this typical result and its proof. For any d, we obtain the rate

$$n(\varepsilon, F_d, \mathrm{LIN}) \asymp \varepsilon^{-d/r}$$

constructively. In fact, it is easy to write down corresponding product rules with error at most ε in an explicit way. Nontrivial product rules need, however, at least 2^d sample points. As long as d is small this is not a problem. On the other hand, if $d = 100$ this can hardly be called a meaningful construction. This is where the curse of dimensionality strikes. Since r is fixed, the exponent d/r in (2.13) goes to infinity with d. Hence, $n(\varepsilon, d, \mathrm{LIN})$ does *not* depend polynomially on ε^{-1}; that is, the problem is intractable.

Assume now that the smoothness parameter r depends on d, say $r = d$. Then the power of ε^{-1} in (2.13) is just 1 and the dependence on ε^{-1} is polynomial. How about the dependence on d? To claim tractability it would be enough to prove that $C_{r,d}$ depends polynomially on d, and to claim intractability it would be enough to prove that $c_{r,d}$ is exponentially large in d. If one checks the proof of (2.13) more carefully then one will notice that, unfortunately, $C_{r,d}$ is exponentially large in d, and $c_{r,d}$ is exponentially small in d. In fact, the gap between $C_{r,d}$ and $c_{r,d}$ is something like

$$\frac{C_{r,d}}{c_{r,d}} \approx 5^{d-1}.$$

This is a typical situation, in which the classical analysis is not enough to claim either tractability or intractability. In fact, tractability of uniform integration for the class $C^r([0,1]^d)$ when $r = d$, or more generally, when $r/d \leq c$ for some positive c, is still open.

3 Discrepancy

The notion of *discrepancy* goes back to the work of Weyl in 1916, see [96], and van der Corput in the thirties, see [14] and [15].

Discrepancy is a quantitative measure of the uniformity of the distribution of points in the d-dimensional unit cube. 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 [5, 6, 19, 38, 46, 71] and [81]. In this paper we limit ourselves to a few basic notions of discrepancy, and show their relation to uniform integration

$$I_d(f) = \int_{[0,1]^d} f(t)\, dt.$$

In Sections 3.1 to 3.3 we study the classical L_2-discrepancy, which is sometimes also called L_2-star discrepancy or L_2-star discrepancy with boundary condition. We first recall the definition of the L_2-discrepancy and then present its absolute and normalised bounds. In Section 3.4 we discuss the weighted L_2-discrepancy (without boundary conditions) and in Section 3.5 we explain its relation to errors for integration in weighted Sobolev classes. Then we discuss the L_p-star discrepancy for finite p, see Section 3.6, and the $*$-discrepancy, where $p = \infty$, see Section 3.7.

For all these discrepancies, boxes of the form $[0, x)$ are used as test-sets, see, for example, the classical L_2-discrepancy (3.2). One can avoid this unsymmetric situation and can study notions of symmetric discrepancy, see [30]. We discuss two such cases: the extreme or unanchored discrepancy in Section 3.8, and the centred discrepancy in Section 3.9.

3.1 L_2*-Discrepancy*

Let $x = [x_1, \ldots, x_d] \in [0,1]^d$. By the box $[0, x)$ we mean the set $[0, x_1) \times \cdots \times [0, x_d)$ whose (Lebesgue) volume is clearly $x_1 \cdots x_d$. For given points $t_1, \ldots, t_n \in [0,1]^d$, we approximate the volume of $[0, x)$ by the fraction of the points t_i which are in the box $[0, x)$. The error of such an approximation is

$$x_1 \cdots x_d \; - \; \frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(t_i),$$

where $1_{[0,x)}(t_i)$ is the indicator (characteristic) function, which is equal to 1 if $t_i \in [0, x)$, and to 0 otherwise.

Observe that we use equal coefficients n^{-1} in the previous approximation scheme. As we shall see, this corresponds to QMC algorithms for integration. We generalise this approach by allowing arbitrary coefficients a_i instead of n^{-1}. That is, we approximate the volume of $[0, x)$ by the weighted sum

$$\mathrm{disc}(x) \; := \; x_1 \cdots x_d \; - \; \sum_{i=1}^{n} a_i 1_{[0,x)}(t_i). \tag{3.1}$$

The L_2-discrepancy of points $t_1, \ldots, t_n$ and coefficients $a_1, \ldots, a_n$ is just the L_2-norm of the error function (3.1),

$$\mathrm{disc}_2(\{t_i\}, \{a_i\}) \; = \; \left(\int_{[0,1]^d} \left(x_1 \cdots x_d \; - \; \sum_{i=1}^{n} a_i 1_{[0,x)}(t_i) \right)^2 dx \right)^{1/2}. \tag{3.2}$$

By direct integration, we have the explicit formula

$$\mathrm{disc}_2^2(\{t_i\}, \{a_i\}) \; = \tag{3.3}$$
$$\frac{1}{3^d} - \frac{1}{2^{d-1}} \sum_{i=1}^{n} a_i \prod_{k=1}^{d} (1 - t_{i,k}^2) + \sum_{i,j=1}^{n} a_i a_j \prod_{k=1}^{d} (1 - \max(t_{i,k}, t_{j,k})) \, ,$$

for the L_2-discrepancy, where $t_i = [t_{i,1}, \ldots, t_{i,d}]$.

Hence, $\mathrm{disc}_2(\{t_i\}, \{a_i\})$ can be computed with a cost of $O(dn^2)$ arithmetic operations. Faster algorithms for computing L_2-discrepancy for relatively small d have been found by Heinrich, see [28] and also [20].

The major problem of L_2-discrepancy is to find points $t_1, \dots, t_n$ and coefficients $a_1, \dots, a_n$ for which $\mathrm{disc}_2(\{t_i\}, \{a_i\})$ is minimised. Let

$$\overline{\mathrm{disc}}_2(n,d) = \inf_{t_1,\dots,t_n} \mathrm{disc}_2(\{t_i\}, \{n^{-1}\})$$

and

$$\mathrm{disc}_2(n,d) = \inf_{t_1,\dots,t_n,\, a_1,\dots,a_n} \mathrm{disc}_2(\{t_i\}, \{a_i\})$$

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

Observe that for $n = 0$ we do not use the points t_i or the coefficients a_i, and obtain the initial L_2-discrepancy

$$\overline{\mathrm{disc}}_2(0,d) \;=\; \mathrm{disc}_2(0,d) \;=\; \left(\int_{[0,1]^d} x_1^2 \cdots x_d^2 \, dx \right)^{1/2} \;=\; 3^{-d/2}. \qquad (3.4)$$

Hence, the initial L_2-discrepancy is exponentially small in d.

3.2 Bounds for the (absolute) L_2-Discrepancy

We now discuss bounds on the minimal (absolute) L_2-discrepancy, which are related to tractability of integration for the absolute error. For a fixed d, the asymptotic behaviour of the minimal L_2-discrepancy as a function of n is known. There exist positive numbers c_d and C_d such that

$$c_d \, \frac{\log^{(d-1)/2} n}{n} \;\le\; \mathrm{disc}_2(n,d) \;\le\; \overline{\mathrm{disc}}_2(n,d) \;\le\; C_d \, \frac{\log^{(d-1)/2} n}{n} \qquad (3.5)$$

for all $n \ge 2$. The lower bound was proven by Roth in 1954 for $a_i = n^{-1}$, see [69], but essentially the same proof covers arbitrary a_i, see [12] and [13]. The upper bound was proven by Roth and Frolov in 1980, see [70] and [22].

The bound (3.5) justifies the definition of *low-discrepancy* sequences (and points which we do not cover here), for the coefficients a_i equal n^{-1}. Namely, the sequence $\{t_i\}$ is a low-discrepancy sequence if there is a positive number C_d such that

$$\mathrm{disc}_2(\{t_i\}, \{n^{-1}\}) \;\le\; C_d \, \frac{\log^d n}{n} \qquad \forall\, n \ge 2. \qquad (3.6)$$

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 a 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 points as their counterparts for the periodic case, see [5, 6, 19, 38, 46, 71, 81]. There are also points and sequences satisfying (3.6) with more general coefficients a_i than n^{-1} in (3.1). An example is provided by hyperbolic points, see [80, 90], although in this case we have $\log^{1.5\,d} n$ instead of $\log^d n$ in (3.6). Explicit bounds on the L_2-discrepancy for hyperbolic points can be found in [90]. In particular, hyperbolic points $t_1, \dots, t_n$ and coefficients $a_1, \dots, a_n$ were constructed such that $\mathrm{disc}_2(\{t_i\}, \{a_i\}) \le \varepsilon$ with

$$\begin{aligned} n \quad &\le \quad 3.304 \left(1.77959 + 2.714 \frac{-1.12167 + \log\, \varepsilon^{-1}}{d-1}\right)^{1.5(d-1)} \frac{1}{\varepsilon} \\ &\le \quad 7.26 \left(\frac{1}{\varepsilon}\right)^{2.454}. \end{aligned}$$

The essence of (3.5) 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 $\log^{(d-1)/2} n$ is essential. Indeed, the function $n^{-1} \log^{(d-1)/2} n$ is increasing for $n \le \exp\left((d-1)/2\right)$. 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 behaviour of $\mathrm{disc}_2(n, d)$ is dubious for practical purposes.

Obviously if d is large, then the numbers c_d and C_d from (3.5) are also very important. We do not know much about them. However, we know that the asymptotic constant

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

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_i and for

$a_i = n^{-1}$, we have

$$\overline{\mathrm{disc}}_2(n,d) \le \left(\int_{[0,1]^{nd}} \mathrm{disc}_2^2(\{t_i\},\{n^{-1}\})\, dt_1 \cdots dt_n \right)^{1/2} \le \frac{2^{-d/2}}{n^{1/2}}. \tag{3.7}$$

The last estimate looks very promising since we have an exponentially small dependence on d through $2^{-d/2}$. 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

$$A_c = \left\{ (t_1,\dots,t_n) \,:\, \mathrm{disc}_2(\{t_i\},\{n^{-1}\}) \le c\, 2^{-d/2} n^{-1/2} \right\}$$

has Lebesgue measure at least $1 - c^{-2}$. Hence, for $c = 10$ we have a set of points $t_1,\dots,t_n$ 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, and compute their discrepancy with $a_i = n^{-1}$. 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 $t_1,\dots,t_n$. If random samples are chosen with uniform distribution, then after a few such selections we will get the desired points.

3.2.1 The Exponent of L_2-Discrepancy

We now briefly discuss the exponent p^* of the absolute L_2-discrepancy, see [92]. This is defined as the smallest number (or the infimum) of positive p for which there exists a positive C such that

$$\mathrm{disc}_2(n,d) \le C\, n^{-1/p} \qquad \forall\, n, d = 1, 2, \dots . \tag{3.8}$$

We stress that the last estimate holds for all n and d; hence both C and p do not depend on n and d. The bound $p^* \ge 1$ is obvious, since for $d = 1$ we have

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

For $a_i = n^{-1}$, it is proven in [37] that p in (3.8) must be at least 1.0669. This means that the case of arbitrary d is harder and the presence of the logarithmic factors in (3.5) cannot be entirely neglected.

The upper bound

$$p^* \le 1.4779$$

is proven in [92]. The proof of this upper bound is *non-constructive.* The best constructive bound currently known is given in [90], and states that

$$\mathrm{disc}_2(n,d) \le \frac{2.244}{n^{0.408}} \qquad \forall\, n,d,$$

which corresponds to $p = 2.454$. This bound is obtained by hyperbolic points.

Recently, it was proven in [65] that using hyperbolic points or more generally nested sparse grids points leads to $p \ge 2.1933$. Hence, to obtain $p < 2.1933$ we must use other than 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_i for which (3.8) holds with $p < 2$.

3.3 Normalised L_2-Discrepancy

We now discuss bounds on the normalised L_2-discrepancy. By the normalised L_2-discrepancy we mean $\mathrm{disc}_2(\{t_i\},\{a_i\})/\mathrm{disc}_2(0,d)$. That is, we normalise by the initial value of the L_2-discrepancy, which is $3^{-d/2}$. This case is directly related to tractability of integration when we reduce the initial error by a factor of ε. Similarly to (2.8), we define

$$\begin{aligned} \overline{n}(\varepsilon,d) &= \min\{\, n \,:\, \overline{\mathrm{disc}}_2(n,d) \le \varepsilon\,\mathrm{disc}_2(0,d)\,\}, && (3.9)\\ n(\varepsilon,d) &= \min\{\, n \,:\, \mathrm{disc}_2(n,d) \le \varepsilon\,\mathrm{disc}_2(0,d)\,\} && (3.10)\end{aligned}$$

and ask whether $\overline{n}(\varepsilon,d)$ and $n(\varepsilon,d)$ behave polynomially in ε^{-1} and d. We wish to stress that polynomial bounds on the absolute value of the L_2-discrepancy, which we presented so far, are useless for the normalised 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 recently solved. First of all, notice that it directly follows from (3.4) and (3.7) that

$$\overline{n}(\varepsilon,d) \le 1.5^d\,\varepsilon^{-2}. \tag{3.11}$$

It was proven in [102], see also [73] and Section 4.1, that

$$\overline{n}(\varepsilon,d) \ge (9/8)^d\,(1-\varepsilon^2). \tag{3.12}$$

The bound (3.12) is also valid if $a_i \ge 0$, which corresponds to the assumption defining the class POS. Hence, we have exponential dependence on d.

For arbitrary a_i, it was proven in [62] that

$$1.0628^d\,(1+o(1)) \le n(\varepsilon,d) \le 1.5^d\,\varepsilon^{-2} \qquad \text{as } d\to\infty, \tag{3.13}$$

the lower bound holding for any fixed $\varepsilon < 1$. Hence in the general case, $n(\varepsilon,d)$ goes to infinity faster than any power of d and the normalised L_2-discrepancy is intractable. The upper bounds in (3.11) and in (3.13) coincide. We do not know whether arbitrary weights are better than positive or equal weights.

3.4 Weighted L_2-Discrepancy

We now discuss a slightly different L_2-discrepancy for functions without a boundary condition, see Section 3.5 and [30]. The *weighted* version of L_2-discrepancy was introduced in [73], and plays a major role for the study of tractability in weighted Sobolev spaces, as will be explained in the next subsection.

We are given a sequence $\gamma = \{\gamma_j\}$ of weights such that

$$\gamma_1 \;\ge\; \cdots \;\ge\; \gamma_j \;\ge\; \cdots \;\ge\; 0.$$

For any subset u of $U_d = \{1,2,\dots,d\}$, we define $\gamma_\emptyset = 1$ and $\gamma_u = \prod_{j\in u}\gamma_j$ for non-empty u. For a vector $x\in[0,1]^d$, we denote x_u as the vector from $[0,1]^{|u|}$, where $|u|$ is the cardinality of u, with the components of x whose indices are in u. For example, for $d=5$ and $u=\{2,4,5\}$ we have $x_u = [x_2,x_4,x_5]$. Then $dx_u = \prod_{j\in u} dx_j$. By $(x_u,1)$ we mean the vector from $[0,1]^d$ with the same components as x for indices in u and with the rest of components being replaced by 1. For our example, we have $(x_u,1) = [1,x_2,1,x_4,x_5]$. Recall that for given points $t_1,\dots,t_n$ and coefficients $a_1,\dots,a_n$, the function $\mathrm{disc}_2(x_u,1)$ is given by (3.1) and takes the form

$$\mathrm{disc}_2(x_u,1) \;=\; \prod_{k\in u} x_k \;-\; \sum_{i=1}^n a_i 1_{[0,x_u)}\left((t_i)_u\right).$$

The *weighted* L_2-discrepancy is then defined as

$$\mathrm{disc}_{2,\gamma}(\{t_i\},\{a_i\}) \;=\; \left(\sum_{u\subset U_d}\gamma_u\int_{[0,1]^{|u|}}\mathrm{disc}_2^2(x_u,1)\,dx_u\right)^{1/2}. \tag{3.14}$$

It is easy to obtain the explicit formula

$$\mathrm{disc}_{2,\gamma}^2(\{t_i\},\{a_i\}) \;=$$

$$\sum_{u\subset U_d} \gamma_u \Bigg(\frac{1}{3^{|u|}} - \frac{1}{2^{|u|-1}} \sum_{i=1}^{n} a_i \prod_{k\in u} (1-t_{i,k}^2) \; + $$
$$\sum_{i,j=1}^{n} a_i a_j \prod_{k\in u} (1-\max(t_{i,k},t_{j,k}))\Bigg)$$

for the weighted L_2-discrepancy. The classical (unweighted) case corresponds to $\gamma = 1$, i.e., $\gamma_j = 1$. Since

$$\begin{aligned} \sum_{u\subset U_d} 3^{-|u|} &= \sum_{k=0}^{d} \binom{d}{k} 3^{-k} = (1+3^{-1})^d, \\ \sum_{u\subset U_d} \prod_{k\in u} \frac{1-t_{i,k}^2}{2} &= \prod_{k=1}^{d} \left(1 + \frac{1-t_{i,k}^2}{2}\right), \end{aligned}$$

we have

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

The square of the weighted L_2-discrepancy is defined as the sum of 2^d terms. One might expect that the cost of computing $\operatorname{disc}_{2,\gamma}(\{t_i\},\{a_i\})$ is exponential in d. Surprisingly, this is *not* the case. It is shown in [33] that the weighted discrepancy can be computed with cost of order $d\,n^2$.

As before, let

$$\begin{aligned} \overline{\operatorname{disc}}_{2,\gamma}(n,d) &= \inf_{t_1,\dots,t_n} \operatorname{disc}_{2,\gamma}(\{t_i\},\{n^{-1}\}), && (3.15) \\ \operatorname{disc}_{2,\gamma}(n,d) &= \inf_{t_1,\dots,t_n\; a_1,\dots,a_n} \operatorname{disc}_{2,\gamma}(\{t_i\},\{a_i\}) && (3.16) \end{aligned}$$

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

$$\overline{\operatorname{disc}}_{2,\gamma}^2(n,d) = \operatorname{disc}_{2,\gamma}^2(n,d) = \sum_{u\subset U_d} \gamma_u 3^{-|u|} = \prod_{j=1}^{d} \left(1 + \frac{1}{3}\gamma_j\right).$$

Observe that for the unweighted case, $\gamma_j = 1$, we have $\operatorname{disc}_2(0,d) = (4/3)^d$, which is exponentially large in d. On the other hand, the initial weighted L_2-discrepancy is uniformly bounded in d iff $\sum_{j=1}^{\infty} \gamma_j < \infty$.

Let us define

$$\begin{aligned} \overline{n}_\gamma(\varepsilon,d) &= \min\{\, n : \overline{\operatorname{disc}}_{2,\gamma}(n,d) \le \varepsilon \operatorname{disc}_{2,\gamma}(n,d)\,\}, && (3.17) \\ n_\gamma(\varepsilon,d) &= \min\{\, n : \operatorname{disc}_{2,\gamma}(n,d) \le \varepsilon \operatorname{disc}_{2,\gamma}(n,d)\,\} && (3.18) \end{aligned}$$

as the minimal n for which we reduce the initial weighted L_2-discrepancy by a factor of ε.

The following bounds on $\overline{n}_\gamma(\varepsilon,d)$ were proven in [73],

$$1.055^{s_d}\,(1-\varepsilon^2)\ \le\ \overline{n}_\gamma(\varepsilon,d)\ \le\ 1.1836^{s_d}\varepsilon^{-2}\,-\,\varepsilon^{-2}\qquad \forall\,\varepsilon\in(0,1),\ d\ge 1. \tag{3.19}$$

Here, $s_d = \sum_{j=1}^d \gamma_j$ is the partial sum of the weights. From these estimates it is easy to determine when $\overline{n}_\gamma(\varepsilon,d)$ depends polynomially on d. It was proven in [62] that the case of arbitrary a_i leads to the same conclusion and we have:

- $n_\gamma(\varepsilon,d)$ is uniformly bounded in d iff $\sum_{j=1}^\infty \gamma_j < \infty$,
- $n_\gamma(\varepsilon,d)$ is polynomially bounded in d iff $\displaystyle\limsup_{d\to\infty}\frac{\sum_{j=1}^d \gamma_j}{\log\,d} < \infty$.

Hence, for the unweighted case, $\gamma_j = 1$, we do not have polynomial dependence on d. We obtain

$$1.0563^d\,(1-\varepsilon^2)\ \le\ \overline{n}_\gamma(\varepsilon,d)\ \le\ 1.125^d\,\varepsilon^{-2}, \tag{3.20}$$

for the case of equal (or positive) weights and

$$1.0463^d\,(1+o(1)) \le n_\gamma(\varepsilon,d) \le 1.125^d\,\varepsilon^{-2}\qquad \text{as } d\to\infty, \tag{3.21}$$

for arbitrary weights. The upper bound is proved by an average case argument, as in Section 3.3. The lower bound in (3.20) follows from the technique in [102], see Section 4.1. The lower bound in (3.21) is proved in [62], see Section 4.2.

On the other hand, for weights γ_j for which the series $s_\gamma = \sum_{j=1}^\infty \gamma_j$ is finite, there is no dependence on d. For example, take $\gamma_j = j^{-2}$. Then $s_\gamma = \pi^2/6$ and

$$\overline{n}_\gamma(\varepsilon,d)\ \le\ 0.3195\,\varepsilon^{-2}.$$

If $a := \limsup_{d\to\infty}\sum_{j=1}^d \gamma_j/\log\,d < \infty$ then there exists a positive C such that

$$\overline{n}_\gamma(\varepsilon,d)\ \le\ C\,d^{a/6}\varepsilon^{-2}.$$

The exponent p_γ^* of the weighted normalised L_2-discrepancy can be defined analogously to the exponent of the L_2-discrepancy. Namely, p_γ^* is defined as the smallest number (or infimum) of positive p for which there exists a positive C such that

$$n_\gamma(\varepsilon,d)\ \le\ C\,\varepsilon^{-p}\qquad \forall\,\varepsilon\in(0,1),\ d\ge 1.$$

Then

$$p_\gamma^* \text{ exists iff } \sum_{j=1}^{\infty} \gamma_j < \infty.$$

Furthermore, if $\sum_{j=1}^{\infty} \gamma_j < \infty$ then $p_\gamma^* \in [1,2]$. It is also known that $\sum_{j=1}^{\infty} \gamma_j^{1/2} < \infty$ implies that $p_\gamma^* = 1$. The proof is non-constructive and is given in [32]. It is not known whether the condition $\sum_{j=1}^{\infty} \gamma_j^{1/2} < \infty$ is necessary for $p_\gamma^* = 1$. We are inclined to believe that it is.

There is a constructive proof that $p_\gamma^* = 1$ in [93] under the assumption that $\sum_{j=1}^{\infty} \gamma_j^{1/3} < \infty$. The optimal exponent is achieved by a weighted Smolyak algorithm using hyperbolic points, see Section 6.

3.4.1 Limiting Discrepancy

We end this subsection by a remark on the *limiting* discrepancy which is formally defined for $d = \infty$, see [73]. For points $t_i^{(\infty)} = [t_{i,1}, t_{i,2}, \dots,] \in [0,1]^\infty$ let $t_i^{(d)} = [t_{i,1}, \dots, t_{i,d}] \in [0,1]^d$ denote their d-dimensional projections. Then $\mathrm{disc}_{2,\gamma}(\{t_i^{(d)}, \{n^{-1}\})$ is a non-decreasing function of d. The limiting discrepancy is then defined as

$$\mathrm{disc}_{2,\gamma}(\{t_i^{(\infty)}\}) = \lim_{d\to\infty} \mathrm{disc}_{2,\gamma}(\{t_i^{(d)}\}, \{n^{-1}\}).$$

We have

$$\mathrm{disc}_{2,\gamma}(\{t_i^{(\infty)}\}) < \infty \qquad \text{iff} \qquad \sum_{j=1}^{\infty} \gamma_j < \infty.$$

We stress that the last condition holds independently of the sample points $t_i^{(\infty)}$. We see once more that the condition $\sum_{j=1}^{\infty} \gamma_j < \infty$ is needed to have a finite limiting discrepancy. Error bounds for quasi-Monte Carlo algorithms in an infinite dimensional setting are also studied in [89].

3.5 L_2-Discrepancy and Errors for Integration

We now show how the different L_2-discrepancies are related to uniform integration for Sobolev spaces and the Wiener sheet measure.

3.5.1 Spaces with Boundary Values

Consider the classical Sobolev space $W_2^1([0,1])$ of absolute continuous univariate functions $f : [0,1] \to \mathbb{R}$ for which $f' \in L_2([0,1])$ and $f(0) = 0$.

The space $W_2^1([0,1])$ is equipped with the inner product

$$\langle f,g\rangle \;=\; \int_0^1 f'(x)g'(x)\,dx.$$

The Sobolev space $W_2^1([0,1])$ is a reproducing kernel Hilbert space with the reproducing kernel

$$K_1(x,t) \;=\; \min(x,t).$$

Take now F_d as the tensor product of $W_2^1([0,1])$,

$$F_d = W_2^{1,\dots,1}([0,1]^d) \;:=\; W_2^1([0,1])\otimes\,\cdots\,\otimes W_2^1([0,1]), \quad (d\,\text{times}).$$

Then F_d is the Sobolev space of multivariate functions $f:[0,1]^d\to\mathbb{R}$ that are differentiable once with respect to each variable, and for which $f(x)=0$ if at least one component of x is zero. It is a reproducing kernel Hilbert space with the reproducing kernel

$$K_d(x,t) \;=\; \prod_{j=1}^d \min(x_j,t_j), \tag{3.22}$$

and the inner product is given by

$$\langle f,g\rangle \;=\; \int_{[0,1]^d} \frac{\partial^d f}{\partial x_1\cdots\partial x_d}\,(x)\,\frac{\partial^d g}{\partial x_1\cdots\partial x_d}\,(x)\;dx.$$

Since F_d is a Hilbert space and I_d is a continuous and linear functional we have by Riesz's theorem,

$$I_d(f) \;=\; \langle f,h_d\rangle\,, \qquad \text{with} \qquad h_d(x)=\int_{[0,1]^d} K_d(x,t)\,dt.$$

Clearly, the initial error satisfies $e(Q_{0,d})=\|h_d\|$ and we have

$$e(Q_{0,d}) \;=\; \left(\int_{[0,1]^{2d}} K_d(x,t)\,dx\,dt\right)^{1/2} \;=\; 3^{-d/2}.$$

Hence, the initial error is the same as the initial L_2-discrepancy.

Let $Q_{n,d}(f)=\sum_{i=1}^n a_i f(t_i)$ be an algorithm for approximating $I_d(f)$. Consider now the worst case error of $Q_{n,d}$, which is given by (2.12) with K_d defined by (3.22). By simple integration we get

$$e^2(Q_{n,d}) \;=\; \frac{1}{3^d} - 2\sum_{i=1}^n a_i \prod_{k=1}^d (t_{i,k}-t_{i,k}^2/2) + \sum_{i,j=1}^n a_i a_j \prod_{k=1}^d \min(t_{i,k},t_{j,k}).$$

Since for any numbers a, b we have

$$\min(a,b) = 1 - \max(1-a, 1-b) \quad \text{and} \quad a - a^2/2 = \left(1-(1-a)^2\right)/2,$$

comparing the last formula with the L_2-discrepancy formula (3.1), we immediately obtain

$$e(Q_{n,d}, F_d) = \mathrm{disc}_2(\{1-t_i\}, \{a_i\}),$$

where $1-t_i = [1-t_{i,1}, \dots, 1-t_{i,d}]$. In this way, we re-proved the classical Koksma–Hlawka inequality, see e.g., [19] and [46], which says that the worst case error of $Q_{n,d}$ is equal to the L_2-discrepancy of points $1-t_i$ and coefficients a_i. Clearly, we also have

$$\begin{aligned} e\left(n, W_2^{1,\dots,1}([0,1]^d), \mathrm{QMC}\right) &= \overline{\mathrm{disc}}_2(n,d), \\ e\left(n, W_2^{1,\dots,1}([0,1]^d), \mathrm{LIN}\right) &= \mathrm{disc}_2(n,d). \end{aligned} \tag{3.23}$$

From this and from the absolute bounds on the L_2-discrepancy we conclude that integration for the Sobolev space $F_d = W_2^{1,\dots,1}([0,1]^d)$ is *strongly tractable* with the exponent at most 1.4779 for the absolute error. On the other hand, from the normalised bounds on the L_2-discrepancy we conclude that integration for $F_d = W_2^{1,\dots,1}([0,1]^d)$ is *intractable* for the normalised error, i.e., for reduction of the initial error. This shows the drastic difference between tractability with respect to absolute and normalised errors.

3.5.2 Average Errors

The L_2-discrepancy is also related to the *average* case error of algorithms $Q_{n,d}$ as shown in [98]. Namely, let $F_d = C([0,1]^d)$ be the space of continuous functions f defined over $[0,1]^d$. The space is equipped with the Wiener sheet measure μ which is the zero mean Gaussian measure with the covariance function

$$K_\mu(x,t) := \int_{C([0,1]^d)} f(x) f(t)\, \mu(df) = \prod_{j=1}^{d} \min(x_j, t_j).$$

That is, the covariance function K_μ is the same as the reproducing kernel of the space $W_2^{1,\dots,1}([0,1]^d)$. Then $f(x) = 0$ with probability one if at least one component of x is zero. We have

$$\left(\int_{C([0,1]^d)} \left(I_d(f) - Q_{n,d}(f)\right)^2 \, \mu(df) \right)^{1/2} = \mathrm{disc}_2(\{1-t_i\}, \{a_i\}).$$

Hence, the bounds on L_2-discrepancy can be also used to study integration and its tractability in the average case setting.

The relation between L_2-discrepancy and the average case setting is an example of a general relation between the worst case setting for the unit ball of a Hilbert space $H(K)$ with a reproducing kernel K and the average case setting of a Banach space equipped with a Gaussian measure with covariance function K. This subject is beyond the scope of this paper and the interested reader is referred to [68, 84].

3.5.3 Spaces without Boundary Values and Weighted Norms

So far, we have assumed that functions satisfy boundary conditions $f(x) = 0$ for x with at least one component equal to zero. We now discuss a more general weighted case without such boundary conditions. As before, we start with univariate functions, but we now drop the assumption that $f(0) = 0$. By $W^1_{2,\gamma}([0,1])$ we denote the space of absolute continuous univariate functions $f : [0,1] \to \mathbb{R}$ for which $f' \in L_2([0,1])$. The inner product is given by

$$\langle f, g\rangle \;=\; f(0)\, g(0) \;+\; \gamma^{-1} \int_0^1 f'(x)\, g'(x)\, dx.$$

Here γ is a positive parameter. The case $\gamma = 1$ corresponds to the classical Sobolev space. The parameter γ is not important for the univariate case. For the multivariate case we will take the parameters γ_j corresponding to successive components and their role will be crucial, as was for the weighted L_2-discrepancy.

The space $W^1_{2,\gamma}([0,1])$ is a reproducing kernel Hilbert space with the kernel

$$K_{1,\gamma}(x,t) \;=\; 1 + \gamma \,\min(x,t).$$

For the multivariate case, we take

$$\gamma_1 \;\ge\; \cdots \;\ge\; \gamma_d \;>\; 0,$$

and consider $F_{d,\gamma}$ as the tensor product of $W^1_{2,\gamma_i}([0,1])$,

$$F_{d,\gamma} \;=\; W^{1,\dots,1}_{2,\gamma}([0,1]^d) \;:=\; W_{2,\gamma_1}([0,1]) \otimes \;\cdots\; \otimes\; W_{2,\gamma_d}([0,1]).$$

The space $W^{1,\dots,1}_{2,\gamma}([0,1]^d)$ is a reproducing kernel Hilbert space with the kernel

$$K_{d,\gamma}(x,t) \;=\; \prod_{k=1}^{d} \left(1 + \gamma_k \,\min(x_k, t_k)\right).$$

The inner product in the space $W_{2,\gamma}^{1,\dots,1}([0,1]^d)$ is given by

$$\langle f,g\rangle \;=\; \sum_{u\in U_d} \gamma_u^{-1} \int_{[0,1]^{|u|}} \frac{\partial^{|u|} f}{\partial x_u}\,(x_u,0)\,\frac{\partial^{|u|} g}{\partial x_u}\,(x_u,0)\,dx_u.$$

For $\gamma_j = 1$ we obtain the classical (tensor product) Sobolev spaces whereas for decreasing γ_j's the weighted spaces consist of functions with diminishing dependence on the jth variable.

It now comes as no surprise that for $Q_{n,d}(f) = \sum_{i=1}^n a_i f(t_i)$ we have

$$e(Q_{n,d}) \;=\; \mathrm{disc}_{2,\gamma}(\{1-t_i\},\{a_i\}).$$

This immediately leads to

$$e\,(n,F_{d,\gamma},\mathrm{QMC}) \;=\; \overline{\mathrm{disc}}_{2,\gamma}(n,d), \quad e\,(n,F_{d,\gamma},\mathrm{LIN}) \;=\; \mathrm{disc}_{2,\gamma}(n,d). \tag{3.24}$$

From the bounds on the weighted L_2-discrepancy presented in the previous subsection, we conclude that

- integration for $W_{2,\gamma}^{1,\dots,1}([0,1]^d)$ is strongly tractable (for the normalised error) iff

$$\sum_{j=1}^{\infty} \gamma_j \;<\; \infty.$$

  If $\sum_{j=1}^\infty \gamma_j < \infty$ then the strong exponent $p^* \in [1,2]$ and $\sum_{j=1}^\infty \gamma_j^{1/2} < \infty$ implies that $p^* = 1$.
- integration for $W_{2,\gamma}^{1,\dots,1}([0,1]^d)$ is tractable (for the normalised error) iff

$$a \;:=\; \limsup_{d\to\infty} \frac{\sum_{j=1}^d \gamma_j}{\log\, d} \;<\; \infty.$$

  If $a < \infty$ then the d-exponent of tractability is at most $a/6$ and the ε-exponent is at most 2.

We stress that as long as $\sum_{j=1}^\infty \gamma_j < \infty$ then absolute and normalised errors are essentially the same, and we have strong tractability in both cases. Also if $\sum_{j=1}^d \gamma_j$ grows proportionally to $\log\, d$, we obtain tractability for both cases. The only difference may be in the d-exponent of tractability.

3.6 L_p-Star Discrepancy

The L_p-star discrepancy is intimately related to the worst case error of multivariate integration for the Sobolev class of functions that are once differentiable in each variable with finite L_q-norm, $1/p + 1/q = 1$, see [19] and [46].

Recall that the L_p-star discrepancy of points $t_1, \dots, t_n \in [0,1]^d$ is defined by

$$\operatorname{disc}_p^*(t_1, \dots, t_n) = \left(\int_{[0,1]^d} \left| x_1 \cdots x_d - \frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(t_i) \right|^p dx) \right)^{1/p}, \quad (3.25)$$

for $1 \le p < \infty$, and

$$\operatorname{disc}_\infty^*(t_1, \dots, t_n) = \sup_{x \in [0,1]^d} \left| x_1 \cdots x_d - \frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(t_i) \right| \quad (3.26)$$

for $p = \infty$. It is customary to denote the L_∞-star discrepancy as the $*$-discrepancy. For $n = 0$, the L_p-star discrepancy $\operatorname{disc}_p^*(0,d)$ is the initial error of multivariate integration without sampling the function. We have $\operatorname{disc}_p^*(0,d) = (p+1)^{-d/p}$, which goes to zero with d exponentially fast for all finite p. This may indicate that the multivariate integration problem is not properly scaled for finite p. For $p = \infty$, this problem disappears since we have $\operatorname{disc}_\infty^*(0,d) = 1$.

The usual bounds on the L_p-star discrepancy are for a fixed dimension d and large n. It is well known that the asymptotic behaviour of $\operatorname{disc}_p^*(n,d)$ with respect to n is of order at most $n^{-1}(\log n)^{d-1}$, see once more [19] and [46]. As before, points for which the L_p-star discrepancy has a bound proportional to $n^{-1}(\log n)^d$ are called low discrepancy points. There is a deep and still evolving theory how to construct such low discrepancy points. This theory is mostly due to Niederreiter and his collaborators.

In this subsection we discuss the L_p-star discrepancy for uniformly distributed points. We consider only even p and define the average L_p-star discrepancy as

$$\operatorname{av}_p(n,d) = \left(\int_{[0,1]^{nd}} \operatorname{disc}_p^*(t_1, \dots, t_n)^p \, dt \right)^{1/p}, \qquad t = (t_1, \dots, t_n).$$

As shown in [29], the average L_p-star discrepancy depends on the Stirling numbers $s(k,i)$ of the first kind, and $S(k,i)$ of the second kind, see [67]. We have

$$\mathrm{av}_p(n,d)^p = \sum_{r=p/2}^{p-1} C(r,p,d)\cdot n^{-r}, \tag{3.27}$$

where

$$C(r,p,d) = (-1)^r \sum_{i=0}^{p-r} \binom{p}{r+i} (-1)^i \sum_{k=i}^{i+r} (p+1-r+k-i)^{-d} s(k,i) S(i+r,k). \tag{3.28}$$

Furthermore,

$$|C(r,p,d)| \le \frac{(r+1)(4p)^p}{(p+1-r)^d},$$

and

$$\mathrm{av}_p(n,d) \le 4\sqrt{2}\, p\,(1+p/2)^{-d/p}\, n^{-1/2} \left(\sum_{i=0}^{p/2-1} n^{-i} \left(\frac{1+p/2}{1+p/2\,-\,i} \right)^d \right)^{1/p}.$$

3.7 ∗-Discrepancy

For $p = 2$, there are a number of negative results for the (normalised) discrepancy, see Section 3.3. The question of dependence on d for $p = \infty$ was raised by [36] who asked whether there exists an $a > 1$ such that $\mathrm{disc}^*_\infty(\lceil a^d \rceil, d)$ tends to 1 as d goes to infinity, and also asked whether, in particular, $\mathrm{disc}^*_\infty(2^d, d)$ goes to 1 as d goes to infinity. Based on the results for $p = 2$ one may be inclined to believe that the answer to at least one of these questions is affirmative.

It was surprising for us that this is *not* the case and that a positive result holds. In [29], it is proven that $\mathrm{disc}^*_\infty(n,d)$ depends only *polynomially* on d and n^{-1}, establishing the upper bound

$$\mathrm{disc}^*_\infty(n,d) \;\le\; C\, d^{1/2}\, n^{-1/2} \qquad \forall\, n,d = 1,2,\dots, \tag{3.29}$$

with an unknown multiplicative factor C.

Hence, the ∗-discrepancy depends at most on $d^{1/2}$. We shall see later that this dependence on d cannot be improved. We do not know whether the dependence on n in (3.29) is sharp. As already mentioned, the asymptotic behaviour of the ∗-discrepancy on n (for a fixed d) is much better, but it does not necessarily mean that the uniform bound which is valid for all d and n cannot be of order $n^{-1/2}$. This problem is open and seems very difficult.

The proof of (3.29) follows directly from deep results from the theory of empirical processes. In particular, we use a result of [79] combined with a result of [27], and the Vapnik–Červonenkis dimension of the family of rational cubes $[0, x)$, see e.g., [18]. The proof is non-constructive, and we do not know for which points the bound (3.29) holds.

The slightly worse upper bound

$$\mathrm{disc}_\infty^*(n,d) \;\le\; 2\sqrt{2}\,n^{-1/2}\,\left(d\log\left(\left\lceil \frac{dn^{1/2}}{2(\log 2)^{1/2}}\right\rceil + 1\right) + \log 2\right)^{1/2} \tag{3.30}$$

follows from Hoeffding's inequality and is quite elementary. Also this proof is non-constructive, but we know that many points almost satisfy the bounds (3.29) and (3.30), see [29] for details. It would be very interesting to *construct* such points.

One can also use the results on the average behaviour of L_p-star discrepancy for an even integer p to obtain upper bounds for the $*$-discrepancy, see [29]. For concrete values of d and ε, these upper bounds seem to be better than those of (3.30).

For the lower bounds, we extend the formula (3.26) by allowing arbitrary a_i's instead of n^{-1}, defining

$$d_\infty^*(t_1,\dots t_n; a_1,\dots,a_n) = \sup_{x\in[0,1]^d}\left| x_1\cdots x_d \;-\; \sum_{i=1}^n a_i 1_{[0,x)}(t_i)\right|$$

and

$$n_\infty(d,\varepsilon) = \min\left\{ n \;:\; \inf_{t_i,a_i} d_\infty^*(t_1,\dots t_n; a_1,\dots,a_n) \;\le\; \varepsilon\right\}.$$

Of course,

$$n_\infty(d,\varepsilon) \le n_\infty^*(d,\varepsilon).$$

In [29], we prove that there exists a positive number c such that for all d and all $\varepsilon \in (0, 1/64]$,

$$n_\infty(d,\varepsilon) \;\ge\; c\,d\log\varepsilon^{-1}. \tag{3.31}$$

From (3.29) and (3.31) we conclude that the $*$-discrepancy depends linearly on the dimension. We also prove that for any $\lambda \in (0, 1)$ there exists a positive c_λ such that

$$n_\infty(d,\varepsilon) \;\ge\; c_\lambda\, d^\lambda\,\varepsilon^{-(1-\lambda)}. \tag{3.32}$$

3.8 Extreme or Unanchored Discrepancy

We already discussed several different discrepancies but always considered boxes of the form $[0,x)$ as our test-sets, see, for example, the classical L_2-discrepancy (3.2). A more symmetric notion of discrepancy can be defined using all boxes $[x,y)$ for $x,y \in [0,1]^d$ with $x \le y$ (in each coordinate). Indeed, instead of (3.1) we consider weighted sums

$$\mathrm{disc}(x,y) = (y_1 - x_1)\cdots(y_d - x_d) - \sum_{i=1}^{n} a_i 1_{[x,y)}(t_i). \tag{3.33}$$

The *extreme or unanchored L_2-discrepancy* of points $t_1,\dots,t_n$ and coefficients $a_1,\dots,a_n$ is just the L_2-norm of the error function (3.33),

$$\mathrm{disc}_2^{\mathrm{ex}}(\{t_i\},\{a_i\}) = \left(\int_{[0,1]^{2d},\ x\le y} \mathrm{disc}^2(x,y)\,dx\,dy\right)^{1/2}. \tag{3.34}$$

By direct integration, we have the explicit formula for the extreme L_2-discrepancy,

$$\begin{aligned}\mathrm{disc}_2^{\mathrm{ex}}(\{t_i\},\{a_i\})^2 &= \frac{1}{12^d} - \frac{1}{2^{d-1}}\sum_{i=1}^{n} a_i \prod_{k=1}^{d} t_{i,k}(1-t_{i,k}) \qquad (3.35)\\ &+ \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \prod_{k=1}^{d} (\min(t_{i,k},t_{j,k}) - t_{i,k}t_{j,k}).\end{aligned}$$

Morokoff and Caflisch introduce the discrepancy (3.34) in [41] and notice that the "L_2-discrepancy over all rectangles had not been previously defined nor used". These authors prefer the new discrepancy because it is "symmetric" and does not prefer a particular vertex, like the classical star discrepancy. So far it was not known that this kind of discrepancy is also directly related to numerical integration, i.e., if it is an error bound for a suitable class of functions. We will see that the extreme discrepancy is the error for periodic functions with a boundary condition.

We begin, however, with periodic functions without a boundary condition and summarise the analysis of [30] and [31] for this case. For $d=1$ and $p=q=2$, we consider the Sobolev space of periodic functions $F_{1,2} = \{f \in W_2^1([0,1]) \mid f(0) = f(1)\}$ with the norm

$$\|f\|^2 = f(1)^2 + \|f'\|_{L_2}^2.$$

This is a rank-1 modification of the space $W_2^1([0,1])$ and we obtain the kernel

$$K_1(x,y) = 1 + (\min(x,y) - xy)\,.$$

The analysis can be extended, by Hölder's inequality, to arbitrary q. The norm in $F_{1,q}$ is given by $\|f\|^q = |f(1)|^q + \|f'\|_{L_q}^q$. For arbitrary $d > 1$, we define $F_{d,q}$ by tensor products of factors $F_{1,q}$ with tensor product norms.

We now discuss the respective error (or discrepancy) of a QMC method $Q_{n,d}$ for the space $F_{d,2}$. The error of $Q_{n,d}$ is given by

$$e(Q_{n,d}) = \left(\sum_{u \subset S} \int_{[0,1)^{2|u|},\, x_u \le y_u} \mathrm{disc}^2((x_u,0),(y_u,1))\, dx_u dy_u \right)^{1/2}.$$

For arbitrary $Q_{n,d} \in \mathrm{LIN}$, we obtain

$$e(Q_{n,d})^2 = \frac{13^d}{12^d} - \sum_{i=1}^n 2a_i \prod_{k=1}^d \left(1 + \frac{t_{i,k}(1-t_{i,k})}{2}\right) + \sum_{i=1}^n \sum_{j=1}^n a_i a_j \prod_{k=1}^d (1 + \min(t_{i,k}, t_{j,k}) - t_{i,k} t_{j,k}).$$

Again we can modify this to spaces with a boundary condition. We start with $d = 1$ and $p = q = 2$, and we take the space $F_{1,2} = \{f \in W_2^1([0,1]) \mid f(0) = f(1) = 0\}$. The kernel for this subspace is given by

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

For $d > 1$, we use tensor product kernels and norms. For $p = 2$, the error $e(Q_{n,d})$ of any $Q_{n,d} \in \mathrm{LIN}$ is

$$e(Q_{n,d}) = \left(\int_{[0,1)^{2d},\, x \le y} \mathrm{disc}^2(x,y)\, dx\, dy \right)^{1/2} = \mathrm{disc}_2^{\mathrm{ex}}(\{t_i\},\{a_i\}). \tag{3.36}$$

Hence, we see that the L_2-extreme discrepancy is also an error bound, this time for a class of periodic functions with a boundary condition. To prove the error bound (3.36) for the kernel

$$K_d(x,y) = \prod_{j=1}^d (\min(x_j,y_j) - x_j y_j)\,,$$

we can simply use the general result (2.12), together with formula (3.35).

We end this subsection with a note on the classical *extreme discrepancy* for the space $F_{d,1}$, which corresponds to $p=\infty$. For $d=1$, the norm of $F_{1,1}$ is given by $\|f\| = 2\|f'\|_1$. We have

$$\mathrm{disc}_{\infty}^{\mathrm{ex}}(\{t_i\},\{a_i\}) = \sup_{x\le y} |\mathrm{disc}(x,y)|. \tag{3.37}$$

The extreme discrepancy (3.37) is tractable, see [29]. It is enough to use equal weights $1/n$ for which we have an upper bound of the form

$$\inf_{t_1,\dots,t_n} \mathrm{disc}_{\infty}^{\mathrm{ex}}(\{t_i\},\{1/n\}) \le C\cdot d^{1/2}\cdot n^{-1/2}, \tag{3.38}$$

where the positive C does not depend on n or d. This bound is the same as for the star discrepancy, we only may have a larger constant C.

3.9 Centred Discrepancy

We now consider the space $F_1 = W^1_{1/2}([0,1])$ which is the Sobolev space of absolutely continuous functions whose first derivatives are in $L_2([0,1])$ and whose function values are zero at the point $1/2$. That is,

$$F_1 \;=\; \{f:[0,1]\to\mathbb{R}\;:f(1/2)=0,\; f \text{ abs. cont. and } f'\in L_2([0,1])\}$$

with the inner product $\langle f,g\rangle_{F_1} = \int_0^1 f'(t)g'(t)\,dt$. It can be checked that this Hilbert space has the reproducing kernel

$$K_1(x,t) = 1_M(x,t)\cdot\min(|x-1/2|,|t-1/2|), \tag{3.39}$$

where $M=[0,1/2]\times[0,1/2]\cup[1/2,1]\times[1/2,1]$. We may also write

$$K_1(x,t) = \tfrac12(|x-1/2|+|t-1/2|-|x-t|). \tag{3.40}$$

The kernel K_1 is *decomposable* in the sense of [62], i.e., $K(x,t)=0$ for all $x\le 1/2\le t$.

For $d>1$, we obtain $F_d = W^{1,1,\dots,1}_{1/2}([0,1]^d) = W^1_{1/2}([0,1])\otimes\cdots\otimes W^1_{1/2}([0,1])$, d times, as the Sobolev space of smooth functions f defined over $D_d=[0,1]^d$ such that $f(x)=0$ if at least one component of x is $1/2$. The inner product of F_d is given by

$$\langle f,g\rangle_{F_d} \;=\; \int_{[0,1]^d} \frac{\partial^d}{\partial x_1\cdots\partial x_d} f(x)\,\frac{\partial^d}{\partial x_1\cdots\partial x_d} g(x)\,dx.$$

It can be checked, see [62], that the error of an algorithm $Q_{n,d}(f) = \sum_{i=1}^n a_i f(z_i)$ is now given by

$$e^2(Q_{n,d}) \;= \tag{3.41}$$

$$\int_{[0,1]^d}\left|\prod_{j=1}^d \min(x_j,1-x_j) - \sum_{i=1}^n a_i\cdot 1_{J(b(x),x)}(z_i)\right|^2 dx,$$

where $J(b(x),x)$ is the cube generated by x and the vertex $b(x)$ of $[0,1]^d$ that is closest, in the sup-norm, to x. Essentially the same formulas are presented by [30], who considers spaces similar to F_d without assuming the condition $f(\frac{1}{2}) = 0$ and considers algorithms with $a_i = 1/n$. The error of the algorithm $Q_{n,d}$ with $a_i = 1/n$ is called by Hickernell the *centred discrepancy* and denoted by $d_2^c(Q_{n,d})$. We will be using the same name for arbitrary coefficients a_i. For the space F_d with the condition $f(\frac{1}{2}) = 0$ we denote the centred discrepancy by $\tilde{d}_2^c(n,d)$. Later we remove this condition and consider the centred discrepancy $d_2^c(Q_{n,d})$ (without the tilde), as originally studied by [30].

The minimal centred discrepancy is defined as

$$\tilde{d}_2^c(n,d)^2 = \inf_{a_i,z_i,i=1,2,\dots,n} \int_{[0,1]^d} \left| \prod_{j=1}^d \min(x_j, 1-x_j) - \sum_{i=1}^n a_i \cdot 1_{J(b(x),x)}(z_i) \right|^2 dx. \tag{3.42}$$

In [62] we prove the lower bound

$$\tilde{d}_2^c(n,d) \ge \left(1 - n\,2^{-d}\right)_+^{1/2} \tilde{d}_2^c(0,d) \quad \text{with} \quad \tilde{d}_2^c(0,d) = 12^{-d/2}. \tag{3.43}$$

The centred discrepancy (and uniform integration) may be also considered for the space $L_q([0,1]^d)$ with $q \in [1,\infty]$. More precisely, for $d=1$ we take the space $F_{1,q}$ as the Sobolev space of absolutely continuous functions whose first derivatives are in $L_q([0,1])$ and that vanish at $\frac{1}{2}$. The norm in $F_{1,q}$ is given by $\|f\|_{F_{1,q}} = \left(\int_0^1 |f'(t)|^q\, dt\right)^{1/q}$ for $q < \infty$, and $\|f\|_{F_{1,\infty}} = \operatorname{ess\,sup}_{t\in[0,1]} |f'(t)|$ for $q = \infty$. For $d > 1$, the space $F_{d,q}$ is taken as a tensor product of factors $F_{1,q}$. Then functions from $F_{d,q}$ vanish at x whenever at least one component of x is $\frac{1}{2}$ and their norm is

$$\|f\|_{F_{d,q}} = \|D^{\vec{1}} f\|_{L_q([0,1]^d} = \left(\int_{[0,1]^d} |D^{\vec{1}} f(x)|^q\, dx \right)^{1/q},$$

where $\vec{1} = [1,1,\dots,1]$ and $D^{\vec{1}} = \partial^d / \partial x_1 \cdots \partial x_d$. We have

$$I_d(f) - Q_{n,d}(f) = \int_{[0,1]^d} D^{\vec{1}} f(t)\, D^{\vec{1}} \left(h_d - \sum_{i=1}^n a_i K_d(\cdot, z_i) \right)(t)\, dt,$$

with

$$h_d(x) = 2^{-d} \prod_{j=1}^d \left(|x_j - \tfrac{1}{2}| - |x_j - \tfrac{1}{2}|^2 \right),$$

$$K_d(x,t) \quad = \quad 2^{-d}\prod_{j=1}^{d}\left(|x_j-\tfrac12|+|t_j-\tfrac12|-|x_j-t_j|\right).$$

From this we conclude that

$$e(Q_{n,d}) \;:=\; \sup_{f\in F_{d,q}\colon \|f\|_{F_{d,q}}\le 1}\left|I_d(f)-Q_{n,d}(f)\right| \;=\; \tilde d_p^c(Q_{n,d}),$$

where $1/p+1/q=1$ and $\tilde d_p^c(Q_{n,d})$ is the centred p-discrepancy given by

$$\tilde d_p^c(Q_{n,d}) = \left(\int_{[0,1]^d}\left|\prod_{j=1}^{d}\min(x_j,1-x_j)-\sum_{i=1}^{n}a_i\cdot 1_{J(b(x),x)}(z_i)\right|^p dx\right)^{1/p}.$$

If $q=1$ then $p=\infty$ and, as usual, the integral is replaced by the essential supremum in the formula above.

Let $e(n,F_{d,q})=\tilde d_p(n,d)$ denote the minimal error, or equivalently the minimal centred p-discrepancy, that can be achieved by using n function values. The initial error, or the initial centred discrepancy, is now given by

$$e(0,F_{d,q},\mathrm{LIN}) \;=\; \tilde d_p^c(0,d) \;=\; \frac{2^{-d}}{(p+1)^{d/p}}$$

for $q>1$, and $e(0,F_{d,1})=2^{-d}$. Hence, for all values of q, the initial centred discrepancy is at most 2^{-d}.

The following result is from [62]. For $n<2^d$ and $p<\infty$, we have

$$\tilde d_p^c(n,d) \;\ge\; \left(1-n\,2^{-d}\right)^{1/p}\,\tilde d_p^c(0,d).$$

Hence, uniform integration is intractable in $F_{d,q}$ since

$$n(\varepsilon,F_{d,q},\mathrm{LIN}) \;\ge\; (1-\varepsilon^p)2^d \quad\text{and}\quad \lim_{d\to\infty}\frac{\tilde d_p^c(d^k,d)}{\tilde d_p^c(0,d)} \;=\; 1 \quad \forall\, k=1,2,\dots .$$

Observe that for $q=1$ we have $p=\infty$, and since $\tilde d_p^c(n,d)$ is a nondecreasing function of p we have

$$\tilde d_\infty^c(n,d) \;=\tilde d_\infty^c(0,d) \;=\; 2^{-d} \quad \forall\, n<2^d.$$

Hence, unlike the $*$-discrepancy and the extreme discrepancy, we find that uniform integration and the centred discrepancy are intractable for $q=1$ and $p=\infty$.

It is known that $\tilde d_\infty^c(n,d)$ goes to zero at least like $n^{-1}(\log n)^{d-1}$. In view of the previous property we must wait, however, exponentially long in d to see this rate of convergence.

3.9.1 Spaces without Boundary Values

We now remove the condition $f(\frac{1}{2}) = 0$. As before, we study the more general case of the weighted L_q norm. That is, $D_1 = [0,1]$ and $F_{1,q,\gamma}$ is the Sobolev space $W_q^1([0,1])$ with the norm

$$\|f\|_{F_{1,q,\gamma}} \;=\; \left(\left|f(\tfrac{1}{2})\right|^q + \gamma^{-q/2}\int_0^1 |f'(x)|^q\,dx\right)^{1/q},$$

where $\gamma > 0$. Observe that for $q = \infty$ we have

$$\|f\|_{F_{1,\infty,\gamma}} \;=\; \max\left(|f(\tfrac{1}{2})|, \gamma^{1/2}\mathrm{ess}\sup_{t\in[0,1]}|f'(t)|\right).$$

For $q = 2$, we have the Hilbert space with the kernel

$$K_{1,\gamma}(x,t) \;=\; 1 + \gamma\,1_M(x,t)\min(|x-\tfrac{1}{2}|, |t-\tfrac{1}{2}|).$$

For $d > 1$, we take $F_{d,q,\gamma} = W_q^{(1,1,\dots,1)}([0,1]^d)$ as the tensor product of $W_q^1([0,1])$. The norm in $F_{d,q,\gamma}$ is given by

$$\|f\|_{F_{d,q,\gamma}} \;=\; \left(\sum_{u\subset\{1,2,\dots,d\}}\gamma_u^{-q/2}\int_{[0,1]^{|u|}}\left|\frac{\partial^{|u|}}{\partial x_u}f(x_u,1/2)\right|^q dx_u\right)^{1/q}. \tag{3.44}$$

The formula for the error of the algorithm $Q_{n,d}(f) = \sum_{i=1}^n a_i f(z_i)$ takes the form, see [30],

$$I_d(f) - Q_{n,d}(f) = \\ \sum_{u\subset\{1,2,\dots,d\}}\int_{[0,1]^{|u|}}\frac{\partial^{|u|}}{\partial x_u}f(x_u,\tfrac{1}{2})\frac{\partial^{|u|}}{\partial x_u}\left(h_d - \sum_{i=1}^n a_i K_d(\cdot,z_i)\right)(x_u,\tfrac{1}{2})\,dx_u,$$

where h_d and the kernel K_d are given as before. Applying the Hölder inequality for integrals and sums to $I_d(f) - Q_{n,d}(f)$ we conclude that

$$e(Q_{n,d}) \;:=\; \sup_{f\in F_{d,q,\gamma}:\,\|f\|_{F_{d,q,\gamma}}\le 1}\left|I_d(f) - Q_{n,d}(f)\right| \;=\; d_{p,\gamma}^c(Q_{n,d}),$$

where $1/p + 1/q = 1$ and the weighted centred p-discrepancy $d_{p,\gamma}^c(Q_{n,d})$ is given by

$$d_{p,\gamma}^c(Q_{n,d}) = \left(\sum_{u\subset\{1,2,\dots,d\}}\gamma_u^{p/2}\int_{[0,1]^{|u|}}|\mathrm{disc}(n,d)^c(x_u,1/2)|^p\,dx_u\right)^{1/p}, \tag{3.45}$$

with

$$\operatorname{disc}(n,d)^c(x_u,1/2) = \prod_{\ell\in u} \min(x_\ell, 1-x_\ell) - \sum_{i=1}^{n} a_i \cdot 1_{J(a(x_u),x_u)}(t_i)_u.$$

Let $e(n, F_{d,q,\gamma}, \mathrm{LIN}) = d_{p,\gamma}(n,d)$ denote the minimal error, or equivalently the minimal weighted centred p-discrepancy, that can be achieved by using n function values. The initial error, or the initial centred weighted p-discrepancy, is now given by

$$e(0, F_{d,q}, \mathrm{LIN}) \;=\; d^c_{p,\gamma}(0,d) \;=$$
$$\left(\sum\nolimits_{u\subset\{1,2,\dots,d\}} \gamma_u^{p/2} \left(\tfrac{2^{-p}}{p+1}\right)^{|u|}\right)^{1/p} \;=\; \prod\nolimits_{j=1}^{d} \left(1+\tfrac{2^{-p}}{p+1}\,\gamma_j^{p/2}\right)^{1/p}.$$

For $q=1$, we have $p=\infty$ and

$$e(0, F_{d,1}, \mathrm{LIN}) \;=\; d^c_{\infty,\gamma}(0,d) \;=\; \max_{k=0,1,\dots,d} \left(2^{-k}(\gamma_1\gamma_2\cdots\gamma_k)^{1/2}\right).$$

It is proved in [62] that

$$d^c_{p,\gamma}(n,d) \ge \left(\sum_{k=0}^{d} C_{d,p,k} \left(\frac{2^{-p}}{p+1}\right)^{k} \cdot (1-n\,2^{-k})_+\right)^{1/p},$$

where

$$C_{d,p,k} \;=\; \sum_{u\subset\{1,2,\dots,d\}\,:\,|u|=k} \gamma_u^{p/2}.$$

For $n \le 2^m < 2^d$, this can be rewritten for $p<\infty$ as

$$d^c_{p,\gamma}(n,d) \ge 2^{-1/p} \left(1 - \frac{\sum_{k=0}^{m} C_{d,p,k}(\frac{2^{-p}}{p+1})^k}{\sum_{k=0}^{d} C_{d,p,k}(\frac{2^{-p}}{p+1})^k}\right)^{1/p} d^c_{p,\gamma}(0,d),$$

and for $p=\infty$ as

$$d^c_{\infty,\gamma} \ge 2^{-(m+1)}\,(\gamma_1\gamma_2\cdots\gamma_{m+1})^{1/2}.$$

For $q>1$, i.e., $p<\infty$, we can check the lack of strong tractability and tractability of I_d in $F_{d,q,\gamma}$, see [62], and obtain the following. If $\sum_{j=1}^{\infty} \gamma_j^{p/2} \;=\; \infty$ then

$$\lim_{d\to\infty} \frac{d^c_{p,\gamma}(n,d)}{d^c_{p,\gamma}(0,d)} \;=\; 1 \qquad \forall\, n,$$

and uniform integration in $F_{d,q,\gamma}$ is not strongly tractable. If

$$\lim^*_{d\to\infty} \frac{\sum_{j=1}^d \gamma_j^{p/2}}{\ln\, d} \;=\; \infty$$

then

$$\lim^*_{d\to\infty} \frac{d^c_{p,\gamma}(d^k,d)}{d^c_{p,\gamma}(0,d)} \;=\; 1 \quad \forall\, k=1,2,\dots\,,$$

and uniform integration in $F_{d,q,\gamma}$ is intractable, where $\lim^*$ is lim or lim sup.

4 Further Intractability Results

4.1 Hilbert Spaces with Positive Kernels

Sometimes it is relatively easy to prove lower bounds for the class POS of positive quadrature formulas whereas for the class LIN of all quadrature formulas, it is either harder to prove similar lower bounds or it is unknown where such bounds hold, see also Sections 4.2 and 5. Positive quadrature formulas were studied in [51, 52, 73, 102]. Here we present a general lower bound of [102] concerning positive quadrature formulas for Hilbert spaces with (pointwise) positive kernels.

Hence we assume that F_d is a reproducing kernel Hilbert space with the reproducing kernel $K_d : D_d \times D_d \to \mathbb{R}$ such that

$$K_d(x,t) \ge 0 \qquad \forall\, x,t \in D_d. \tag{4.1}$$

We consider the linear functional I_d given by

$$I_d(f) \;=\; \int_{D_d} f(t)\,\varrho_d(t)\,dt = \langle f, h_d\rangle\,,$$

where $h_d \in F_d$ and

$$h_d(x) \;=\; \int_{D_d} K_d(x,t)\varrho_d(t)\,dt \qquad \forall\, x \in D_d,$$

see Section 2.3. The lower bound will be expressed in terms of K_d and h_d by the quantity

$$\kappa_d := \sup_{x\in D_d} \frac{|h_d(x)|}{(K_d(x,x)\int_{D_d} h_d(y)\varrho_d(y)\,dy)^{1/2}}. \tag{4.2}$$

One can conclude that $\kappa_d \le 1$; it may happen that $\kappa_d = 1$. If $\kappa_d = 1$, however, then multivariate integration is trivial since $I_d(f)$ can be approximated with an arbitrary small error using only one function value, see [59]. The lower bound

$$e^2(Q_{n,d}) \geq e^2(Q_{0,d})\,(1 - n\,\kappa_d^2) \tag{4.3}$$

is proved in [102] for any $Q_{n,d} \in \mathrm{POS}$ and therefore

$$n(\varepsilon, F_d, \mathrm{POS}) \geq (1 - \varepsilon^2)\,\kappa_d^{-2}. \tag{4.4}$$

For a discussion of this result, see [102]. Here we mention only one example.

4.1.1 Example: Tensor Product Spaces

Assume that $F_d = F_1 \otimes F_1 \otimes \ldots \otimes F_1$ is the tensor product of d copies of the same space F_1. Then $D_d = D_1^d$ and $K_d(x,t) = \prod_{j=1}^d K_1(x_j, t_j)$. We assume that $K_1 \geq 0$ and that the weight ϱ_d is also of product form, $\varrho_d(t) = \prod_{j=1}^d \varrho_1(t_j)$. Then it follows that $\kappa_d = \kappa_1^d$. If univariate integration is not trivial, i.e., $\kappa_1 < 1$, then

$$n(\varepsilon, F_d, \mathrm{POS}) \geq (1 - \varepsilon^2)\,\kappa_1^{-2d}, \tag{4.5}$$

i.e., the minimal number of samples of positive quadrature formulas goes exponentially fast to infinity with d.

We apply this result to the kernel

$$K_1(x,t) = 1 + \min(x,t)$$

on $[0,1]^2$, which corresponds to the Sobolev space $W_2^1([0,1])$. In this case we have

$$h_1(x) \;=\; \int_0^1 K_1(x,t)\,dt = 1 + x - x^2/2$$

and

$$\kappa_1 := \sup_{x\in[0,1]} \frac{h_1(x)}{(K_1(x,x)\int_0^1 h_1(y)\,dy)^{1/2}}.$$

A trivial computation yields

$$\kappa_1 = \frac{1}{3}\,\frac{1+2\sqrt{7}}{\sqrt{2+\sqrt{7}}} \approx 0.97298 \qquad \text{and} \qquad \kappa_1^{-2} \approx 1.0563.$$

This proves formula (3.20) in Section 3.4.

The bound (4.5) can also be applied to "very small" classes of analytic functions. For example we can consider $I_d(f) = \int_{[0,1]^d} f(x)\,dx$ and the norm

$$\|f\|^2 = \sum_{\alpha \in \mathbb{N}_0^d} \|D^\alpha f\|_{L_2}^2,$$

see [52]. For this last example we do not know whether arbitrary linear algorithms are more efficient than positive algorithms.

4.2 Classes of Smoother Functions

In Section 3, we presented lower bounds for different discrepancies and for related integration problems. The integrands of these classes were not very smooth, since the highest existing derivative was $f^{(1,1,\dots,1)}$. Positive quadrature formulas for smooth functions were previously discussed in Section 4.1. In this section, we present lower bounds for smoother functions for the class LIN of all quadrature formulas.

In [62] we studied multivariate (uniform and Gaussian) integration defined for integrand spaces F_d, such as weighted Sobolev spaces of functions of d variables with smooth mixed derivatives. The weight γ_j moderates the behaviour of functions with respect to the jth variable. For $\gamma_j = 1$, we obtain the classical Sobolev spaces whereas for decreasing γ_j's the weighted Sobolev spaces consist of functions with diminishing dependence on the jth variable.

The results of [62] are obtained for arbitrary linear tensor product functionals I_d defined over weighted tensor product reproducing kernel Hilbert spaces F_d of functions of d variables.

The kernel of F_d is denoted by K_d, and is fully determined by the kernel K_1 of the space F_1 of univariate functions. Our point of departure is the notion of a *decomposable* kernel. We say that the kernel K_1 is decomposable if $K_1(x,t) = 0$ for all $x \le a \le t$ for some a. For example, kernels of the Sobolev spaces of smooth functions vanishing at a are decomposable. For such kernels, the problem of approximating the linear functional I_d naturally decomposes into 2^d subproblems. These subproblems are not trivial iff the two subproblems for $d = 1$ are not trivial. This is usually the case when the point a belongs to the interior of the function domain. If so, we find that $n(\varepsilon, F_d, \mathrm{LIN})$ depends exponentially on d, and so the problem is intractable.

We then extend our analysis to certain non-decomposable kernels. We assume first that the kernel $K_1 = R_1 + \gamma R_2$ where R_i are kernels of Hilbert spaces $H(R_i)$ such that $H(R_1) \cap H(R_2) = \{0\}$, and R_2 is decomposable. For the d dimensional case, we take the weighted tensor product Hilbert space whose kernel is

$$K_d(x,t) = \prod_{j=1}^{d} \left(R_1(x_j,t_j) + \gamma_j R_2(x_j,t_j) \right),$$

where x_j and t_j denote the successive components of x and t. We stress that we may use different weights γ_j for different components. Assuming, as before, that I_d has non-trivial subproblems with respect to $H(R_2)$, we prove that $\sum_{j=1}^{\infty} \gamma_j = \infty$ implies that I_d is not strongly tractable, and that $\limsup_d \sum_{j=1}^{d} \gamma_j / \ln\, d = \infty$ implies that I_d is intractable.

Sobolev spaces without the condition that functions vanish at a have reproducing kernels of this form, provided that a does not belong to the boundary of the domain. This holds for the centred discrepancy, which is also analysed in the Banach case where the norm is taken in the L_p sense with $p \in [1,\infty]$, see Section 3.9.

Finally, we consider the case when $K_1 = R_1 + \gamma R_2$, where R_2 is not necessarily decomposable but is the sum of two kernels with one decomposable term. This last assumption holds for the Sobolev spaces with or without boundary conditions when the kernel R_2 is modified by an operator of finite rank. In particular, this decomposition holds for the L_2-star discrepancy. In this case, we obtain the same sufficient conditions for the lack of strong tractability and intractability of I_d.

The approach in [62] requires us to identify a part of the reproducing kernel that is decomposable. As demonstrated by many examples in this paper, this can be done for many classical spaces. Decomposability means, roughly speaking, that F_1 contains two orthogonal subspaces $F_{(0)}$ and $F_{(1)}$ such that

$$f(x) = 0 \quad \text{for} \quad x \notin D_{(0)} \ \forall\, f \in F_{(0)},$$

$$f(x) = 0 \quad \text{for} \quad x \notin D_{(1)} \ \forall\, f \in F_{(1)},$$

where the set $D_{(0)} \cap D_{(1)}$ has at most one element and both the $D_{(i)}$ have positive Lebesgue measure. If the spaces $F_{(i)}$ are non-trivial (i.e., they contain elements different from zero) then our results are non-trivial and lead to intractability.

This theory cannot be applied, however, to spaces of analytic functions, since these functions cannot be decomposed as above. For example, the kernel

$$K_1(x,t) = \sum_{j=1}^{\infty} \frac{x^j t^j}{(j!)^2},$$

corresponds to a Hilbert space of entire functions. In general, for spaces of analytic functions, tractability issues of linear functionals are still open.

4.2.1 Sobolev Spaces

In this subsection, we illustrate the results of [62] for smooth Sobolev spaces with and without boundary conditions. Let r be an arbitrary (maybe large) natural number. We consider the Sobolev space $F_1 = W_0^r([0,1])$ with the boundary condition $f(0) = \ldots = f^{(r-1)}(0) = 0$ and the inner product

$$\langle f, g \rangle_{F_1} = \int_0^1 f^{(r)}(t) g^{(r)}(t)\, dt.$$

The reproducing kernel is

$$K_1(x,t) = \int_0^1 \frac{(x-u)_+^{r-1}}{(r-1)!} \frac{(t-u)_+^{r-1}}{(r-1)!}\, du.$$

An explicit form of K_1 is

$$K_1(x,t) = \frac{x^r}{r!} \sum_{j=0}^{r-1} \frac{r}{r+j} \frac{x^j}{j!} \frac{(t-x)^{r-1-j}}{(r-1-j)!},$$

for $x \le t$. This means that for a fixed x, the function $g(t) = K_1(x,t)$ is a polynomial in $t \in [x,1]$ of degree at most $r-1$.

For an arbitrary $a^* \in (0,1)$, consider the subspace F_{a^*} of F_1 consisting of functions f for which $f^{(i)}(a^*) = 0$ for $i = 0, 1, \ldots, r-1$. Observe that $f^{(i)}(a^*) = \left\langle f, K_1^{(i,0)}(a^*, \cdot) \right\rangle_{F_1}$, where $K_1^{(i,0)}$ denotes i times differentiation with respect to the first variable. Hence, $f \in F_{a^*}$ iff f is orthogonal to

$$A_{r-1} = \mathrm{span}(K^{(0,0)}(a^*, \cdot), K^{(1,0)}(a^*, \cdot), \ldots, K^{(r-1,0)}(a^*, \cdot)).$$

Let $\{g_0, \ldots, g_{r-1}\}$ be an orthonormal basis of A_{r-1}. Since each function $K_1^{(i,0)}(a^*, \cdot)$ is a polynomial of degree at most $r-1$, the same holds for g_i in the interval $[a^*, 1]$.

Let K_{a^*} be the reproducing kernel of F_{a^*}. It is known that

$$K_{a^*}(x,t) \;=\; K_1(x,t) - \sum_{j=0}^{r-1} g_j(x)g_j(t).$$

One can check that K_{a^*} is decomposable with a^*. That is, for $x \le a^* \le t$ we have $K_{a^*}(x,t) = 0$. Hence, we have $K_1 = R_{1,a^*} + R_{2,a^*}$, where $R_{1,a^*}(x,t) = \sum_{j=1}^{r-1} g_j(x)g_j(t)$ and $R_{2,a^*} = K_{a^*}$ is decomposable.

Consider now uniform integration, $I_1(f) = \int_0^1 f(t)\,dt$. Then

$$h_1(x) \;=\; \int_0^x \frac{(1-u)^r}{r!}\,\frac{(x-u)^{r-1}}{(r-1)!}\,du$$

is a polynomial of degree $2r$. Since $h_{1,2}$ differs from h_1 by a polynomial of degree $r-1$, the function $h_{1,2}$ is a polynomial of degree $2r$. Its restrictions $h_{1,2,(0)}$ and $h_{1,2,(1)}$ are also polynomials of degree $2r$ over $[0,a^*]$ and $[a^*,1]$ and therefore their norms must be non-zero. It follows from the general results of [62] that uniform integration is intractable.

We now remove the boundary conditions and consider the Sobolev space $F_{1,\gamma} = W^r([0,1]^d)$ with the inner product

$$\langle f,g\rangle_{F_{1,\gamma}} \;=\; \sum_{j=0}^{r-1} f^{(j)}(0)\,g^{(j)}(0) \;+\; \gamma^{-1}\int_0^1 f^{(r)}(t)\,g^{(r)}(t)\,dt.$$

The kernel is now

$$K_{1,\gamma}(x,t) \;=\; R_1(x,t) + \gamma\int_0^1 \frac{(x-u)_+^{r-1}}{(r-1)!}\,\frac{(t-u)_+^{r-1}}{(r-1)!}\,du$$

with

$$R_1(x,t) \;=\; \sum_{j=0}^{r-1} \frac{x^j}{j!}\,\frac{t^j}{j!}\,.$$

The kernel $K_{1,\gamma}$ can be written as

$$K_{1,\gamma}(x,t) \;=\; R_1(x,t) \;+\gamma\sum_{j=1}^{r-1} g_j(x)g_j(t) \;+\; \gamma\,K_{a^*}(x,t).$$

For uniform integration we have

$$h_1(x) \;=\; \sum_{j=0}^{r-1}\frac{x^j}{j!(j+1)!} \;+\; \gamma\int_0^x \frac{(1-u)^r}{r!}\,\frac{(x-u)^{r-1}}{(r-1)!}\,du.$$

So h_1 is still a polynomial of degree $2r$. We summarise the results for this case.

Uniform integration for the Sobolev space $F_d = W_0^{r,r,\dots,r}([0,1]^d)$ is intractable for any $r \ge 1$. In particular,

$$\lim_{d\to\infty} \frac{e(d^q, F_d)}{e(0,F_d)} = 1 \quad \forall\, q = 1,2,\dots, .$$

If $\sum_{j=1}^{\infty} \gamma_j = \infty$ then uniform integration in the Sobolev space $F_{d,\gamma} = W^{r,r,\dots,r}([0,1]^d)$ is not strongly tractable and

$$\lim_{d\to\infty} \frac{e(n, F_{d,\gamma})}{e(0,F_{d,\gamma})} = 1 \quad \forall\, n.$$

If $\lim^*_{d\to\infty} \sum_{j=1}^{d} \gamma_j / \ln d = \infty$ then uniform integration in the Sobolev space $F_{d,\gamma} = W^{r,r,\dots,r}([0,1]^d)$ is intractable and

$$\lim{}^*_{d\to\infty} \frac{e(d^q, F_{d,\gamma})}{e(0,F_{d,\gamma})} = 1 \quad \forall\, q = 1,2,\dots,$$

where $\lim^* \in \{\lim, \limsup\}$.

If $\gamma_j \ge \gamma_1 > 0$ for all j then uniform integration in the Sobolev space $F_{d,\gamma} = W^{r,r,\dots,r}([0,1]^d)$ is intractable and

$$\lim_{d\to\infty} \frac{e(\lfloor b^d \rfloor, F_{d,\gamma})}{e(0,F_{d,\gamma})} = 1$$

for $b \in (1, b^*)$ with $b^* > 1$. The proof of this result gives a bound for b^*. This allows to prove the lower bound in (3.13) and in (3.21).

5 On the Power of Negative Weights

In Section 4.1 we studied linear functionals defined on Hilbert spaces and obtained the general lower bounds (4.3) and (4.4) for quadrature formulas modulo two assumptions:

- We assumed that the kernel of the Hilbert space is (pointwise) positive.
- We assumed that the quadrature formula only uses positive weights.

We may ask whether the second assumption is really necessary. One might be inclined to believe that for positive functionals, such as the unweighted Lebesgue integral, positive quadrature formulas should be optimal. This guess is supported by the results reported in Section 3.4: the (normalised) weighted L_2-discrepancy is strongly tractable (or tractable) for the class POS iff it is strongly tractable (or tractable) for the class LIN.

On the other hand, there is the very powerful algorithm of Smolyak for general multivariate tensor product problems which uses negative weights, see Section 6. In some cases the Smolyak algorithm is better than all *explicitly known* quadrature formulas with positive weights. Of course this does not prove that general linear algorithms are better than positive ones. Proving the optimality of positive algorithms for natural function spaces is a challenging problem.

One way to define such natural function spaces is as follows. We assume that F_d is a tensor product Hilbert space consisting of functions $f : [0,1]^d \to \mathbb{R}$. We consider quadrature formulas for the usual (unweighted) integral and assume that the kernel of F_d is positive. Moreover, we assume that F_d is a subspace of the Sobolev space $W^{1,1,\dots,1}([0,1]^d)$. We will see that these assumptions are not enough to prove the optimality of positive algorithms. In fact, it may happen that general linear algorithms are exponentially better than positive algorithms. Our example is rather artificial, however, and we still do not know whether optimal algorithms for classes such as the $F_{d,\gamma}$ from Section 3.5.3 use negative weights. We begin with a result from [59]. We consider a Hilbert space F_d of functions defined on $[0,1]^d$ that is a tensor product, $F_d = F_1 \otimes \ldots \otimes F_1$. The space F_1 is two-dimensional and generated by the orthonormal functions e_1 and e_2. We may take $e_1(x) = |x - 1/2|$ and $e_2(x) = (x - 1/2) \cdot g(x)$, where $|g(x)| \le 1$ is symmetric about $1/2$, $g(1/2 - x) = g(1/2 + x)$, and takes infinitely many values. We can take g of the form $g(x) = 1/2 + \alpha\,(x - 1/2)^2$. Note that

$$K_1(x,t) \;=\; e_1(x)e_1(t) + e_2(x)e_2(t)$$

is pointwise non-negative.

Then the (partially defined) function e_1/e_2 has a range that is infinite but not dense in $\mathbb{R}$, since $|e_1(x)/e_2(x)| \ge 1$. Let I_1 be a linear functional satisfying $e(1, F_1, \mathrm{LIN}) > 0$. Note that $I_1(f) = \int_0^1 f(t)\,dt$ is such a functional. For any such I_1, let $I_d = I_1 \otimes \ldots \otimes I_1$ be the d-fold tensor product functional. Then

$$e(d, F_d, \mathrm{LIN}) > 0 \quad \text{ and } \quad e(d+1, F_d, \mathrm{LIN}) = 0.$$

Hence,

$$n(\varepsilon, F_d, \mathrm{LIN}) \;\le\; d+1 \qquad \forall\, \varepsilon \in [0,1).$$

This means that all tensor product functionals on F_d are tractable. If we only allow positive quadrature formulas then we obtain a completely different result. Since all the assumptions of Section 4.1.1 hold, we obtain

$$n(\varepsilon, F_d, \mathrm{POS}) \ge (1 - \varepsilon^2)\,\kappa_1^{-2d},$$

where $\kappa_1 < 1$, i.e., the minimal number of samples of positive quadrature formulas goes exponentially fast to infinity with d.

This example proves that, at least for some spaces, allowing negative weights breaks intractability of integration using positive quadrature formulas. It would be of practical importance to characterise for which spaces F_d this phenomenon occurs.

6 Smolyak-Type Algorithms

The known constructions of polynomial time algorithms for multivariate tensor product problems are based on a construction of Smolyak (1963) and on certain modifications of this algorithm, see also [4, 11, 16, 17, 20, 21, 24, 25, 26, 53, 54, 55, 56, 57, 58, 64, 65, 66, 76, 77, 78, 80, 90, 93, 95]. Even this long list is not complete since similar algorithms (under names such as *blending algorithm* or *sparse grid algorithm* or *Boolean algorithm* or *hyperbolic cross points algorithm*) were studied by many authors in different contexts.

Explicit cost bounds and error estimates (not including any unknown constants) for Smolyak formulas are proved in [90]. These estimates are quite general. For a given problem, one only has to know suitable algorithms and error bounds for the univariate case $d = 1$. Then the Smolyak algorithm and the bounds from [90] automatically give the algorithm and the error bounds for the tensor product problem in any dimension $d > 1$. The Smolyak algorithm has been modified in different papers, see the papers which we already mentioned.

For simplicity, we describe the Smolyak algorithm only for multivariate integration

$$I_d(f) = \int_{[0,1]^d} f(x)\, dx.$$

We assume that for $d = 1$ quadrature formulas†

$$U^i(f) = \sum_{j=1}^{m_i} a_j^i\, f(t_j^i)$$

are given. Here we have $m_{i+1} > m_i$. For $d > 1$ a *(tensor) product*

† Hence the a_j^i are real numbers. In case of interpolation or operator equations the a_i^j are functions.

formula is given by

$$(U^{i_1} \otimes \cdots \otimes U^{i_d})(f) = \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_d=1}^{m_{i_d}} a_{j_1}^{i_1} \cdots a_{j_d}^{i_d} \; f(t_{j_1}^{i_1}, \ldots, t_{j_d}^{i_d}).$$

To evaluate this formula one needs

$$n = \prod_{\ell=1}^{d} m_{i_\ell}$$

function values. The algorithm of Smolyak is a suitable linear combination of such tensor products with the aim of only using a small number of function values. It is given by

$$A(q,d) = \sum_{q-d+1 \le |i| \le q} (-1)^{q-|i|} \cdot \binom{d-1}{q-|i|} \cdot (U^{i_1} \otimes \cdots \otimes U^{i_d}),$$

where $q \ge d$. Here $i = [i_1, i_2, \ldots, i_d]$ with integers $i_j \ge 1$, and $|i| = i_1 + i_2 + \cdots + i_d$.

The information used by the algorithm $A(q,d)$ consists of the function values $f(t_{j_1}^{i_1}, \ldots, t_{j_d}^{i_d})$ with $j_k \le m_{i_k}$. Assume that $m_i \le M^i$ for some $M > 1$. Then the indices of the sample points $t_{j_1}^{i_1}, \ldots, t_{j_d}^{i_d}$ satisfy

$$j_1 j_2 \cdots j_d \;\le\; m_{i_1} m_{i_2} \cdots m_{i_d} \;\le\; M^{|i|} \;\le\; M^q.$$

Hence, the indices $j = [j_1, j_2, \ldots, j_d]$ satisfy the so called *hyperbolic cross* inequality, and therefore the information used by $A(q,d)$ is called *hyperbolic cross* information.

We now discuss the (worst case) error of the Smolyak algorithm $A(q,d)$ for the unit ball of a tensor product Hilbert space $F_d = F_1 \otimes F_1 \otimes \cdots \otimes F_1$ (d times). The error of $A(q,d)$ depends on the error of the one-dimensional algorithms U^i for the unit ball of the Hilbert space F_1. If we assume that *nested* information is used, i.e., the sample points of U^i are also used by U^{i+1}, and the weights a_j^i are chosen optimally, then the algorithm $A(q,d)$ remains optimal for all d. This holds independently of the choice of nested sample points t_j^i. Obviously, the smaller the one-dimensional error of U^i is, the smaller the error of $A(q,d)$ will be. The error of $A(q,d)$ can be expressed in terms of the errors of U^i, and the explicit formulas can be found in [90].

We now discuss the cost $n(\varepsilon, F_d)$ of the Smolyak algorithm $A(q,d)$ which is defined as the minimal number of function values for which the worst case error $\|I_d - A(q,d)\|$ is at most ε. Assume that the error of

U^i is polynomial in m_i. That is, there exist positive C and p such that

$$\|I_1 - U^i\| \le C\, m_i^{-1/p} \qquad \forall\, i.$$

Then, see [90],

$$n(\varepsilon, F_d) \le \beta_1 \left(\beta_2 + \beta_3 \frac{\log \varepsilon^{-1}}{d-1}\right)^{\beta_4(d-1)} \left(\frac{1}{\varepsilon}\right)^p \tag{6.1}$$

for some positive β_i which are fully determined by the error of the one-dimensional algorithms U^i.

We stress that the leading factor of (6.1) is ε^{-p} and the exponent of ε^{-1} does *not* depend on d and is the same as for $d = 1$. The dependence on d is through $\log \varepsilon^{-1}$. Note that for a fixed d, we have

$$C_d := \limsup_{\varepsilon \to 0} \frac{n(\varepsilon, F_d)}{(\log \varepsilon^{-1})^{\beta_4(d-1)}\, \varepsilon^{-p}} \le \beta_1 \left(\frac{\beta_3}{d-1}\right)^{\beta_4(d-1)}.$$

Hence, the asymptotic constant C_d goes to zero super-exponentially fast with d.

For a fixed ε and varying d, or for both ε and d varying, the bound (6.1) may depend exponentially on d. It turns out that if $\|I_1\| < 1$, then $n(\varepsilon, F_d)$ has a bound

$$n(\varepsilon, F_d) \le C\, \varepsilon^{-p^*} \qquad \forall \varepsilon > 0$$

for some positive C and p^*. (In general, $p^* > p$.) This bound is independent of d and depends polynomially on ε^{-1}, see [90]†,

Hence, we have *strong tractability* (for the absolute error) and the Smolyak algorithm is a *constructive* algorithm that computes an ε-approximation using a polynomial number of function values. We stress that the Smolyak algorithm is fully constructive. We only need to know the one-dimensional algorithms U^i with good error properties, which usually is easy to obtain for many spaces F_1. The total computing time of the Smolyak algorithm is proportional to the number of sample points; this holds for any dimension d. The additional cost (for example, the cost of computing the binomials) is negligible.‡

† The assumption $\|I_1\| < 1$ is also a necessary condition for strong tractability provided that I_1, instead of being an integration operator, is an arbitrary continuous linear operator of rank at least two.

‡ There exist other constructive algorithms whose errors serve as upper bounds on the complexity. For example, this holds for algorithms based on constructive versions of Tchakalov's theorem. This theorem says the following. If $V(d, k)$ denotes the space of polynomials of d variables and degree at most k then there exists a quadrature formula with positive weights and a number of sample points not larger than the dimension of $V(d, k)$ which is exact for all polynomials from

We illustrate the Smolyak algorithm by an example. Many more examples can be found in the papers cited previously. We take

$$m_1 = 1 \quad \text{and} \quad m_i = 2^{i-1} + 1 \quad \text{for } i > 1 \tag{6.2}$$

and choose Clenshaw–Curtis formulas

$$U^i(f) = \sum_{j=1}^{m_i} a_j^i \, f(t_j^i)$$

with (6.2) and the sample points

$$t_j^i = -\cos \frac{\pi(j-1)}{m_i - 1}$$

($t_1^1 = 0$). We then obtain an algorithm which was presented in [54] and studied further in several other papers. Here one uses, for $d = 1$, *non-equidistant* sample points that are *nested*. Both properties improve the error bounds, as well as the stability of the algorithm. The $A(q, d)$ are almost optimal for many different function classes, see [54] and [55]. For all the spaces from Section 3 we obtain

$$e(A(q,d)) \preceq n^{-1} \cdot (\log n)^{2(d-1)},$$

where n is the number of sample points of $A(q, d)$. The weights of $A(q, d)$ are relatively small since $\|A(q,d)\|_\infty \preceq (\log n)^{d-1}$. The formula $A(k + d, d)$ is exact for all polynomials of degree $2k + 1$ and uses, for large d, about $n \approx 2^k \, d^k / k!$ sample points, see [56].

The Smolyak algorithm (together with the midpoint rule for $d = 1$) can be used to prove *constructive* upper bounds for the star-discrepancy. From the results of [90] we obtain

$$n(\varepsilon, F_d, \mathrm{LIN}) \le C_\varepsilon \cdot 62^d. \tag{6.3}$$

This upper bound is very bad if compared with the best non-constructive upper bound $n(\varepsilon, F_d, \mathrm{LIN}) \le C_\varepsilon \cdot d$. To prove (6.3) we use the formulas of [90], p. 32. For the nested midpoint rule U_i in one dimension we use $n_i = (3^i - 1)/2$ sample points to obtain the error 3^{-i}. Using the notation of [90], we have $\alpha_1 = \alpha_2 \cdot \log(3e/\log 3)$, with $\alpha_2 = e^2/((e-1)\log 3)$. Hence we obtain $\beta_2 = \alpha_1 < 7.85$ and (6.3) follows from formula (50) of [90] since $C = 1$ and $\beta_4 = 2$.

A major modification of the original Smolyak algorithm for the case

$V(d, k)$. The computation of this quadrature formula is, however, so expensive that it cannot be implemented for, say, $d > 10$.

of weighted tensor product problems together with explicit error bounds was given in [93]. Using the *WTP-algorithm* from this paper one can constructively prove strong tractability of weighted problems provided the weights go to zero sufficiently fast. Such a result was already mentioned in Section 3.4.

7 Path Integration

Many applications require approximate values of path integrals. Often path integrals are approximated by a high dimensional integrals and then a *Monte Carlo* (or randomised) algorithm is applied. Typical Monte Carlo algorithms need roughly ε^{-2} integrand evaluations and the error bound ε is guaranteed only in a stochastic sense. Do we really need to use randomised algorithms for path integrals? In this section we present results concerning *deterministic algorithms* together with *worst case error (and complexity) bounds.*

7.1 Tractability of Path Integration

We consider path integration with respect to a Gaussian measure μ for different classes F of integrands. That is, we want to compute an ε-approximation to

$$I(f) = \int_X f(x)\, d\,\mu(x), \qquad f \in F,$$

where X is a Banach (usually infinite dimensional) space equipped with a Gaussian measure μ.

Tractability of path integration means that the complexity (the minimal number of integrand evaluations) depends polynomially on ε^{-1}, where ε is the worst case error. The results depend on the Gaussian measure, mainly through the eigenvalues λ_i of the covariance operator, and on the class F of integrands. We present two results from [91].

For the first result we assume that $r \in \mathbb{N}$, and F^r is the class of all $f : X \to \mathbb{R}$ such that $f^{(r)}$ is continuous with $\|f^{(k)}(x)\| \le 1$ for all $x \in X$ and $k = 0, 1, \ldots, r$. Then the tractability of path integration depends on the eigenvalues λ_i, which we assume to be ordered,

$$\lambda_1 \ge \lambda_2 \ge \ldots \ge 0.$$

There are two cases.

First case: If $r \ge 1$ and only k eigenvalues are positive, i.e., $\lambda_k > 0$ and

$\lambda_{k+1} = 0$, then the path integration problem is tractable with exponent k/r, i.e.,

$$n(\varepsilon, F^r) \asymp \varepsilon^{-k/r}.$$

Second case: If $r = 0$ or if infinitely many eigenvalues are positive, then the path integration problem is intractable.

Observe that the first case is not very typical. In this case the measure is concentrated on a k-dimensional space, which contradicts the essence of the path integration problem since we really have a finite dimensional integration problem. Hence the second case is typical and we have a negative result on tractability of path integration. This result indicates that the class F^r is simply too large.

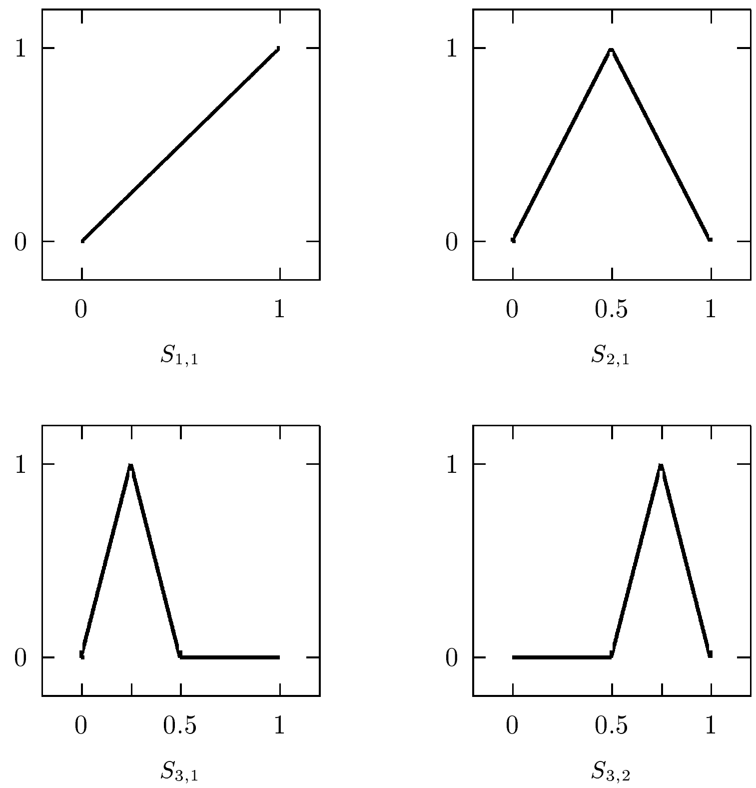
To obtain a positive result on tractability of path integration, one may study classes of entire integrands. In [91] the authors assume additionally that one can compute the derivatives of integrands at zero. For a class of entire integrands the optimal algorithm is given in [91], together with error bounds that prove the tractability of the problem. It turns out that the complexity of path integration is at most of the order ε^{-p}, where p depends on the Gaussian measure and is always less than or equal to 2. For the classical Wiener measure, $\lambda_k \asymp k^{-2}$, the worst case upper bound is of the order $\varepsilon^{-2/3}$.

7.2 A Smolyak Type algorithm for Wiener Integrals

In this section we present a Smolyak type algorithm for Wiener integrals, i.e., integrals with respect to the Wiener measure W. Such integrals are usually computed with the Monte Carlo algorithm or with number theoretic (or quasi-Monte Carlo) algorithms, based on low discrepancy sequences. These algorithms usually are reliable but slow. In particular, these algorithms can hardly exploit the smoothness properties of the integrand. Here we discuss the computation of Wiener integrals by a Smolyak type algorithm following [58] and [78]. The new algorithm is superior if the integrand is sufficiently smooth.

We begin with the *Lévy Ciesielski decomposition* of the Wiener measure by means of the *Schauder or hat functions.*

The Schauder functions $S_{j,k}$ are defined for $k = 1, \dots, m(j)$, where $m(j) = 2^{j-2}$ for $j \geq 2$. Consider independent standard normal random variables $Y_{j,k}$ and variances $\sigma^2(1) = 1$, $\sigma^2(j) = 2^{-j}$ for $j \geq 2$. Then for



$d = 2^{s-1}$ let

$$W^{(d)} = \sum_{j=1}^{s} \sigma(j) \sum_{k=1}^{m(j)} Y_{j,k} \cdot S_{j,k}$$

be the *s-th Lévy Ciesielski approximation* of the Wiener measure. For $s \to \infty$ we get weak convergence to the Wiener measure. We approximate Wiener integrals

$$E(f) = \int_{C[0,1]} f(x)\, dW(x)$$

by the finite dimensional expectations $E(f(W^{(d)}))$. Here the influence of the variables $Y_{j,k}$ decreases with increasing j.

For approximation of $E(f(W^{(d)}))$ we use

$$A(s)(f) = A(s-1+d,\, d)(f \circ W^{(d)}), \quad d = 2^{s-1} \tag{7.1}$$

where $A(q,d)$ is a modification of the Smolyak algorithm of the form

$$A(q,d) = \sum_{q-d+1 \le |i| \le q} (-1)^{q-|i|} \cdot \binom{d-1}{q-|i|} \cdot (U^{i_1-\beta_1} \otimes \cdots \otimes U^{i_d-\beta_d})$$

with Gauss Hermite formulas U^k that use $2k-1$ sample points for $k \in \mathbb{N}$ and 1 sample point for $k \le 0$. The β_j are defined by $\beta_j = n$ for $j \in \{2^n+1, \ldots, 2^{n+1}\}$ if $n \in \mathbb{N}_0$ and $\beta_1 = 0$.

Worst case error bounds for the algorithm (7.1) and a class F of analytic functions are proved in [78]. The class F contains all Lipschitz continuous functionals with the property that all even partial derivatives in the direction of Schauder functions are bounded by 1. The error bounds are of the form $e(A(s)) \le 2^{-s/2}$ and it is proved that $A(s)$ uses at most 2^{2s-1} integrand evaluations (sample paths).

7.3 Feynman–Kac Path Integration

Path integration is intractable for the class of C^r-integrands, see 7.1. Often, however, the integrals have a particular form. The *Feynman–Kac integrals* are of the form

$$S(v,V) = \int_{C[0,1]} v(x(1)) \exp\left(\int_0^1 V(x(s))\, ds\right) dw(x), \tag{7.2}$$

where v and V belong to a Banach space F. Recall that Feynman–Kac integrals occur as the solution of heat equations by the use of the Feynman–Kac formula.

We report some results from [66], where a slightly more general problem is studied. For example, the function exp in (7.2) can be replaced by any smooth function. Observe that the solution operator S depends linearly on the function v, but nonlinearly on the function V. In [66] it is assumed that the functions v and V belong, in particular, to a subclass F of $C^r(\mathbb{R})$ functions. The authors present an algorithm A_ε that computes, for inputs v and V from F, the number $S(v,V)$ up to the error bound ε. The algorithm A_ε is again based on the Smolyak algorithm and uses the values of v and V at deterministically chosen sample points derived from from a certain approximation problem for the class F. An explicit bound on the cost of the algorithm A_ε is given. By the cost, we mean the total number of function values of v and V plus all arithmetic operations needed to compute $A_\varepsilon(v,V)$.

For the given subclass F of C^r-functions the cost is roughly of the order $\varepsilon^{-1/r}$ and this is almost optimal. This should be compared with the classical Monte Carlo algorithm whose cost is at least of the order ε^{-2}. Hence there is a great improvement, even for a modest smoothness $r = 2$ or $r = 3$.

To apply the new algorithm A_ε, one must precompute certain constants that are needed by the algorithm. Precomputation is typical for many algorithms. Consider, for example, the classical Gaussian quadrature formula

$$Q_n(f) = \sum_{i=1}^{n} a_i\, f(x_i) \quad \text{for} \quad \int_{-1}^{1} f(x)\, dx. \tag{7.3}$$

To compute the sample points x_i and the weights a_i one has to solve certain equations. This is usually done just once (for example with Maple or Mathematica and high precision) and the results of this precomputation are stored in a file that defines the algorithm. The cost of precomputation is usually ignored since it is done only once for all functions f for which we use (7.3). The cost of the algorithm (7.3) is then of order n.

The problem with the new algorithm A_ε is that it requires calculation of many multivariate integrals. Hence it is by itself a difficult problem. So far the precomputation has been done for $r = 1$, 2, and 4 and the precomputed values have an error of the order 10^{-4}. This is enough for the very fast computation of $S(v, V)$ if the needed accuracy is not too large. Nevertheless, for higher accuracy or for higher r, the problem of precomputation currently limits the use of this algorithm.

Acknowledgments

We thank Klaus Ritter, Joseph Traub and Arthur Werschulz for valuable comments. We also thank Arieh Iserles for editing our paper for the FOCM proceedings. The first author was supported by a Heisenberg scholarship of the German Research Council (DFG) and by a scholarship of the KU Leuven. The second author was partly supported by the National Science Foundation.

References

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

[2] Bakhvalov, N.S. (1959). On approximate computation of integrals, *Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem.* **4**, 3–18 (Russian).

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

[4] Barthelmann, V., Novak, E. and Ritter, K. (1998). High dimensional polynomial interpolation on sparse grids. *Adv. in Comput. Math.,* to appear.

[5] Beck, J. and Chen, W. (1987). *Irregularities of Distribution* (Cambridge University Press, Cambridge).

[6] Beck, J. and Sós, V.T. (1995). Discrepancy theory, in *Handbook of Combinatorics* (R. Graham, M. Grötschel and L. Lovász, eds) (Elsevier, Amsterdam).

[7] Bellman, R. (1957). *Dynamic Programming* (University Press, Princeton NJ).

[8] Blum, L., Shub, M. and Smale, S. (1989). On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines, *Bull. AMS* **21**, 1–46.

[9] Blum, L., Cucker, F., Shub, M. and Smale, S. (1998). *Complexity and Real Computation* (Springer-Verlag, New York).

[10] Boender, C.G.E. and Romeijn, H.E. (1995). Stochastic methods. in *Handbook of Global Optimization* (R. Horst, P. M. Pardalos, eds) (Kluwer, Dordrecht), 829–869.

[11] Bungartz, H.-J. and Griebel, M. (1999). A note on the complexity of solving Poisson's equation for spaces of bounded mixed derivatives, *J. Complexity* **15**, 167–199.

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

[13] Chen, W.W.L. (1987). On irregularities of distribution and approximate evaluation of certain functions, II, in *Analytic Number Theory and Diophantine Problems, Proceedings of a Conference at Oklahoma State University in 1984* (Birkhäuser, Basel), 75–86.

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

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

[16] Cools, R., Novak, E. and Ritter, K. (1999). Smolyak's construction of cubature formulas of arbitrary trigonometric degree, *Computing* **62**, 147–162.

[17] Delvos, F.-J. and Schempp, W. (1989). *Boolean Methods in Interpolation and Approximation,* Pitman Research Notes in Mathematics Series **230** (Longman, Essex).

[18] Dudley, R.M. (1984). A course on empirical processes (École d'Été de Probabilités de Saint-Flour XII-1982), LNiM 1097 (Springer-Verlag, New York), 2–141,

[19] Drmota, M. and Tichy, R.F. (1997). *Sequences, Discrepancies and Applications,* LNiM **1651** (Springer-Verlag, Berlin).

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

[21] Frank, K., Heinrich, S. and Pereverzev, S. (1996). Information complexity of multivariate Fredholm integral equations in Sobolev classes. *J. Complexity* **12**, 17–34.

[22] Frolov, K.K. (1980). Upper bounds on the discrepancy in L_p, $2 \leq p < \infty$, *Dokl. Akad. Nauk USSR* **252**, 805–807.
[23] Gal, S. and Micchelli, C.A. (1980). Optimal sequential and non-sequential procedures for evaluating a functional, *Appl. Anal.* **10**, 105–120.
[24] Genz, A.C. (1986). Fully symmetric interpolatory rules for multiple integrals, *SIAM J. Numer. Anal.* **23**, 1273–1283.
[25] Gerstner, Th. and Griebel, M. (1998). Numerical integration using sparse grids, *Num. Algorithms* **18**, 209–232.
[26] Griebel, M., Schneider, M. and Zenger, Ch. (1992). A combination technique for the solution of sparse grid problems, in *Iterative Methods in Linear Algebra* (R. Beauwens R. and P. de Groen, eds) (Elsevier, Amsterdam), 263–281.
[27] Haussler, D. (1995). Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Červonenkis dimension, *J. Combinatorial Theory* **A 69**, 217–232.
[28] Heinrich, S. (1995). Efficient algorithms for computing the L_2-discrepancy, *Math. Comput.* **65**, 1621-1633.
[29] Heinrich, S., Novak, E., Wasilkowski, G.W. and Woźniakowski, H. (1999). The inverse of the star-discrepancy depends linearly on the dimension, *Acta Arithmetica,* to appear.
[30] Hickernell, F.J. (1998). A generalized discrepancy and quadrature error bound, *Math. Comput.* **67**, 299–322.
[31] Hickernell, F.J. (1998). What affects the accuracy of quasi-Monte Carlo quadrature? Preprint.
[32] Hickernell, F.J. and Woźniakowski, H. (1999). Integration and approximation in arbitrary dimension, *Adv. Comput. Math.* **12**, 25–58.
[33] Joe, S. (1997). Formulas for the computation of the weighted L_2-discrepancy. Manuscript.
[34] Kon, M.A. and Novak, E. (1990). The adaption problem for approximating linear operators, *Bull. AMS* **23**, 159–165.
[35] Larcher, G. (1998a). Digital point sets: analysis and application, in *Random and quasi-random point sets* (P. Hellekalek, G. Larcher, eds), LNiS **138** (Springer-Verlag, Berlin) 167–222.
[36] Larcher, G. (1998b). Talk at Dagstuhl-Seminar.
[37] Matoušek, J. (1998). The exponent of discrepancy is at least 1.0669, *J. Complexity* **14**, 448–453.
[38] Matoušek, J. (1999). *Geometric Discrepancy* (Springer-Verlag, Berlin).
[39] Meer, K. and Michaux, C. (1997). A survey on real structural complexity theory, *Bull. Belg. Math. Soc.* **4**, 113–148.
[40] Morokoff, W.J. (1998). Generating quasi-random paths for stochastic processes, *SIAM Review* **40**, 765–788.
[41] Morokoff, W.J. and Caflisch, R.E. (1994). Quasi-random sequences and their discrepancies, *SIAM J. Sci. Comput.* **15**, 1251–1279.
[42] Nemirovsky, A. (1996). Polynomial time methods in convex programming, in*The Mathematics of Numerical Analysis, 1995 AMS–SIAM Summer Seminar in Applied Mathematics* (J. Renegar, M. Shub and S. Male, eds), AMS Lect. Appl. Math. **32**, 543–589.
[43] Nemirovsky, A.S. and Yudin, D.B. (1983). *Problem Complexity and Method Efficiency in Optimization* (Wiley-Interscience, New York).
[44] Nesterov, Yu. and Nemirovskii, A. (1994). *Interior-Point Polynomial*

Algorithms in Convex Programming, SIAM Studies in Applied Mathematics (SIAM, Philadelphia).

[45] Neumayer, A. (1997). Molecular modeling of proteins and mathematical prediction of protein structure, *SIAM Review* **39**, 407–460.

[46] Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods* (SIAM, Philadelphia).

[47] Niederreiter, H. and Xing, C. (1998). Nets, (t, s)-sequences, and algebraic geometry, in *Random and quasi-random point sets* (P. Hellekalek, G. Larcher, eds), LNiS **138** (Springer-Verlag, Berlin), 267–302.

[48] Novak, E. (1988). *Deterministic and Stochastic Error Bounds in Numerical Analysis,* LNiM **1349** (Springer-Verlag, Berlin).

[49] Novak, E. (1995). The real number model in numerical analysis, *J. Complexity* **11**, 57–73.

[50] Novak, E. (1996). On the power of adaption, *J. Complexity* **12**, 199–237.

[51] Novak, E. (1999). Intractability results for positive quadrature formulas and extremal problems for trigonometric polynomials, *J. Complexity* **15**, 299–316.

[52] Novak, E. (1999). Is there a curse of dimension for integration? in *Eleventh Int. Conf. on Domain Decomposition Methods* (C.-H. Lai, P. E. Bjørstad, M. Cross and O. B. Widlund, eds), 89–96.

[53] Novak, E. and Ritter, K. (1996). Global optimization using hyperbolic cross points, in *State of the Art in Global Optimization: Computational Methods and Applications* (C. A. Floudas and P. M. Pardalos, eds) (Kluwer, Dordrecht), 19–33.

[54] Novak, E. and Ritter, K. (1996). High dimensional integration of smooth functions over cubes, *Numer. Math.* **75**, 79–97.

[55] Novak, E. and Ritter, K. (1997). The curse of dimension and a universal method for numerical integration, in *Multivariate Approximation and Splines* (G. Nürnberger, J. W. Schmidt and G. Walz, eds), ISNM **125** (Birkhäuser, Basel), 177–188.

[56] Novak, E. and Ritter, K. (1999). Simple cubature formulas with high polynomial exactness, *Construct. Approx.* **15**, 499–522.

[57] Novak, E., Ritter, K., Schmitt, R. and Steinbauer, A. (1999). On an interpolatory method for high dimensional integration, *J. Comput. Appl. Math.* **112**, 215–228.

[58] Novak, E., Ritter, K. and Steinbauer, A. (1998). A multiscale method for the evaluation of Wiener integrals, in *Approximation Theory IX, Volume 2* (C.K. Chui and L.L. Schumaker, eds), 251–258.

[59] Novak, E., Sloan, I.H. and Woźniakowski, H. (1997). Tractability of tensor product linear operators, *J. Complexity* **13**, 387–418.

[60] Novak, E. and Woźniakowski, H. (1996). Topological complexity of zero finding, *J. Complexity* **12**, 380–400.

[61] Novak, E. and Woźniakowski, H. (1999). On the cost of uniform and nonuniform algorithms, *Theoretical Comp. Sc.* **219**, 301–318.

[62] Novak, E. and Woźniakowski, H. (1999). Intractability results for integration and discrepancy, *J. Complexity,* to appear.

[63] Novak, E. and Woźniakowski, H. (2000). Complexity of linear problems with a fixed output basis, *J. Complexity* **16**, 333–362.

[64] Petras, K. (2000). On the Smolyak cubature error for analytic functions, *Adv. in Comput. Math.* **12**, 71–93.

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

[66] Plaskota, L., Wasilkowski, G.W. and Woźniakowski, H. (1999). A new algorithm and worst case complexity for Feynman-Kac path integration. Preprint.

[67] Riordan, J. (1958). *An Introduction to Combinatorial Analysis,* (Wiley and Sons, New York).

[68] Ritter, K. (1999). *Average Case Analysis of Numerical Problems,* LNiM (Springer-Verlag, Berlin), to appear.

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

[70] Roth, K.F. (1980). On irregularities of distributions IV, *Acta Arithm.* **37**, 67–75.

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

[72] Sloan, I.H. and Woźniakowski, H. (1997). An intractability result for multiple integration, *Math. Comp.* **66**, 1119–1124.

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

[74] Schäffler, S. (1995). Unconstrained global optimization using stochastic integral equations, *Optimization* **35**, 43–60.

[75] Smale, S. (1998). Mathematical problems for the next century, *Math. Intell.* **20**, 7–15.

[76] Smolyak, S.A. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions, *Sov. Math. Doklady* **4**, 240–243.

[77] Sprengel, F. (1997). Interpolation and wavelets on sparse Gauss–Chebyshev grids, in *Multivariate Approximation* (W. Haussmann et al., eds), Math. Res. **101** (Akademie Verlag, Berlin), 269–286.

[78] Steinbauer, A. (1999). Quadrature formulas for the Wiener measure, *J. Complexity* **15**, 476–498.

[79] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes, *Ann. Probability* **22**, 28–76.

[80] Temlyakov, V.N. (1994). *Approximation of Periodic Functions* (Nova Science, New York).

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

[82] Tezuka, S. (1998), Financial applications of Monte Carlo and Quasi-Monte Carlo methods, in *Random and Quasi-random Point Sets* (P. Hellekalek, G. Larcher, eds), LNiM **138** (Springer-Verlag, Berlin).

[83] Thiémard, E. (1999). Computing bounds for the star discrepancy. Preprint.

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

[85] Traub, J.F. and Werschulz, A.G. (1998). *Complexity and Information* (Cambridge University Press, Cambridge).

[86] Traub, J.F. and Woźniakowski, H. (1980). *A General Theory of Optimal Algorithms* (Academic Press, New York).

[87] Traub, J.F. and Woźniakowski, H. (1982). Complexity of linear programming, *Oper. Res. Lett.* **1**, 59–62.

[88] Wahba, G. (1990). *Spline Models for Observational Data,* SIAM-NSF Regional Conference Series in Appl. Math. **59** (SIAM, Philadelphia).

[89] Wang, X. and Hickernell, F.H. (1999). Error bounds and tractability of quasi-Monte Carlo algorithms in infinite dimension. Preprint.

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

[91] Wasilkowski, G.W. and Woźniakowski, H. (1996). On tractability of path integration, *J. Math. Phys.* **37**, 2071–2088.

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

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

[94] Weihrauch, K. (1998). A refined model of computation for continuous problems, *J. Complexity* **14**, 102–121.

[95] Werschulz, A.G. (1996). The complexity of the Poisson problem for spaces of bounded mixed derivatives, in *The Mathematics of Numerical Analysis* (J. Renegar, M. Shub, S. Smale, eds.), Lect. in Appl. Math. **32** (AMS, Providence, RI) 895–914.

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

[97] Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods* (Springer-Verlag, Berlin).

[98] Woźniakowski, H. (1991). Average case complexity of multivariate integration, *Bull. AMS (new series)* **24**, 185–191.

[99] Woźniakowski, H. (1994). Tractability and strong tractability of linear multivariate problems, *J. Complexity* **10** (1994), 96–128.

[100] Woźniakowski, H. (1994). Tractability and strong tractability of multivariate tensor product problems, *J. of Computing and Information* **4**, 1–19.

[101] Woźniakowski, H. (1998). Why does information-based complexity use the real number model? *Theoretical Comp. Sc.* **219**, 451–466.

[102] Woźniakowski, H. (1999). Efficiency of quasi-Monte Carlo algorithms for high dimensional integrals, in *Monte Carlo and Quasi-Monte Carlo Methods 1998* (H. Niederreiter and J. Spanier, eds.) (Springer-Verlag, Berlin), 114–136.