IBC-PROBLEMS RELATED TO STEVE SMALE

ERICH NOVAK AND HENRYK WOŹNIAKOWSKI

ABSTRACT. The first problem deals with topological complexity; that is, with the minimal number of comparisons which have to be performed to solve certain numerical problems. We present recent results on solving scalar nonlinear equations to within ε for different classes F of continuous functions. Depending on F and the set of permissible arithmetic operations, the topological complexity may be zero or roughly $\log\log\varepsilon^{-1}$ or even $\log\varepsilon^{-1}$.

The second problem concerns uniform versus nonuniform algorithms. In particular, we discuss the role of the error threshold ε. If ε is regarded as a fixed parameter we permit built-in constants of algorithms depending on ε, and this is the case of nonuniform algorithms. On the other hand, if ε is treated as a varying parameter and as one of the inputs, the built-in constants of algorithms are independent of ε, and this is the case of uniform algorithms. For some problems, the minimal costs of nonuniform and uniform algorithms may be quite different. We present conditions under which they are roughly the same.

We started these two projects during the Park City conference organized by Steve Smale and the discussions with Steve in Park City and later in Hong Kong, Dagstuhl and by e-mail were very helpful for us.

We end this paper by discussing tractability which is a central theme of discrete and continuous computational complexity. This problem has been also studied by Steve Smale. We want to know for which problems there exists an algorithm which computes an ε-approximation with cost bounded by a polynomial in ε^{-1} and in the input size. In contrast to many areas of complexity, we *prove* intractability or tractability of some problems using information-based arguments. We illustrate this in the worst case setting for numerical integration and discrepancy. We report on recent results that numerical integration is intractable in the L_2 norm and tractable in the L_1 norm. The latter is done jointly with S. Heinrich and G. W. Wasilkowski and follows from the fact that the inverse of the star-discrepancy depends linearly on the dimension.

1. INTRODUCTION

Information-based complexity, IBC, is part of continuous computational complexity. IBC is usually developed over abstract infinite-dimensional linear spaces such as Hilbert or Banach spaces. The applications are typically multivariate problems sometimes in hundreds or thousands of variables. Examples of such problems include multivariate integration or approximation, ordinary or partial differential equations,

Date: February, 2001.

integral equations, optimization, and nonlinear equations. For instance, consider a partial differential equation whose boundary conditions and coefficients are specified by multivariate functions. Since functions cannot be entered into a digital computer, we may only compute a finite number of functionals, such as function values. These functionals are sometimes called oracles or subroutines. Thus only partial information about the problem is available in the computer and the original problem can be only approximately solved. Furthermore, this partial information is often contaminated with noise. Contaminated information is studied by Plaskota (1996); for simplicity we assume here that the partial information is not contaminated.

The goal of IBC is to compute an approximation of the original problem at minimal cost. The error and the cost of approximation can be defined in different settings. In this paper we concentrate on the worst case setting. The ε-complexity is defined as the minimal cost of computing an approximation with error at most ε.

We use the *real number model with an oracle*. Roughly speaking, we add an oracle (or subroutine) for the computation of function values to the BSS-model over the reals, see Blum, Shub, Smale (1989), Blum, Cucker, Shub, Smale (1998), Meer, Michaux (1997), Novak (1995), as well as Traub, Woźniakowski (1980), and Traub, Wasilkowski, Woźniakowski (1988). In this model we assume that we can exactly compute arithmetic operations over reals, comparisons of real numbers and function values by an oracle. 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 Novak, Woźniakowski (1996, 1999a, 2000a), and in Woźniakowski (1998).

Usually we are interested in the total cost of an algorithm which is given by a weighted sum of all operations and often it turns out that sharp complexity bounds are obtained by studying only the information cost. But sometimes it is interesting to count only one kind of admissible operations. Section 2 deals with topological complexity or the question of how many comparisons have to be performed to solve certain numerical problems. We present recent results on solving scalar nonlinear equations to within ε for different classes F of continuous functions. Depending on F and the set of permissible arithmetic operations, the topological complexity may be zero or roughly $\log\log\varepsilon^{-1}$ or even $\log\varepsilon^{-1}$.

In Section 3 we discuss the problem of uniform versus nonuniform algorithms. In particular, we discuss the role of the error threshold ε. If ε is regarded as a fixed parameter we may have built-in constants of algorithms depending on ε, and this is the case of nonuniform algorithms. On the other hand, if ε is treated as a varying parameter and as one of the inputs, the built-in constants of algorithms are independent of ε, and this is the case of uniform algorithms. For some problems, the minimal costs of nonuniform and uniform algorithms may be quite different. We present conditions under which they are roughly the same.

We started our research concerning topological complexity and uniform algorithms during the Park City conference organized by Steve Smale. The discussions with Steve in Park City and later in Hong Kong, Dagstuhl and by e-mail were very helpful for us.

We end this paper by discussing tractability which is a central theme of discrete and continuous computational complexity. This problem has been also studied by Steve Smale. We want to know for which problems there exists an algorithm which computes an ε-approximation with cost bounded by a polynomial in ε^{-1} and in the input size. In contrast to many areas of complexity, we *prove* intractability or tractability of some problems using information-based arguments. We illustrate this in the worst case setting for numerical integration and discrepancy. We report on recent results that numerical integration is intractable in the L_2 norm and tractable in the L_1 norm. The latter is done jointly with S. Heinrich and G. W. Wasilkowski and follows from the fact that the inverse of the star-discrepancy depends linearly on the dimension.

2. Topological Complexity

Our interest in the topological complexity is motivated by the work of Smale (1987) and Vassiliev (1992, 1996). In particular, they consider zero-finding for monic univariate polynomials of degree d with complex coefficients whose absolute values are at most one. They prove that the minimal total number of comparison nodes in the computational graph is roughly d for small ε.

We deal with zero-finding for univariate functions defined on the interval $[0,1]$, and by the topological complexity we mean the minimal depth of the computational graph. We present some results from our paper Novak, Woźniakowski (1996) where the proofs and more details can be found. We analyze different classes F of functions that are subsets of the class

$$F^* = \{f \in C[0,1] \mid f(0) < 0 \text{ and } f(1) > 0\}. \tag{1}$$

We consider the *root error* criterion[1], which was used in Smale and Vassiliev's papers. For this criterion, x is an ε-approximation of a zero of the function f if

$$|x - x^*| \le \varepsilon \quad \text{for some } x^* \text{ such that } f(x^*) = 0.$$

We now comment on arithmetic operations. Usually the standard operations of addition, subtraction, multiplication and division are called arithmetic operations. As in many papers in this area, we understand arithmetic operations in a more general sense. Namely, by an *arithmetic operation* we mean an operation from a given set

[1] In our paper Novak, Woźniakowski (1996) we also consider the *residual* error criterion, where x is an ε-approximation of a zero of the function f if $|f(x)| \le \varepsilon$. The results for these two error criteria can be very different.

ARI. We consider various sets ARI. We always assume that

$$\mathrm{ARI}_{\min} = \{+, -, *\},$$

i.e., addition, subtraction and multiplication, is a subset of the set ARI. As we shall see, division plays a special role and the results depend on whether division belongs to the set ARI.

Since we are primarily interested in the topological complexity, we assume that the cost of each comparison is taken as unity, whereas information and arithmetic operations are free of charge. By

$$\mathrm{comp}^{\mathrm{TOP}}(F, \mathrm{ARI}, \varepsilon)$$

we denote the topological complexity, which is defined as the minimal number of comparisons necessary to compute an ε-approximation for any function from the class F by using finitely many information operations (function values), and finitely many arithmetic operations from the set ARI.

Observe that we define the topological complexity here as the depth of the computation graph which corresponds to the maximal number of comparisons used in one particular computation. The total number of comparison nodes in the computation graph can be much bigger. This latter quantity is also called topological complexity, see Smale (1987).

We will see that the topological complexity depends crucially on the class F of functions and on the choice of arithmetic operations ARI. Sometimes apparently innocent changes of F or ARI lead to completely different topological complexities.

For all F and ARI studied in this paper we have

$$0 \le \mathrm{comp}^{\mathrm{TOP}}(F, \mathrm{ARI}, \varepsilon) \le \lceil -\log_2 \varepsilon - 1 \rceil. \tag{2}$$

The lower bound is trivial, and the upper bound follows from the bisection algorithm which is well defined for any $f \in F^*$ and only uses operations from $\mathrm{ARI}_{\min}$. For $\varepsilon \ge 1/2$, bisection gives $x = 1/2$ with $0 = \lceil -\log_2 \varepsilon - 1 \rceil$ comparisons. For a positive $\varepsilon < 1/2$, bisection uses $k = \lceil -\log_2 \varepsilon - 1 \rceil$ function values and comparisons. The number of additions and multiplications is $2k$.

It is known, see Kung (1976), that at least k function values must be used by any algorithm which computes an ε-approximation for all functions from F^*. Hence, bisection is optimal with respect to the number of function values. Optimality of bisection is preserved also for subsets of F^* which consist of smooth functions, see Sikorski (1985). On the other hand, bisection is *not* optimal in an average case setting (with a Brownian bridge on smooth functions as the probability measure), see Novak, Ritter, Woźniakowski (1995).

We stress that for most classes F of functions and sets ARI studied in this paper, the total complexity, i.e., when all information, arithmetic and comparison operations cost unity, is insensitive to these changes and is always of order $-\log_2 \varepsilon$. Hence, the sensitivity of the topological complexity usually does not correspond to the sensitivity of the total complexity.

We ask whether the upper bound of (2) is sharp, or equivalently, is bisection also optimal with respect to the number of comparisons? The answer depends on F and ARI. The bound is sharp even for small subsets F of F^* if we only allow arithmetic operations that are Hölder on bounded domains. More precisely, let

$$\mathrm{F}_{\mathrm{lin}} \;=\; \{f \in F^* :\; f(x) = ax + b,\, \forall\, x \in [0,1],\; \text{for some } a, b\; \}.$$

That is, $\mathrm{F}_{\mathrm{lin}}$ is a subclass of the class of linear functions. Observe that $f \in F^*$ implies that $f(0) = b < 0$ and $f(1) = a + b > 0$. This implies $a > 0$, however, $|a|$ can be arbitrarily small. The only zero of f is

$$x^*(f) = \frac{-b}{a} = \frac{-f(0)}{f(1) - f(0)} \;\in\; (0,1).$$

If the set ARI contains division then the exact solution may be computed without comparisons, and therefore the topological complexity for $\mathrm{F}_{\mathrm{lin}}$ is zero.

On the other hand, if division is *not* in ARI then, as we shall see, comparisons are needed and the upper bound (2) is sharp. This holds for the set $\mathrm{ARI}_{\mathrm{hol}}$ which is defined as the set of all operations that are Hölder on bounded domains. That is, the operation op belongs to $\mathrm{ARI}_{\mathrm{hol}}$ iff $\mathrm{op} : D_{\mathrm{op}} \to \mathbb{R}$, where the domain $D_{\mathrm{op}} \subset \mathbb{R}^j$ for some j, and for any bounded set $M \subset D_{\mathrm{op}}$ there are constants $\beta, \gamma > 0$ such that

$$|\mathrm{op}(x) - \mathrm{op}(y)| \le \gamma \cdot \|x - y\|^{\beta} \qquad \forall\, x, y \in M. \tag{3}$$

The set $\mathrm{ARI}_{\mathrm{hol}}$ is very large. It obviously contains $\mathrm{ARI}_{\mathrm{min}}$ as well as many of the standard functions. Note that the operation $\mathrm{op}(x,y) = x/y$ with the domain $D_{\mathrm{op}} \subset \mathbb{R}^2$ belongs to $\mathrm{ARI}_{\mathrm{hol}}$ if there exists a positive δ such that the domain D_{op} satisfies the condition

$$(x,y) \,\in\, D_{\mathrm{op}} \quad \text{implies} \quad |y| \,\ge\, \delta.$$

Of course, the division $\mathrm{op}(x,y) = x/y$ with the domain $D_{\mathrm{op}} = \{(x,y) :\; y \ne 0\}$ does *not* belong to $\mathrm{ARI}_{\mathrm{hol}}$. Similarly, the natural logarithm $\mathrm{op}(x) = \log(x)$ with the domain $D_{\mathrm{op}} = \{x : \, x > \delta\}$ belongs to $\mathrm{ARI}_{\mathrm{hol}}$ if $\delta > 0$, but not for $\delta = 0$.

Theorem 1.

$$\mathrm{comp}^{\mathrm{TOP}}(\mathrm{F}_{\mathrm{lin}}, \mathrm{ARI}_{\mathrm{hol}}, \varepsilon) \;=\; \lceil -\log_2 \varepsilon \;-\; 1\rceil.$$

Observe that, contrary to the case of Smale and Vassiliev, the topological complexity depends on ε and goes to infinity as ε approaches zero.

Assume that we want to approximate $-b/a$ for pairs $a, b \in \mathbb{R}$ such that $b < 0$ and $a + b > 0$. If we allow Hölder operations, i.e., no division, then comparisons are necessary and bisection is optimal. Hence, one division can be avoided at the expense of $\lceil -\log_2 \varepsilon \;-\; 1\rceil$ comparisons to compute an ε-approximation of $-b/a$. This result can be compared with other results in complexity theory, see e.g., Strassen (1973).

We now show that the assumptions of Theorem 1 are essential. First we replace the set $\mathrm{ARI}_{\mathrm{hol}}$ by $\mathrm{ARI}_{\mathrm{sgn}} = \mathrm{ARI}_{\min} \cup \{\mathrm{sgn}\}$, where $\mathrm{sgn}(y) = 1$ for $y > 0$, $\mathrm{sgn}(0) = 0$, and $\mathrm{sgn}(y) = -1$ for $y < 0$. That is, we can now compute signs of the function values, $\mathrm{sgn}(f(x))$ for $x \in [0,1]$. How many comparisons do we now need? It turns out that no comparisons are needed. Indeed, this follows from a modification of the bisection algorithm presented in Novak and Woźniakowski (1996).

Hence, the signum function makes the topological complexity zero. Obviously, the signum function is discontinuous. The use of this particular discontinuous operation allows us to eliminate comparisons.

How about continuous operations which are *not* Hölder? Once more, Theorem 1 is not true for the class $\mathrm{F}_{\mathrm{lin}}$ and the set ARI of continuous operations since division is continuous. But what will happen if we enlarge the class $\mathrm{F}_{\mathrm{lin}}$? Again, Theorem 1 does not hold for the class of polynomials of bounded degree. This follows from the result of Renegar (1987) who proved that bisection is far from being optimal for polynomials of bounded degree assuming that the four standard arithmetic operations are used, i.e., $\mathrm{ARI}_{\min}$ and division.

We now show that Theorem 1 does not hold and, in fact, the topological complexity is zero for the class of strictly increasing functions,

$$\mathrm{F}_{\mathrm{inc}} \;=\; \{f \in F^* \,|\; f \text{ is strictly increasing }\}$$

and for the set $\mathrm{ARI}_{\mathrm{ext}}$ of the four standard arithmetic operations, the exponential function $\exp(x)$, $\forall\, x \in \mathbb{R}$, and the natural logarithm $\log(x)$, $\forall x > 0$,

$$\mathrm{ARI}_{\mathrm{ext}} \;=\; \mathrm{ARI}_{\min} \,\cup\, \{\, / \,,\, \exp \,,\, \log \,\}.$$

Observe that all operations in $\mathrm{ARI}_{\mathrm{ext}}$ are continuous.

Theorem 2. *For every $\varepsilon > 0$,*

$$\mathrm{comp}^{\mathrm{TOP}}(\mathrm{F}_{\mathrm{inc}}, \mathrm{ARI}_{\mathrm{ext}}, \varepsilon) \;=\; 0.$$

We present an algorithm (machine) P_ε which computes an ε-approximation and which does not use comparisons. At the computation nodes, the algorithm P_ε uses function values and operations from $\mathrm{ARI}_{\mathrm{ext}}$. Let

$$n \;=\; \lceil 1 + 9/(4\varepsilon) \rceil.$$

The idea of the algorithm is to compute a suitable weighted mean of the equally spaced $x_i = (i-1)/(n-1)$, $i = 1, 2, \ldots, n$. The weights depend on the function values $f(x_i)$ and are chosen such that the weighted mean approximates the zero of the function f. Hence, the weights are now used instead of comparisons to localize the position of the zero. It is convenient to assume that

$$f(0)^2 + f(1)^2 = 1. \tag{4}$$

This can be done without loss of generality since, in what follows, we can replace the $f(x_i)$ by $y_i = f(x_i)/(f(x_1)^2 + f(x_n)^2)$. The algorithm $P_\varepsilon(f)$ is defined as follows:

Step 1: Compute $f(x_1), \dots, f(x_n)$

Step 2: Compute the numbers

$$d_{jk} := (f(x_j) - f(x_k))^2 > 0$$

for all $j, k = 1, \dots, n$ with $j \neq k$.

It is important that each d_{jk} is *strictly* positive. This holds since $f \in F_{\text{inc}}$. We will see later that Theorem 2 does not hold for the slightly larger class of nondecreasing functions.

Step 3: Compute

$$w_{ijk} := \left(\frac{1}{f(x_i)^2 + d_{jk}} \right)^{2n^4/d_{jk}}$$

for $j \neq k$, and

$$w_i := \sum_{j \neq k} w_{ijk},$$

and finally,

$$P_\varepsilon(f) := \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}.$$

All quantities are well defined since $d_{jk} > 0$ and division by 0 does not occur in the definition of w_{ijk}. Since we use $\mathrm{ARI}_{\text{ext}}$, the computation of powers is possible using log and exp, i.e., by using $x^y = \exp(y \log(x))$. We stress that no comparison is used during the computation of $P_\varepsilon(f)$. One can show that $P_\varepsilon(f)$ is an ε-approximation of the zero $x^* = x^*(f)$ for f from F_{inc}, see Novak, Woźniakowski (1996) for details.

Comparing Theorem 1 and 2 we see that Theorem 1 holds for all subsets of F^* that contain the class F_{lin} of linear functions, while Theorem 2 is only proved for the class F_{inc} of strictly increasing functions. Actually, Theorem 2 is not true for the slightly larger class F_{nd} of nondecreasing functions,

$$\mathrm{F}_{\text{nd}} = \{ f \in F^* \mid f \text{ is nondecreasing}, \ f(0) = -1, \ f(1) = 1 \}.$$

We close this section with some results of Peter Hertling. In his paper Hertling (1996) proved results about the power of the set of arithmetic operations

$$\mathrm{ARI}_{\text{abs}} = \{ +, -, *, /, |\cdot| \}.$$

That is, he permits the four standard arithmetic operations and the absolute value. First of all, Hertling (1996) modified Theorem 2 and proved that

$$\mathrm{comp}^{\text{TOP}}(\mathrm{F}_{\text{inc}}, \mathrm{ARI}_{\text{abs}}, \varepsilon) = 0.$$

Hertling also proved that the arithmetic operations $\mathrm{ARI}_{\text{abs}}$ are exponentially better than $\mathrm{ARI}_{\text{hol}}$ for all classes F between F_{nd} and F^*.

Theorem 3 (Hertling 1996). *Assume that* $\mathrm{F_{nd}} \subset F \subset F^*$. *Then*

$$\mathrm{comp}^{\mathrm{TOP}}(F, \mathrm{ARI_{abs}}, \varepsilon) \;=\; \left\lceil \log_2 \left(\left\lceil \log_2(\varepsilon^{-1} + 2) \right\rceil \;-\; 1 \right) \right\rceil .$$

Hence, the topological complexity is roughly $\log_2 \log_2 \varepsilon^{-1}$ which is exponentially smaller than $\log_2 \varepsilon^{-1}$. This also shows that bisection is not optimal for the set $\mathrm{ARI_{abs}}$. Hertling (1996) also proved that adding more continuous operations to the set $\mathrm{ARI_{abs}}$ does not help. Namely, if $\mathrm{ARI_{con}}$ denotes the set of all continuous operations then

$$\mathrm{comp}^{\mathrm{TOP}}(\mathrm{F_{nd}}, \mathrm{ARI_{con}}, \varepsilon) \;=\; \mathrm{comp}^{\mathrm{TOP}}(\mathrm{F_{nd}}, \mathrm{ARI_{abs}}, \varepsilon).$$

One may argue that the most interesting set of arithmetic operations is the *standard set*

$$\mathrm{ARI_{std}} \;=\; \{\, +, -, *, / \,\}.$$

Until recently, not much was known about this set. The following results are from Hertling (2000).

Theorem 4 (Hertling 2000).

- *Bisection is optimal for* F^* *and* $\mathrm{ARI_{std}}$.
- *For the class* $\mathrm{F_{nd}}$, *the topological complexity for the sets* $\mathrm{ARI_{abs}}$ *and* $\mathrm{ARI_{std}}$ *differs at most by 1.*
- *For the class* $\mathrm{F_{inc}}$ *and* $\mathrm{ARI_{std}}$, *the topological complexity is 1 for all* $0 < \varepsilon < 1/2$.

It is interesting to note that the addition of the operation $x \mapsto |x|$ to the set $\mathrm{ARI_{std}}$ does not change much the topological complexity for the class $\mathrm{F_{nd}}$ but gives an exponential improvement for the class F^*.

3. Uniform versus Nonuniform Algorithms

Different notions of computation over the reals were considered before 1989 in algebraic complexity theory and in IBC. These approaches were nonuniform, however, and only problems with a fixed input dimension or a fixed error threshold could be studied. The BSS-model, introduced in Blum, Shub, Smale (1989), is a uniform model which permits varying input dimension or error threshold and generalizes many ideas of discrete complexity theory to computation over more general rings.

Similar as in algebraic complexity theory, most work in IBC assumes the *nonuniform* model. A typical question is the following: given a class F_d of functions $f : [0,1]^d \to \mathbb{R}$ and given a positive ε, what is the complexity (minimal cost) of computing values $S_d(f)$ of a given functional $S_d : F_d \to \mathbb{R}$ up to some error ε? We will illustrate such a problem by the *integration* problem for which $S_d(f) = \int_{[0,1]^d} f(x)\, dx$, see also Section 4. Observe that usually the dimension d, the class F_d, as well as S_d and ε are *given and fixed.* Any algorithm to solve this problem may use this a priori information and the only input of such an algorithm is f from F_d.

In this section we study the “uniform versus nonuniform” alternatives for the BSS-model with an oracle, i.e., for the IBC-model of computation. An example of an uniform algorithm for the integration problem is an algorithm that works for different classes F_d and for different positive ε. Input for such an algorithm might consist of (d, ε, f), where $f \in F_d$. As in Section 2 we assume that $f \in F_d$ is given by an oracle (black box, subroutine) for function values.

In particular, we discuss the role of the error threshold ε. If ε is regarded as a fixed parameter we may have built-in constants of algorithms depending on ε, and this is the case of nonuniform algorithms. On the other hand, if ε is treated as a varying parameter and as one of the inputs, the built-in constants of algorithms are independent of ε, and this is the case of uniform algorithms. We compare the costs of uniform and nonuniform algorithms for approximate solutions of continuous problems assuming the real number model. By the complexity we mean the minimal cost of nonuniform algorithms for computing an ε-approximation. In our paper Novak, Woźniakowski (1999a) we show that, in general, anything can happen:

(1) For some problems, the class of uniform algorithms that compute an ε-approximation is empty even though the class of nonuniform algorithms is non-empty and the complexity is relatively small.
(2) For some problems, the cost of any uniform algorithm is arbitrarily larger than the complexity.
(3) For some problems, there exist uniform algorithms with essentially the same cost as the complexity.

Our examples for problems for which the negative results (1) and (2) hold are rather artificial. Therefore we concentrate here on problems for which the positive result (3) hold. Our first result guarantees the existence of uniform algorithms if we assume the existence of *continuous* or *weakly continuous* nonuniform algorithms. To define the concept of continuity we proceed as follows.

Let $S : F \to G$ be a operator, possibly nonlinear. Here F is a given set and G is assumed to be a metric space with the metric ϱ. We assume that G is a subset of

$$\mathbb{R}^* = \bigcup_{i=1}^{\infty} \mathbb{R}^i$$

consisting of finite sequences of real numbers. We want to approximate $S(f)$ for all f from F by means of oracles $L(f)$ for some functionals L, i.e., $L : F \to \mathbb{R}$. Assume that for all positive ε, there exists a nonuniform algorithm P_ε which computes an ε-approximation. That is,

$$\sup_{f \in F} \varrho\left(S(f), P_\varepsilon(f)\right) \leq \varepsilon.$$

We assume that the built-in constants of the nonuniform algorithm P_ε

$$g_{1,\varepsilon}, \ldots, g_{k(\varepsilon),\varepsilon}$$

are real numbers.

By continuity, we mean that a small change of the numbers $g_{i,\varepsilon}$ does not cause much change in the computed result $P_\varepsilon(f)$. More precisely, for $\delta > 0$ we define a nonuniform algorithm $P_{\varepsilon,\delta}$ in the following way. The algorithms P_ε and $P_{\varepsilon,\delta}$ are identical except that the built-in real numbers $g_{i,\varepsilon}$ of P_ε are replaced by built-in *rational* numbers $g_{i,\varepsilon}^\delta$ for $P_{\varepsilon,\delta}$ such that

$$|g_{i,\varepsilon} - g_{i,\varepsilon}^\delta| \le \delta, \qquad \forall\, i = 1, 2, \ldots, k(\varepsilon). \tag{5}$$

We say that the nonuniform algorithm P_ε is *continuous* if for any positive ε there exists a positive δ such that for *all* rationals $g_{i,\varepsilon}^\delta$ satisfying (5) we have

$$\sup_{f \in F} \varrho\,(S(f), P_{\varepsilon,\delta}(f)) \le 2\varepsilon. \tag{6}$$

We say that the nonuniform algorithm P_ε is *weakly continuous* if for any positive ε there exists a positive δ such that for *some* rationals $g_{i,\varepsilon}^\delta$ satisfying (5) we have (6). We now illustrate continuous and weakly continuous algorithms by an example.

Example 1. We discuss these concepts for integration and quadrature formulas. We show that continuity and weak continuity depend on how the quadrature formulas are implemented. Consider the class

$$F = \{f \in C^1([0,1]) \;\mid\; \|f'\|_\infty \le 1\}$$

with $S(f) = \int_0^1 f(t)\,dt$, $\varrho(a,b) = |a-b|$, and the quadrature formulas

$$P_\varepsilon(f) \;=\; Q_n(f) \;=\; \frac{1}{n}\sum_{i=1}^n f\left(\frac{2i-1}{2n}\right).$$

It is well known that for $n = \lceil 1/(4\varepsilon)\rceil$, the error of P_ε is at most ε.

Observe that any linear algorithm A_ε with a finite worst case error must exactly integrate the constant functions $f_c(\cdot) = c$. This means, $A_\varepsilon(f_c) = S(f_c) = c$, $\forall\, c \in \mathbb{R}$. This condition, obviously, holds for Q_n.

Assume for a moment that $\varepsilon = 1/8$ so that $n = 2$. An obvious way to compute $Q_2(f)$ is by the formula

$$Q_2(f) = g_1\, f(g_2) + g_3\, f(g_4)$$

with the four constants $g_1 = g_3 = \frac{1}{2}$ and $g_2 = \frac{1}{4}$ and $g_4 = \frac{3}{4}$. If programmed this way then Q_2 is *not* continuous. Indeed, if we use g_i^δ such that $g_1^\delta + g_3^\delta \ne 1$ then we do not integrate f_c exactly and the error is infinity. Hence (6) does not hold.

Let us now return to an arbitrary ε. We have to guarantee that the sum of the weights is 1. Clearly, for *some* rational approximations g_i^δ we have $g_1^\delta + g_3^\delta + \cdots + g_{2n-1}^\delta = 1$. In fact we can now even take $g_{2i-1}^\delta = 1/n$. So P_ε is *weakly* continuous.

To obtain continuity we can eliminate one of the constants g_i. For instance, we take $g_{2i-1} = 1/n$ for $i = 2, 3, \ldots, n$ and $g_{2i} = (2i-1)/(2n)$ for $i = 1, 2, \ldots, n$. Then we compute $g_1 = 1 - g_3 - g_5 - \cdots - g_{2n-1}$. A small error in the constants $g_2, g_3, \ldots, g_{2n}$ now leads to a small error in the result, which yields continuity of P_ε.

To illustrate further this point, let us consider integration for the class

$$F = F_k = \{f \in C^k[0,1] \mid \|f^{(k)}\|_\infty \le 1\}.$$

Then

$$Q_n(f) = \sum_{i=1}^n g_i f(t_i)$$

has finite worst case error iff Q_n is exact for all polynomials of degree less than k. Furthermore, for $n = \Theta(\varepsilon^{-1/k})$ there exist $g_i = g_{i,\varepsilon}$ and $t_i = t_{i,\varepsilon}$ for which $P_\varepsilon = Q_n$ has error at most ε.

Hence, for distinct sample points t_i, the constants g_i must satisfy the linear equations

$$Q_n(t^j) = \sum_{i=1}^n g_i t_i^j = S(t^j) = \frac{1}{j+1}, \qquad j = 0, 1, \ldots, k-1.$$

To obtain continuity of P_ε, we have to guarantee that the perturbed knots $t_{i,\varepsilon}^\delta$ and weights $g_{i,\varepsilon}^\delta$ satisfy the same system of linear equations, i.e.,

$$\sum_{i=1}^n g_{i,\varepsilon}^\delta \, (t_{i,\varepsilon}^\delta)^j = \frac{1}{j+1}, \qquad j = 0, 1, \ldots, k-1. \tag{7}$$

Hence, if all $g_{i,\varepsilon}$ and $t_{i,\varepsilon}$ are regarded as built-in constants of P_ε then P_ε is *not* continuous since not all small rational perturbations $g_{i,\varepsilon}^\delta$ and $t_{i,\varepsilon}^\delta$ satisfy (7). On the other hand, P_ε *is* weakly continuous since some small rational perturbations $g_{i,\varepsilon}^\delta$ and $t_{i,\varepsilon}^\delta$ satisfy (7). As before, we can obtain continuity of P_ε by eliminating, say, the first k constants $g_1, g_2, \ldots, g_k$ and compute them as the solution of the linear equations

$$\sum_{i=1}^k g_i (t_{i,\varepsilon}^\delta)^j = \frac{1}{j+1} - \sum_{i=k+1}^n g_{i,\varepsilon}^\delta \, (t_{i,\varepsilon}^\delta)^j, \qquad j = 0, 1, \ldots, k-1.$$

We stress that such elimination of built-in constants is *not* always needed. For instance, if F_k is replaced by the smaller class

$$\widetilde{F}_k = \{f \in C^k[0,1] \mid \|f^{(r)}\|_\infty \le 1 \text{ for } r = 0, \ldots, k\}$$

then we can implement quadrature formulas in the obvious way and still get continuous algorithms.

We are ready to prove that weak continuity implies the existence of uniform algorithms. This will be shown by coding and a magic constant.

Theorem 5. *Assume that for each positive ε we can approximate S to within ε by weakly continuous nonuniform algorithms P_ε. Then there exists a uniform algorithm which computes an ε-approximation of S.*

To obtain a uniform algorithm, we consider the sequence of weakly continuous nonuniform algorithms $\widetilde{P}_{2^{-k}}$ with error $2{\cdot}2^{-k}$. Since each $\widetilde{P}_{2^{-k}}$ needs only finitely many rational built-in numbers, we can code these numbers as well as the whole program $\widetilde{P}_{2^{-k}}$ into a single (natural or rational) number G_k. As in the classical theory of computable functions, this number G_k can be called a *Gödel number*. All these Gödel numbers can be coded into one (magic) real number G and a *universal BSS-machine* can run (simulate) a computation $\widetilde{P}_{2^{-k}}(f)$ on an input G, f and ε with $k = \lceil \log_2 2/\varepsilon \rceil$. See Blum, Shub, Smale (1989) for the construction of a universal machine over the reals.

Theorem 5 does not address the cost of uniform algorithms. We now analyze the cost of uniform algorithms in terms of the minimal cost of nonuniform algorithms. This will be done for linear functionals. That is, $S : F_1 \to \mathbb{R}$ is a linear functional and F_1 is a linear normed space of real functions. We want to approximate $S(f)$ for all f from a given subset F of F_1. We assume that F is convex and balanced. By oracles, we now mean function values.

Let the (nonuniform) complexity be

$$\mathrm{comp}(\varepsilon) \;=\; O\left(\varepsilon^{-1/q} \cdot (-\log \varepsilon)^{\alpha}\right), \quad 0 < \varepsilon < 1/2 \tag{8}$$

for some positive q and nonnegative α.

It is known that the complexity in the nonuniform case is obtained by *linear* algorithms using nonadaptive function values, see Traub, Wasilkowski, Woźniakowski (1988). That is, there exists an integer $n = n(\varepsilon) = O(\varepsilon^{-1/q}(-\log \varepsilon)^{\alpha})$ and a linear P_ε such that

$$P_\varepsilon(f) = \sum_{i=1}^{n} g_{2i-1,\varepsilon}\, f(g_{2i,\varepsilon}) \tag{9}$$

has worst case error $\sup_{f\in F} |S(f) - P_\varepsilon(f)| \le \varepsilon$.

We assume that the built-in constants $g_{i,\varepsilon}$ of P_ε are uniformly bounded,

$$|g_{i,\varepsilon}| \le M, \qquad \forall\, i = 1, 2, \ldots, 2n(\varepsilon),\ \forall \varepsilon > 0, \tag{10}$$

for some $M \ge 2$.

We assume that P_ε is weakly continuous. That is, for a positive ε there exists a positive δ such that the linear algorithm

$$P_{\varepsilon,\delta}(f) = \sum_{i=1}^{n} g^{\delta}_{2i-1,\varepsilon}\, f(g^{\delta}_{2i,\varepsilon})$$

with some rational $g^{\delta}_{i,\varepsilon}$ satisfying $|g_{i,\varepsilon} - g^{\delta}_{i,\varepsilon}| \le \delta$ has error at most 2ε,

$$\sup_{f\in F} |S(f) - P_{\varepsilon,\delta}(f)| \le 2\varepsilon. \tag{11}$$

Without loss of generality we may assume that the perturbed constants $g^\delta_{i,\varepsilon}$ are bounded by M, i.e., $|g^\delta_{i,\varepsilon}| \le M$. We additionally assume that

$$O(\log M + \log n) \tag{12}$$

bits are sufficient to represent each $g^\delta_{i,\varepsilon}$ exactly.

We add in passing that for the classes F_k discussed above it is relatively easy to check that (12) holds.

Theorem 6. *Under the above assumptions (8-12) there exists a uniform algorithm P which computes an ε-approximation to S with cost*

$$\mathrm{cost}(\varepsilon) \;=\; O\left(\varepsilon^{-1/q} \cdot (-\log\varepsilon)^{\alpha+1}\right).$$

That is, the cost of P is at most $\log 1/\varepsilon$ larger than the (nonuniform) complexity.

We do not claim that the positive results of Theorem 5 and Theorem 6 (both are from Novak, Woźniakowski (1999a)) are always practical. All these results are based on some kind of coding and to use such uniform algorithms one has, in general, to find out and store at least one *magic* real number. This magic number is needed *with arbitrary precision*, i.e., exactly. Such an assumption is not realistic in practical computation where usually floating point arithmetic is used, and all numbers are rounded and only their approximations are used. This suggests that one should also study algorithms where only integer or rational numbers are allowed as built-in constants, see Hemmerling (1998).

To address the practical side, we note that the difference between uniform and nonuniform algorithms basically disappears if we do not let ε go to zero, and instead consider ε in the interval $[\varepsilon_{\mathrm{low}}, \varepsilon_{\mathrm{upp}}]$ with, say, $\varepsilon_{\mathrm{low}} = 10^{-8}$ and $\varepsilon_{\mathrm{upp}} = 10^{-2}$. Then no magic numbers are needed to construct a uniform algorithm which works well for all $\varepsilon \in [\varepsilon_{\mathrm{low}}, \varepsilon_{\mathrm{upp}}]$, see Novak, Woźniakowski (1999a) for more details.

4. When are Integration and Discrepancy Tractable?

One of the most important problems in computer science and mathematics concerns the tractability of problems. Usually this problem, and the famous $P = NP$ problem, is posed for discrete decision problems and the Turing machine. The same problem was generalized to the real number model in Blum, Shub, Smale (1989).

One can also study the class of tractable problems in the case of partial information, i.e., in the IBC model, see Woźniakowski (1994a, 1994b). The basic question is the following:

For which problems does there exist an algorithm which computes an ε-approximation with cost bounded by a polynomial in ε^{-1} and the input size?

By "input size" we mean here the number d of variables of a function. 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. 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.

In this section we present some recent results for numerical integration in the worst case setting, see Novak, Woźniakowski (2000b) for more details and a survey of recent results. For some classes of functions integration is related to various notions of *discrepancy*, such as the L_2-discrepancy and the $*$-discrepancy. Discrepancy is widely used and studied in many areas of mathematics.

Let F_d be a normed space of integrable functions $f : D_d \to \mathbb{R}$ where $D_d = [0,1]^d$. For $f \in F_d$ we want to approximate the multivariate integral

$$I_d(f) = \int_{D_d} f(t)\, dt. \tag{13}$$

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 for which d is often very large, say in the hundreds or thousands. QMC algorithms are of the form

$$Q_{n,d}(f) = \frac{1}{n} \sum_{i=1}^{n} f(t_i). \tag{14}$$

  The sample points t_i are deterministic, belong to D_d, and may depend on n and d as well as on 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.
- 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

$$Q_{n,d}(f) = \sum_{i=1}^{n} a_i f(t_i), \qquad a_i \geq 0. \tag{15}$$

  The coefficients a_i and the sample points t_i are deterministic and may depend on n, d and F_d.
- The class LIN is the class of *linear* algorithms of the form (15) with arbitrary real weights a_i. In this case, some 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 long as the class F_d is convex and symmetric, it is not necessary to study more general algorithms, see Traub, Wasilkowski, Woźniakowski (1988). That is, the worst case error is minimized by linear algorithms using non-adaptive sample points. This important result is due to Bakhvalov and Smolyak.

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,

$$e(Q_{n,d}) \;=\; \sup_{f\in F_d,\,\|f\|_d\le 1} \left|I_d(f) \;-\; Q_{n,d}(f)\right|. \tag{16}$$

For $n = 0$ we do not sample the function and we[2] set $Q_{0,d} = 0$. Then

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

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 *normalized 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 or LIN, and let

$$e(n, F_d, \mathrm{A}) \;=\; \inf\{\, e(Q_{n,d}) \;:\; Q_{n,d} \;\in\; \mathrm{A} \,\} \tag{17}$$

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

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

Since QMC $\subset$ POS $\subset$ LIN we also have

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

We now formally define tractability for the class A$\in$ $\{$QMC, POS, LIN$\}$. We would like to reduce the initial 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,

$$n(\varepsilon, F_d, A) \;=\; \min\{\, n \;:\; e(n, F_d, A) \;\le\; \varepsilon\, e(Q_{0,d}) \,\}\,. \tag{18}$$

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

$$n(\varepsilon, F_d, A) \;\le\; C\, d^{\,q}\, \varepsilon^{-p} \qquad \forall\, d = 1, 2, \ldots\,,\ \forall\, \varepsilon \in (0,1). \tag{19}$$

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}. If $q = 0$ in the bound above we say that the problem is *strongly* tractable, and the infimum of p satisfying the bound above is called the *strong exponent*.

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

We stress that the minimal number $n(\varepsilon, F_d, \mathrm{LIN})$ of function samples is directly related to the complexity, which is the minimal cost of computing an approximation with error ε.

Discrepancy is a quantitative measure of the lack of uniformity of the 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 Drmota, Tichy (1997), Matoušek (1999), Niederreiter (1992), and Tezuka (1995).

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 some bounds. Then we discuss the L_p-star discrepancy for finite p and the $*$-discrepancy, where $p = \infty$.

Let $x = [x_1, \dots, x_d] \in [0,1]^d$. 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 $t_1, \dots, t_n \in [0,1]^d$, 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 generalize 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

$$\mathrm{disc}(x) \;:=\; x_1 \cdots x_d \;-\; \sum_{i=1}^n a_i 1_{[0,x)}(t_i). \tag{20}$$

The 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 (20),

$$\mathrm{disc}_2(\{t_i\}, \{a_i\}) \;=\; \left(\int_{[0,1]^d} \left(x_1 \cdots x_d \;-\; \sum_{i=1}^n a_i 1_{[0,x)}(t_i) \right)^2 \, dx \right)^{1/2}. \tag{21}$$

By direct integration, we have the explicit formula

$$\mathrm{disc}_2^2(\{t_i\}, \{a_i\}) \;= \tag{22}$$
$$\frac{1}{3^d} - \frac{1}{2^{d-1}} \sum_{i=1}^n a_i \prod_{k=1}^d (1 - t_{i,k}^2) + \sum_{i,j=1}^n a_i a_j \prod_{k=1}^d \left(1 - \max(t_{i,k}, t_{j,k})\right),$$

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

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 minimized. 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

$$\overline{\mathrm{disc}}_2(0,d) \;=\; \mathrm{disc}_2(0,d) \;=\; \left(\int_{[0,1]^d} x_1^2 \cdots x_d^2 \, dx\right)^{1/2} \;=\; 3^{-d/2}. \tag{23}$$

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

We now discuss bounds on the normalized L_2-discrepancy. By the normalized L_2-discrepancy we mean $\mathrm{disc}_2(\{t_i\}, \{a_i\})/\mathrm{disc}_2(0,d)$. That is, we normalize by the initial value of the L_2-discrepancy, which is $3^{-d/2}$. As we shall see, this case is directly related to tractability of integration for some spaces F_d when we reduce the initial error by a factor of ε. Similarly to (18), we define

$$\overline{n}(\varepsilon,d) \;=\; \min\{\, n \;:\; \overline{\mathrm{disc}}_2(n,d) \;\le\; \varepsilon\, \mathrm{disc}_2(0,d) \,\}, \tag{24}$$

$$n(\varepsilon,d) \;=\; \min\{\, n \;:\; \mathrm{disc}_2(n,d) \;\le\; \varepsilon\, \mathrm{disc}_2(0,d) \,\} \tag{25}$$

and ask whether $\overline{n}(\varepsilon,d)$ and $n(\varepsilon,d)$ are polynomial in ε^{-1} and d. Recent results show that the normalized L_2-discrepancy is intractable.

Theorem 7.

$$\overline{n}(\varepsilon,d) \;\ge\; (9/8)^d\,(1-\varepsilon^2). \tag{26}$$

The bound (26) is also valid if $a_i \ge 0$, which corresponds to the assumption defining the class POS. For arbitrary a_i,

$$n(\varepsilon,d) \;\ge\; 1.0628^d\,(1+o(1)) \qquad \text{as } d \to \infty, \tag{27}$$

the lower bound holding for any fixed $\varepsilon < 1$.

The bound (26) was proven in Woźniakowski (1999), see also Sloan, Woźniakowski (1998). The bound (27) is from Novak, Woźniakowski (1999b).

The L_2-discrepancy is related to uniform integration for Sobolev spaces and the Wiener sheet measure. Consider the classical Sobolev space $W_2^1([0,1])$ of absolutely 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 $W_2^1([0,1])$,

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

i.e., the completion of the algebraic tensor product. 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

$$K_d(x,t) \;=\; \prod_{j=1}^d \min(x_j,t_j). \tag{28}$$

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 $e(Q_{n,d})$ of $Q_{n,d}$. It is known that

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

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 (20), we immediately obtain

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

where $1-t_i=[1-t_{i,1},\dots,1-t_{i,d}]$.

From the normalized 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 normalized error, i.e., for reduction of the initial error.

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

$$\mathrm{disc}_p^*(t_1,\dots,t_n) = \left(\int_{[0,1]^d} \left| x_1\cdots x_d - \frac{1}{n}\sum_{i=1}^n 1_{[0,x)}(t_i)\right|^p dx\right)^{1/p}, \tag{29}$$

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

$$\mathrm{disc}_\infty^*(t_1,\dots,t_n) = \sup_{x\in[0,1]^d} \left| x_1\cdots x_d - \frac{1}{n}\sum_{i=1}^n 1_{[0,x)}(t_i)\right| \tag{30}$$

for $p=\infty$. It is customary to denote the L_∞-star discrepancy as the $*$-discrepancy. Let

$$\operatorname{disc}_p^*(n,d) \;=\; \inf_{t_1,\dots,t_n} \operatorname{disc}_p^*(t_1,\dots,t_n)$$

denote the minimal L_p-discrepancy for n points. Note that for $n=0$ we get the initial discrepancy $\operatorname{disc}_p^*(0,d) = (1/(p+1))^{d/p}$. This shows that for $p<\infty$, the initial discrepancy in the L_p-norm goes exponentially fast to zero as d approaches infinity. For $p=\infty$, we have $\operatorname{disc}_\infty^*(0,d)=1$ and the initial discrepancy is properly normalized. Finally, let

$$n_p^*(\varepsilon,d) \;=\; \min\{\, n \,:\, \operatorname{disc}_p^*(n,d) \;\le\; \varepsilon\,\}.$$

The usual bounds on the L_p-star discrepancy are for fixed dimension d and large n. The asymptotic behavior of $\operatorname{disc}_p^*(n,d)$ with respect to n is known, see once more Drmota, Tichy (1997), Matoušek (1999), and Niederreiter (1992). This yields to

$$\begin{aligned} n_p^*(\varepsilon,d) &= \Theta\left(\varepsilon^{-1}(\log 1/\varepsilon)^{(d-1)/2}\right) \quad \text{for } p\in(1,\infty),\\ n_p^*(\varepsilon,d) &= O\left(\varepsilon^{-1}(\log 1/\varepsilon)^{(d-1)a_p}\right) \quad \text{for } p\in\{1,\infty\}, \end{aligned}$$

where $a_1=1/2$ and $a_\infty=1$, as ε tends to zero. Here O- and Θ-factors depend on d.

The question of dependence on d for $p=\infty$ was raised by Larcher who asked whether there exists an $a>1$ such that $\operatorname{disc}_\infty^*(\lceil a^d\rceil,d)$ tends to 1 as d goes to infinity, and also asked whether, in particular, $\operatorname{disc}_\infty^*(2^d,d)$ goes to 1 as d goes to infinity. Based on the results for $p=2$ one might have been 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 Heinrich, Novak, Wasilkowski, Woźniakowski (1999), it is shown that $n_\infty^*(\varepsilon,d)$ depends only *polynomially* on d and n^{-1}.

Theorem 8. *There exists a positive number C such that*

$$n_\infty^*(\varepsilon,d) \;\le\; C\,d\,\varepsilon^{-2} \qquad \forall\, n,d=1,2,\dots. \tag{31}$$

Hence, the inverse of the $*$-discrepancy depends at most linearly on d. This dependence on d cannot be improved. The proof of (31) follows from deep results from the theory of empirical processes. The proof is non-constructive, and we do not know for which points the bound (31) holds.

Acknowledgment: We wish to thank Joseph Traub for valuable comments.

References

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

2. L. Blum, F. Cucker, M. Shub, S. Smale (1998): *Complexity and Real Computation.* Springer, New York.

3. M. Drmota, R. F. Tichy (1997): *Sequences, Discrepancies and Applications.* Lecture Notes in Mathematics **1651**, Springer.

4. S. Heinrich, E. Novak, G. W. Wasilkowski, H. Woźniakowski (2000): The inverse of the star-discrepancy depends linearly on the dimension. To appear in ACTA Arithmetica.

5. A. Hemmerling (1998): Computability of string functions over algebraic structures. Math. Logic Quarterly **44**, 1–44.

6. P. Hertling (1996): Topological complexity with continuous operations. J. Complexity **12**, 315–338.

7. P. Hertling (2000): Topological complexity of zero finding with algebraic operations. Submitted for publication.

8. H. T. Kung (1976): The complexity of obtaining starting points for solving operator equations by Newton's method. In "Analytical Computational Complexity" (J. F. Traub, ed.), 35-75, Academic Press, New York.

9. J. Matoušek (1999): *Geometric Discrepancy.* Springer-Verlag, Berlin.

10. K. Meer, C. Michaux (1997): A survey on real structural complexity theory. Bull. Belg. Math. Soc. **4**, 113–148.

11. H. Niederreiter (1992): *Random Number Generation and Quasi-Monte Carlo Methods.* SIAM, Philadelphia.

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

13. E. Novak, K. Ritter, H. Woźniakowski (1995): Average Case Optimality of a Hybrid Secant-Bisection Method. Math. Computation **64**, 1517–1539.

14. E. Novak, H. Woźniakowski (1996): Topological complexity of zero finding, J. Complexity **12**, 380–400.

15. E. Novak, H. Woźniakowski (1999a): On the cost of uniform and nonuniform algorithms. Theoretical Computer Science **219**, 301–318.

16. E. Novak, H. Woźniakowski (1999b): Intractability results for integration and discrepancy. J. Complexity, to appear.

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

18. E. Novak, H. Woźniakowski (2000b): When are integration and discrepancy tractable? To appear in the FOCM Proceedings, Oxford, 1999.

19. L. Plaskota, *Noisy Information and Computational Complexity* Cambridge Univ. Press, Cambrigde, 1996.

20. J. Renegar (1987): On the worst-case arithmetic complexity of approximating zeros of polynomials. J. Complexity **3**, 90–113.

21. K. Sikorski (1985): Optimal solution of nonlinear equations. J. Complexity **1**, 197-209.

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

23. S. Smale (1987): On the topology of algorithms. J. Complexity **3**, 81–89.

24. V. Strassen (1973): Vermeidung von Divisionen. J. Reine Angew. Math. **264**, 184–202.

25. S. Tezuka (1995): *Uniform Random Numbers: Theory and Practice.* Kluwer, Boston.

26. J. F. Traub, G. W. Wasilkowski, H. Woźniakowski (1988): *Information-Based Complexity.* Academic Press, New York.

27. J. F. Traub, H. Woźniakowski (1980): *A general theory of optimal algorithms.* Academic Press, New York.

28. V. A. Vassiliev (1992): Complements of discriminants of smooth maps: topology and applications. Transl. of Math. Monographs 98, 1992, revised 1994, Amer. Math. Soc., Providence, R.I.

29. V. A. Vassiliev (1996): Topological complexity of root-finding algorithms. In: The Mathematics of Numerical Analysis, J. Renegar, M. Shub, S. Smale (eds.), Lectures in Applied Mathematics 32, AMS, 831–856.

30. H. Woźniakowski (1994a): Tractability and strong tractability of linear multivariate problems. J. Complexity **10** (1994), 96–128.

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

32. H. Woźniakowski (1998): Why does information-based complexity use the real number model? Theoretical Computer Science, **219**, 451–466.

33. H. Woźniakowski (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.

Mathematisches Institut, Universität Jena, Ernst-Abbe-Platz 4, D-07740 Jena, Germany

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

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

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