Tractability of Tensor Product Linear Operators

Erich Novak, Ian H. Sloan and Henryk Woźniakowski

April 1997

Abstract. This paper deals with the worst case setting for approximating multivariate tensor product linear operators defined over Hilbert spaces. Approximations are obtained by using a number of linear functionals from a given class of information. We consider the three classes of information: the class of all linear functionals, the Fourier class of inner products with respect to given orthonormal elements, and the standard class of function values.

We wish to determine which problems are tractable and which strongly tractable. The complete analysis is provided for approximating operators of rank two or more. The problem of approximating linear functionals is fully analyzed in the first two classes of information. For the third class of standard information we show that the possibilities are very rich. We prove that tractability of linear functionals depends on the given space of functions. For some spaces all nontrivial normed linear functionals are intractable, whereas for other spaces all linear functionals are tractable. In "typical" function spaces, some linear functionals are tractable and some others are not.

1 Introduction

We study multivariate tensor product linear operators defined over Hilbert spaces. The d-variate linear operator S_d is obtained by taking d-fold tensor product of the continuous linear operator S_1. We consider the worst case setting, in which we want to approximate S_d over the unit ball with error at most ε. Approximations of S_d are obtained by using a number of continuous linear functionals from a given class of information.

The problem is said to be *tractable* iff the number of linear functionals needed to approximate S_d with error at most ε is polynomial in d and $1/\varepsilon$, and is said to be *strongly tractable* iff the number of linear functionals does not depend on d and is polynomial in $1/\varepsilon$.

We are mainly interested in characterizing which problems are tractable and which strongly tractable.

We consider *three* classes of information. The first class is the class of all linear functionals. For this class, it is known, see [8, 9], that tractability is equivalent to strong tractability, and that strong tractability holds iff either S_1 is a linear functional, or $\|S_1\| < 1$ and singular values of S_1 go polynomially to zero[1].

The second class is the *Fourier* class of information. This class consists of inner products with respect to products of given orthonormal elements. The analysis of the Fourier class seems to be new. If the domain space of S_1 is not spanned by the given orthonormal elements then we may be not able to approximate S_d even if S_1 is a linear functional. On the other hand, if the domain space of S_1 is spanned by the given orthonormal elements then we can approximate S_d by using finitely many linear functionals iff S_1 is compact. We provide necessary and sufficient conditions for tractability and strong tractability. As with the first class, tractability and strong tractability are equivalent in the Fourier class. Strong tractability holds iff either S_1 can be approximated with an arbitrarily small error by one inner product from the Fourier class, or $\|S_1\| < 1$ and the nth minimal errors[2] of approximating S_1 go polynomially to zero.

The third class is the class of standard information, which consists of function values. In this case we assume that the domain space of S_1 is a reproducing kernel Hilbert space of univariate functions. Standard information is probably the most important from a practical point of view. There are many papers analyzing this class. In particular, it is known, see [7], that if S_1 is at least two-dimensional[3] then tractability is again equivalent to strong tractability, and strong tractability holds iff $\|S_1\| < 1$ and the nth minimal errors of approximating S_1 go polynomially to zero.

The unresolved case for the class of standard information is when S_1 is a linear functional. We show that the results for this case are very rich in possibilities. First of all, there exist domain spaces of S_1 (even of infinite dimension) such that all problems S_d are strongly tractable. In fact, it is enough to compute only *one* function value to get an ε-approximation, and this holds for arbitrarily small positive ε. Such spaces can even be subspaces of continuous functions. Their construction is related to Peano curves.

Let us now assume that one function value is not enough to get an ε-approximation for arbitrarily small positive ε. We then have two cases. The first one is $\|S_1\| < 1$. Then tractability and strong tractability are equivalent, and strong tractability holds iff the nth

[1]A sequence λ_n goes polynomially to zero iff there exists a positive k such that $\lambda_n = O(n^{-k})$

[2]By the nth minimal error we mean the minimal error of approximations that use at most n linear functionals from the given class of information.

[3]We assume here that $\dim(S_1(F_1)) \geq 2$, where F_1 is the domain space of S_1.

minimal errors for approximating S_1 go polynomially to zero, see [7].

The second case is $\|S_1\| \geq 1$. Then the problem is *not* strongly tractable. (This result has been proven in [7] under an additional assumption.) To approximate S_d we have to compute at least d function values for small ε. The last bound is sharp, since for some domain spaces it is enough to compute $d+1$ linear functionals to solve *all* S_d even exactly, i.e., with $\varepsilon = 0$. In this case, we have tractability. On the other hand, for some other domain spaces, *all* problems S_d are intractable. Hence, tractability of linear functionals with $\|S_1\| \geq 1$ depends on the given space of functions. We provide conditions on tractability and intractability of linear functionals. In "typical" function spaces these conditions are satisfied for some linear functionals. That is, the classes of tractable and intractable linear functionals are each in general nonempty. For a given linear functional, such as integration or weighted integration, it is usually hard to verify to which class it belongs. Recently, an intractability result for multivariate integration was proved in [4] for the Korobov class of functions, which is different from the classes studied here.

2 Formulation of the Problem

In this section we define multivariate linear tensor product problems, as well as the three classes of information which are used for their approximation. Then we define the concepts of tractability and strong tractability for such problems.

Let F_1 and G_1 be Hilbert spaces over the real field. The inner products in F_1 and in G_1 are denoted by $\langle \cdot, \cdot \rangle_{F_1}$ and $\langle \cdot, \cdot \rangle_{G_1}$. We stress that F_1 or G_1 need not be separable.

For $d \geq 2$, define the Hilbert space $F_d = F_1 \otimes \cdots \otimes F_1$ (d times) as a tensor product of F_1's. That is, F_d is the completion of linear combinations of tensor products $f_1 \otimes \cdots \otimes f_d$, which we write for simplicity as $f_1 f_2 \cdots f_d$, with $f_i \in F_1$. For the reader's convenience we recall that the tensor product $f = f_1 \otimes \cdots \otimes f_d = f_1 f_2 \cdots f_d$ of numbers f_k is just the product $\prod_{k=1}^{d} f_k$, while for univariate functions f_k it is a function of d variables $f(t_1, \cdots, t_d) = \prod_{k=1}^{d} f_k(t_k)$. The inner product in F_d is defined for $f = f_1 \cdots f_d$ and $h = h_1 \cdots h_d$ with $f_i, h_i \in F_1$ as

$$\langle f, h \rangle_{F_d} = \prod_{j=1}^{d} \langle f_i, h_i \rangle_{F_1}.$$

Similarly we define the Hilbert space $G_d = G_1 \otimes \cdots \otimes G_1$ with the inner product $\langle \cdot, \cdot \rangle_{G_d}$.

Let $S_1 : F_1 \to G_1$ be a linear continuous operator. For $d \geq 2$, we define $S_d : F_d \to G_d$ as a linear continuous operator which is the d-fold tensor product of S_1. More precisely, $S_d = S_1 \otimes \cdots \otimes S_1$ (d times) such that for $f = f_1 f_2 \cdots f_d$ with $f_i \in F_1$ we have

$$S_d \, f \; = \; S_1 f_1 \, S_1 f_2 \, \cdots \, S_1 f_d \, \in G_d.$$

By a multivariate linear tensor product problem (or shortly the problem) we mean the sequence of the triples $\{S_d, F_d, G_d\}$.

We shall devote considerable attention to the case of a linear continuous functional S_1. That is, $G_1 = \mathbb{R}$. Then there exists an element $h \in F_1$ such that $S_1 f = \langle f, h \rangle_{F_1}, \forall\, f \in F_1$, and $\|S_1\| = \|h\|$. For $d \geq 2$, S_d is also a linear continuous functional, $G_d = \mathbb{R}$, and

$$S_d \, f \; = \; \left\langle f, h^d \right\rangle_{F_d},$$

where $h^d = h \, h \cdots h \in F_d$.

If F_1 is a space of univariate functions defined on, say, the interval $[0,1]$ then F_d is a space of multivariate functions defined on the d-dimensional unit cube $[0,1]^d$ and $h^d(t) = h(t_1)h(t_2)\cdots h(t_d)$ with $t = [t_1, t_2, \ldots, t_d]$.

We wish to approximate the elements $S_d f$ for f from the unit ball of F_d. That is, for a given nonnegative ε, we want to compute for each f an approximation $U_d(f)$ such that the worst case error $e(U_d)$ does not exceed ε. Here the error is given by

$$e(U_d) \; = \sup_{f \in F_d, \, \|f\| \leq 1} \| \, S_d f - U_d(f) \, \|_{G_d}. \tag{1}$$

We now explain how the elements $U_d(f)$ can be constructed. We assume that the element f is not known explicitly. Instead we may gather information about f by computing a number of linear continuous functionals on f. These functionals are from a specific class Λ_d of information which is always a subset of F_d^*. As mentioned in the introduction, in this paper we consider three classes of information:

1. The class of all linear information, $\Lambda_d = F_d^*$. That is, we can now compute linear continuous functionals $\langle \cdot, \eta \rangle_{F_d}$ for arbitrary η's from F_d. The results for this class, summarized in Section 3, are not new, but are included here for completeness.

2. The class of Fourier information, $\Lambda_d = \Lambda_d^{\mathrm{Fou}}$. This class, which does not seem to have been considered before, is defined as follows. Let $d = 1$. Here we assume that an arbitrary

orthonormal system $\{\eta_i \mid i \in I\}$ is given, where I is a set of indices which may be finite, countable or even uncountable. We assume that we can compute the inner products $\langle f, \eta_i \rangle_{F_1}$ for $i \in I$. The set $\{\eta_i\}$ may or may not form an orthogonal basis of F_1. If this is true then we can compute Fourier coefficients of f with respect to the given basis.

For $d \geq 2$, we assume that we can compute the tensor products of the one-dimensional functionals. That is, we can now compute inner products of the form $\langle f, \eta_{i_1}\eta_{i_2}\cdots\eta_{i_d} \rangle_{F_d}$ for any $i_j \in I$.

Once more, if the η_i form an orthonormal basis of F_1 then the $\{\eta_{i_1}\eta_{i_2}\cdots\eta_{i_d}\}$ form an orthonormal basis of F_d. In this case, we can compute Fourier coefficients of elements from the space F_d.

That is why we call this class the class of Fourier information, even if the system $\{\eta_i\}$ is not complete. We stress that for the Fourier class the elements η_i are fixed and the inner products with η_i are used for approximation of all linear operators S_d. Obviously, Λ_d^{Fou} is a proper subset of F_d^*. The difference between the classes of linear and Fourier information is that for the class F_d^* we may select an orthonormal system which is suitable for approximating the operators S_d, whereas for the class Λ_d^{Fou} we use the same orthonormal system based on η_i, independently of the operators S_d. Properties for this class are established in Section 4.

3. The class of standard information, $\Lambda_d = \Lambda_d^{\mathrm{std}}$. This class consists of function values. More precisely, for this class we assume that the space F_1 consists of univariate functions f defined on a given domain, say D, and for which the linear functional $f(t), \forall\, f \in F_1$, is continuous for any $t \in D$. This is equivalent, see [1], to the assumption that F_1 has a reproducing kernel $K_1 : D^2 \to \mathbb{R}$ with $K_1(\cdot, t) \in F_1$ and

$$f(t) \;=\; \langle f, K_1(\cdot, t) \rangle_{F_1}, \;\; \forall\, f \in F_1, \forall\, t \in D.$$

For $d \geq 2$, the space F_d has also a reproducing kernel $K_d : D^{2d} \to \mathbb{R}$ and

$$K_d(x, t) \;=\; \prod_{i=1}^{d} K_1(x_i, t_i), \;\; \forall\, x, t \in D^d,$$

with $x = [x_1, \ldots, x_d]$ and $t = [t_1, \ldots, t_d]$.

For the class of standard information we assume that we can compute function values $f(t) = \langle f, K_d(\cdot, t) \rangle_{F_d}$ for any $t \in D^d$. Obviously, Λ_d^{std} is a subset of F_d^*. Relations between the classes Λ_d^{std} and Λ_d^{Fou} depend on the space F_1 and on the sequence of the η_i. As we shall see, for some cases these two classes are the same, whereas for other cases they are different and yield completely different results for multivariate linear tensor product problems.

In particular, we shall see in Section 5 that the results for the approximation of continuous linear functionals are very rich in the possibilities they allow in the case of standard information. This is the largest part of the paper.

We are finally ready to define tractability and intractability concepts. Let Λ_d be one of the three classes defined above. Suppose we compute n such functionals,

$$N(f) \,=\, [L_1(f),\dots,L_n(f)]\,, \quad L_i \in \Lambda_d.$$

Then the approximation $U_d(f)$ is given as a linear combination[4] of $L_i(f)$,

$$U_d(f) \,=\, \sum_{i=1}^{n} g_i L_i(f), \qquad \text{for some} \quad g_i \in G_d. \tag{2}$$

The worst case error of U_d is defined by (1). For fixed n and the class Λ_d, let $e(n,\Lambda_d)$ denote the minimal error which can be achieved by computing n functionals from the class Λ_d,

$$e(n,\Lambda_d) \,=\, \min\{e(U_d) :\ U_d \text{ is of the form (2)}\,\}.$$

We want to guarantee that the worst case error is at most ε. The smallest n for which this holds is called the complexity[5] of the multivariate linear tensor product problem $\{S_d,F_d,G_d\}$,

$$\mathrm{comp}(\varepsilon,\Lambda_d) \,=\, \min\{n :\ e(n,\Lambda_d) \,\le\, \varepsilon\,\}.$$

We listed as the arguments of the complexity only ε and Λ_d since we want to study the dependence on ε, d and the class Λ_d of information.

We say that the multivariate linear tensor product problem $\{S_d,F_d,G_d\}$ is *tractable* (in the class Λ_d) if its complexity is bounded by a polynomial in $1/\varepsilon$ and d, i.e., there exist nonnegative numbers C,p and q such that

$$\mathrm{comp}(\varepsilon,\Lambda_d) \,\le\, C\,\varepsilon^{-p}\,d^q, \quad \forall\,\varepsilon\le 1,\ \forall\,d=1,2\dots\,.$$

The smallest (or the infimum) of p or q, respectively, is called the exponent with respect to ε^{-1} or the exponent with respect to d.

[4]We restrict ourselves to nonadaptive information and linear algorithms since for the worst case setting considered in this paper, adaption and nonlinear algorithms are not better, see e.g., [6].

[5]Usually, the complexity is defined as the minimal cost needed to compute approximations with error at most ε. In our case, the minimal cost is proportional to the smallest n and therefore we choose this simplified definition of complexity.

The problem is *strongly tractable* (in the class Λ_d) iff q above is zero, i.e.,

$$\mathrm{comp}(\varepsilon, \Lambda_d) \le C\,\varepsilon^{-p} \quad \forall\, \varepsilon \le 1,\ \forall\, d = 1, 2 \ldots .$$

The smallest (or the infimum) of p above is called the strong exponent.

Finally, the problem is called *intractable* iff it is not tractable. For more detailed discussion of these concepts the reader is referred to [8, 9].

3 The Class of Linear Information

The class of linear information has been studied in many papers, and the complexity of many problems is known for this class, see e.g., [6] and papers cited there. Tractability and strong tractability issues have been studied in [8, 9]. In this section we briefly review necessary and sufficient conditions on tractability and strong tractability for the multivariate linear tensor product problems in the class of linear information.

Let $d = 1$. It is well known that $\mathrm{comp}(\varepsilon, F_1^*)$ is finite for all positive ε iff the linear operator S_1 is compact, see e.g., Chapter 4 of [6]. Hence, for a noncompact S_1 the problem is intractable. Assume then that S_1 is compact. Let $W_1 = (S_1^* S_1)^{1/2} : F_1 \to F_1$. Then W_1 is a compact, self adjoint and nonnegative definite operator. Let (ζ_i, λ_i) be its orthonormal eigenpairs,

$$W_1 \zeta_i = \lambda_i\, \zeta_i,$$

with $\lambda_1 \ge \lambda_2 \ge \ldots \ge 0$, and $\langle \zeta_i, \zeta_j \rangle_{F_1} = \delta_{i,j}$. Here i and j vary from $1, 2, \ldots, \dim(F_1)$. If $\dim(F_1)$ is finite then we formally set $\lambda_i = 0$ for $i > \dim(F_1)$.

For $d \ge 2$, the operator S_d is also compact and $W_d = (S_d^* S_d)^{1/2}$ is compact, self adjoint and nonnegative definite with orthonormal eigenpairs $(\zeta_{i_1}\zeta_{i_2}\cdots\zeta_{i_d}, \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_d})$. As with the case $d = 1$, the complexity $\mathrm{comp}(\varepsilon, F_d^*)$ is finite for all positive ε and all d.

The behavior of $\mathrm{comp}(\varepsilon, F_d^*)$ depends on the singular values λ_i of S_1. Assume first that $\lambda_1 = 0$. Then $S_d = 0$, and the problem is trivially strongly tractable with strong exponent zero.

Assume next that $\lambda_1 > 0$ and that $\lambda_2 = 0$. This implies, for $j \ge 2$, that $\|S_1\zeta_j\|^2 = (S_1\zeta_j, S_1\zeta_j) = (\zeta_j, W_1^2\zeta_j) = 0$, and hence $S_1\zeta_j = 0$. This means that S_1 is an operator of rank 1, i.e., its image has dimension 1, and $S_1 f = \langle f, \zeta_1 \rangle_{F_1} g$ with $g = S_1\zeta_1 \in G_1$ and $\lambda_1 = \|g\|_{G_1}$. Hence, $S_1 f$ can be recovered exactly by computing one linear functional $\langle f, \zeta_1 \rangle_{F_1}$. For $d \ge 2$, we have $S_d f = \langle f, \zeta_1\zeta_1\cdots\zeta_1 \rangle_{F_d} g^d$ for all $f \in F_d$, which can be recovered exactly by computing one linear functional. Hence, the problem is also strongly tractable with strong exponent zero.

Hence, it is enough to consider the case $\lambda_2 > 0$. Then the dimension of $S_1(F_1)$ is at least two, and the dimension of $S_d(F_d)$ is at least 2^d. The following theorem is proven in [9].

Theorem 1 *Consider the problem $\{S_d, F_d, G_d\}$ in the class of linear information with $\lambda_2 > 0$. Then*

(i) the problem is tractable iff it is strongly tractable.

(ii) the problem is strongly tractable iff

$$\lambda_1 = \|S_1\| < 1, \quad \textit{and} \quad \lambda_n = O(n^{-k})$$

for some positive k. For the strong exponent p we have

$$p \leq \max\left\{k^{-1}, s\right\}, \tag{3}$$

where s is given by the equation $\sum_{i=1}^{\infty} \lambda_i^s = 1$.

For some eigenvalue sequences we have equality in (3). This holds, for example, for $\lambda_n = 1/(a\,n + b)^r$ with positive a and r, and $a + b > 1$. In this case $p = s = \kappa/r$, where κ is given by $\sum_{n=1}^{\infty}(a\,n + b)^{-\kappa} = 1$. For fixed a and b, the strong exponent p goes to infinity as r goes to zero, and it goes to zero as r goes to infinity.

Hence, we have strong tractability if the sequence of eigenvalues of W_1 goes to zero like a polynomial in n^{-1}, and the norm of the operator S_1 (or W_1) is strictly less than one. Note that the norm of S_d is λ_1^d which is exponentially small in d for strongly tractable problems. It might seem more natural to scale the problem by taking $\lambda_1 = \|S_1\| = 1$, but we would then lose even tractability. Scaling of linear multivariate problems and their tractability are interrelated with some surprising consequences, see [8].

4 The Class of Fourier Information

We believe that the class of Fourier information has not yet been studied in the literature, and that the analysis presented in this section is new.

Since Λ_d^{Fou} is a subset of F_d^*, all the negative results for the class of linear information are also true for the class of Fourier information. Hence, without loss of generality we assume that S_1 is compact since otherwise the problem is intractable.

Let $d = 1$. For the class of Fourier information we have available to us the inner products $\langle f, \eta_i \rangle_{F_1}$ for $i \in I$. For the operator S_1 and fixed n, suppose we compute

$$N(f) = [\langle f, \eta_{i_1} \rangle_{F_1}, \langle f, \eta_{i_2} \rangle_{F_1}, \ldots, \langle f, \eta_{i_n} \rangle_{F_1}]$$

for some indices $i_1, i_2, \ldots, i_n \in I$.

Let $r(N)$ denote the minimal error of the approximations U_1 having the form (2) that use the information N. It is known, see e.g., Chapter 4 of [6], that

$$r(N) = \sup_{f \in F_1, N(f)=0, \|f\| \le 1} \|S_1 f\| = \|S_1|_{X_{\vec{i}}}\|, \tag{4}$$

where $X_{\vec{i}} = \operatorname{span}^{\perp}(\eta_{i_1}, \eta_{i_2}, \ldots, \eta_{i_n})$.

The optimal choice of information N corresponds to choosing the vector $\vec{i}$ for which the norm of S_1 over $X_{\vec{i}}$ is minimal. Hence,

$$\gamma_n := e(n, \Lambda_1^{\mathrm{Fou}}) = \min_{\vec{i}=[i_1,i_2,\ldots,i_n],\ i_j \in I} \|S_1|_{X_{\vec{i}}}\|. \tag{5}$$

Clearly, the complexity $\operatorname{comp}(\varepsilon, \Lambda_1^{\mathrm{Fou}})$ is finite for all positive ε iff the sequence γ_n tends to zero as n goes to infinity. Unfortunately, the compactness of S_1 does not necessarily imply this. Let X be the closed linear hull of the η_i; we write $X = \operatorname{span}\{\eta_i \mid i \in I\}$. If X is a proper subset of F_1 then the sequence γ_n need not converge to zero. Assume then that $X = F_1$, i.e., the η_i form an orthogonal basis of F_1. Then compactness of S_1 implies convergence to zero of γ_n.

This discussion illustrates the difference between the classes of linear and Fourier information for $d = 1$. For F_1^*, the complexity is finite for all ε iff S_1 is compact. For Λ_1^{Fou} we need to assume also that the η_i's form an orthogonal basis. Then the complexity is finite for all ε iff S_1 is compact.

We now discuss the multivariate case $d \ge 2$. We approximate the linear operator S_d by the class of Fourier information consisting of inner products $\langle \cdot, \eta_{i_1}\eta_{i_2}\cdots\eta_{i_d}\rangle_{F_d}$, where $i_j \in I$.

Assume first that $\gamma_0 = 0$ in (5). Then S_1 as well as all S_d are zero and the problem is trivially strongly tractable.

Assume thus that $\gamma_0 > 0$ and that $\gamma_1 = 0$. This means that S_1 is of rank 1, i.e., of the form $S_1 f = \langle f, \eta_{i^*}\rangle_{F_1} g$ for some i^*, with $g = S_1\eta_{i^*} \in G_1$ and $\gamma_0 = \|g\|_{G_1}$. Then $S_d f = \left\langle f, \eta_{i^*}^d \right\rangle_{F_d} g^d$ and it can be computed in one evaluation. Once more, the problem is strongly tractable.

We now consider the case $\gamma_1 > 0$. We begin by discussing nonzero linear functionals, $S_1 f = \langle f, h\rangle_{F_1}$. Then $S_d f = \left\langle f, h^d \right\rangle_{F_d}$ is also a linear functional. This problem is trivial for the class of linear information since it can be solved exactly in one evaluation. For the class of Fourier information the situation may be quite different.

Assume that $h \in X$ since otherwise the problem cannot be solved. Let

$$h = \sum_{i=1}^{\infty} a_i \eta_i, \quad \text{with } \|h\|^2 = \sum_{i=1}^{\infty} a_i^2,$$

where, with a possible permutation of η_i, we can assume that

$$|a_1| \;\geq\; |a_2| \;\geq\; \cdots \;\geq\; 0.$$

Consider the information $N(f) = [\langle f, \eta_{i_1}\rangle_{F_1}, \langle f, \eta_{i_2}\rangle_{F_1}, \dots, \langle f, \eta_{i_n}\rangle_{F_1}]$ for some indices i_j. It is easy to check that the approximation

$$U_1(f) \;=\; \sum_{j=1}^{n} \left\langle f, \eta_{i_j} \right\rangle_{F_1} \, a_{i_j}$$

minimizes the error among all approximations that use the information N, and the minimal error $r(N)$, see (4), is given by

$$e(U_1) \;=\; r(N) \;=\; \sqrt{\|h\|^2 \;-\; \sum_{j=1}^{n} a_{i_j}^2}.$$

This shows that the optimal choice of η_{i_j} corresponds to the largest weights a_j, i.,e., $\eta_{i_j} = \eta_j$, and

$$\gamma_n \;=\; \sqrt{\sum_{i=n+1}^{\infty} a_i^2} \;=\; \sqrt{\|h\|^2 \;-\; \sum_{i=1}^{n} a_i^2}.$$

Hence, $\gamma_1 > 0$ implies that $|a_2| > 0$, or equivalently, that $\|h\| > |a_1|$.

For $d \geq 2$, we have

$$h^d \;=\; \sum_{i_1, i_2, \dots, i_d = 1}^{\infty} a_{i_1} a_{i_2} \cdots a_{i_d} \, \eta_{i_1} \eta_{i_2} \cdots \eta_{i_d}.$$

We order the coefficients $a_{i_1}\, a_{i_2} \,\cdots\, a_{i_d}$ in decreasing order, i.e., let $\{\beta_{i,d}\}$ be a rearrangement of the products $\{a_{i_1}\, a_{i_2} \,\cdots\, a_{i_d}\}$ such that

$$|\beta_{1,d}| \;\geq\; |\beta_{2,d}| \;\geq\; \cdots \;\geq\; 0.$$

Clearly, $\beta_{1,d} = a_1^d$, and $\sum_{i=1}^{\infty} \beta_{i,d}^2 = \|h\|^{2d}$.

It is easy to check that the minimal error is now given by

$$e(n, \Lambda_d^{\mathrm{Fou}}) \;=\; \sqrt{\|h\|^{2d} - \sum_{i=1}^{n} \beta_{i,d}^2}$$

and the best approximation U_d that has error $e(n,\Lambda_d^{\rm Fou})$ is of the form

$$U_d(f) \;=\; \sum_{i=1}^n \langle f, \eta_{i_1}\eta_{i_2}\cdots\eta_{i_d}\rangle_{F_d}\;\beta_{i,d}.$$

Hence, $e(n,\Lambda_d^{\rm Fou})$ tends to zero as n goes to infinity. We now check that a necessary condition for tractability (and strong tractability) is $\|h\|\,<\,1$. Indeed, to illustrate the necessity of this condition, assume that $\|h\|\ge 1$. Since $|\beta_{i,d}|\le|\beta_{1,d}|=|a_1|^d$, we have

$$e(n,\Lambda_d^{\rm Fou})^2 \;\ge\; \|h\|^{2d}\left(1\;-\;n\,\left(\frac{|a_1|}{\|h\|}\right)^{2d}\right).$$

If we want to guarantee that the error is at most ε, with $\varepsilon<1$, then n must satisfy

$$n\;\ge\;\left(1\;-\;\frac{\varepsilon}{\|h\|^{2d}}\right)\;\left(\frac{\|h\|}{|a_1|}\right)^{2d}.$$

Since $\|h\|/|a_1|>1$, the number n of computed functionals is bounded below by an exponential function of d, and therefore the problem is intractable.

We stress that even for $d=1$, the speed of convergence $e(n,\Lambda_1^{\rm Fou})=\gamma_n$ can be arbitrarily slow for some h, and equivalently, the complexity even for $d=1$ can go to infinity arbitrarily quickly as ε approaches zero. Indeed, let $g:[0,\infty)\to\mathbb{R}_+$ be a convex decreasing function such that $g(0)=1$ and $\lim_{x\to\infty}g(x)=0$. Define

$$a_i\;=\left(g(i-1)-g(i)\right)^{1/2}\quad\text{for } i\;=\;1,2,\dots \tag{6}$$

and, as before, $h=\sum_{i=1}^\infty a_i\eta_i$. Note that monotonicity and convexity of g yield that a_i are positive and $a_i\ge a_{i+1}$. We have $\|h\|=1$, and

$$\gamma_n^2=e(n,\Lambda_1^{\rm Fou})^2\;=\;g(n),$$

and therefore

$${\rm comp}(\varepsilon,\Lambda_1^{\rm Fou})\;=\;\min\{n:\;g(n)\;\le\;\varepsilon^2\,\}.$$

For example, take an integer k and define the function $g(x)=1/\ln(k,x)$, where $\ln(k,x)=\ln\,\ln\,\cdots\,\ln(x+c_k)$, (with ln occurring k times), $\forall\,x\ge 0$, with $c_k=\exp(\exp(\,\cdots\,(\exp(1)\,\cdots)$. The number c_k is chosen in such a way that $\ln(k,\cdot)$ is well defined and $\ln(k,0)=1$. For such g, we have

$${\rm comp}(\varepsilon,\Lambda_1^{\rm Fou})\;=\;\left\lceil\exp(\exp(\cdots\exp(\varepsilon^{-2})\,\cdots)\;-\;c_k\right\rceil.$$

We are ready to present necessary and sufficient conditions for tractability and strong tractability of general operators S_d in the class of Fourier information.

Theorem 2 *Consider the problem* $\{S_d, F_d, G_d\}$ *in the class of Fourier information with* $F_1 = \mathrm{span}\{\eta_i \mid i \in I\}$ *and* $\gamma_1 > 0$*. Then*

(i) the problem is tractable iff it is strongly tractable.

(ii) the problem is strongly tractable iff

$$\gamma_0 = \|S_1\| < 1, \quad \textit{and} \quad \gamma_n = O(n^{-k})$$

for some positive k*.*

Proof: Assume first that the problem is tractable, $\mathrm{comp}(\varepsilon, \Lambda_d^{\mathrm{Fou}}) \le C\,\varepsilon^{-p}\,d^q$. Tractability in the class of Fourier information implies tractability in the class of linear information.

If S_1 is a linear functional we proved in Section 3 that $\lambda_1 = \gamma_0 = \|S_1\| < 1$. If S_1 is not a linear functional ($\lambda_2 > 0$ in the notation of Section 3) then tractability in the class of linear information implies that $\lambda_1 = \gamma_0 < 1$.

For $d = 1$, we have, because the problem is tractable,

$$\mathrm{comp}(\varepsilon, \Lambda_1^{\mathrm{Fou}}) = \min\{n : \gamma_n \le \varepsilon\} \le C\,\varepsilon^{-p}.$$

This implies that $\gamma_n = O(n^{-k})$ with $k = 1/p$.

Consider now the Smolyak algorithm, see [5], for approximation of S_d as analyzed in [7]. The Smolyak algorithm is linear and uses as its information the tensor product of linear functionals used in the one-dimensional case. Thus, the Smolyak algorithm uses $\langle \cdot, \eta_{i_1}\eta_{i_2}\cdots\eta_{i_d}\rangle_{F_d}$ for some indices i_j. This information is allowed in the class of Fourier information.

As proven in [7], $\gamma_0 < 1$ and $\gamma_n = O(n^{-k})$ implies that the cost of the Smolyak algorithm with error at most ε is bounded by $C\,\varepsilon^{-m}$ for some C and m both independent of d. Hence, the problem is strongly tractable.

Both parts of Theorem 2 easily follow from the above reasoning. $\square$

Theorem 2 specifies conditions on strong tractability for the class of Fourier information. It does not, however, specify the strong exponent. An upper bound on the strong exponent can be found in Theorem 2 of [7]. In general, this bound is not sharp. The problem of finding the strong exponent for the class of Fourier information is open.

We stress that conditions on tractability in both classes of linear and Fourier information are similar. Excluding trivial cases ($\lambda_2 = 0$ and $\gamma_1 = 0$), tractability is equivalent to strong tractability. Strong tractability holds under the same conditions on the sequence of λ_n or γ_n, respectively.

5 The Class of Standard Information

The class of standard information is probably the most practical one and has been studied in many papers for many specific problems. As we shall see, tractability and strong tractability in this class depend on the dimension of $S_1(F_1)$. If the latter is at least two, i.e., S_1 is not a linear functional, there is a simple criterion for tractability. In particular, as with the linear and Fourier classes of information, tractability is equivalent to strong tractability. If, however, S_1 is a linear functional then the situation is much more complex. We shall show by constructing examples that in this case tractability is not, in general, equivalent to strong tractability. Furthermore, the structure of the Hilbert space F_1 plays a much more decisive role than in the previous cases.

5.1 Linear Operators

First of all observe that even for a nonseparable space F_1 we have $\lim_{n\to+\infty} e(n,\Lambda_d^{\rm std}) = 0$ iff S_1 is compact. We already know from the discussion in Section 2 that compactness of S_1 is a necessary condition for $e(n,\Lambda_d^{\rm std})$ to converge to zero. It is enough to check the sufficiency for linear functionals[6], $S_d f = \left\langle f, h^d \right\rangle_{F_d}$. The element h^d is from F_d, which has the reproducing kernel K_d. Since F_d is the completion of linear combinations of elements $K_d(\cdot,t_i)$, see [1], we know that for any positive ε there exists a finite $n = n(\varepsilon)$ and there exist $t_1, t_2, \ldots, t_n \in D^d$ and $a_1, a_2, \ldots, a_n \in \mathbb{R}$ such that

$$\left\| h^d - \sum_{i=1}^n a_i\, K_d(\cdot, t_i) \right\|_{F_d} \le \varepsilon.$$

Define the approximation $U_d(f) = \sum_{i=1}^n a_i\, f(t_i)$ which uses only function values, i.e., information from the class $\Lambda_d^{\rm std}$. Then $f(t_i) = \langle f, K_d(\cdot,t_i)\rangle_{F_d}$ yields

$$|S_d f - U_d(f)| \;=\; \left| \left\langle f, h^d - \sum_{i=1}^n a_i\, K_d(\cdot,t_i) \right\rangle_{F_d} \right| \;\le\; \|f\|_{F_d}\, \varepsilon.$$

This means that the error of U_d is at most ε. Therefore, $e(m,\Lambda_d^{\rm std}) \le \varepsilon$ for all $m \ge n$. Since ε is arbitrary we have $\lim_{n\to+\infty} e(n,\Lambda_d^{\rm std}) = 0$, as claimed.

This also implies that $\mathrm{comp}(\varepsilon,\Lambda_d^{\rm std})$ is finite for all positive ε, although the proof presented above does not supply any specific bounds on $\mathrm{comp}(\varepsilon,\Lambda_d^{\rm std})$. In general, $\mathrm{comp}(\varepsilon,\Lambda_d^{\rm std})$ may

[6]This follows from the fact that a compact operator can be approximated with an arbitrarily small error by finite rank operators. An operator of rank k is determined by k linear functionals. Hence, if we can approximate linear functionals then we can approximate finite rank operators, and compact operators.

go arbitrarily quickly to infinity as ε approaches zero even for $d = 1$ and a linear functional S_1, see Section 4.2.

To control the behavior of $\mathrm{comp}(\varepsilon, \Lambda_d^{\mathrm{std}})$, assume that for $d = 1$ we have a polynomial dependence on $1/\varepsilon$, $\mathrm{comp}(\varepsilon, \Lambda_1^{\mathrm{std}}) = O(\varepsilon^{-p})$. For $d \geq 2$, we may use the Smolyak algorithm, see [7], which yields

$$\mathrm{comp}(\varepsilon, \Lambda_d^{\mathrm{std}}) \;\leq\; \alpha_1 \left(\alpha_2 \,+\, \frac{\ln\, \varepsilon^{-1}}{d-1} \right)^{d-1} \varepsilon^{-p}$$

with α_i independent of d and fully determined by the one-dimensional S_1. The above estimate can be rewritten as follows. For any positive η there exists a constant C_η (which could be larger than one) such that

$$\mathrm{comp}(\varepsilon, \Lambda_d^{\mathrm{std}}) \;\leq\; C_\eta^d \, \varepsilon^{-p-\eta}.$$

The essence of the last estimate is that we have at most exponential dependence on d, and that the dependence on ε is roughly the same as for the one-dimensional case. For small d, this estimate is always fine.

The last estimate does not answer the question of when the problem is tractable or strongly tractable. We now address this issue. Let

$$\sigma_n \;:=\; e(n, \Lambda_1^{\mathrm{std}}) \;=\; \inf_{t_1, t_2, \dots, t_n \in D} \; \sup_{f \in F_1,\, f(t_1) = f(t_2) = \cdots = f(t_n) = 0} \|S_1 f\|_{G_1}$$

be the nth minimal error for the one-dimensional case, $d = 1$. For standard information, $f(t_i) = \langle f, K_1(\cdot, t_i) \rangle_{F_1} = 0$ means that f is orthogonal to $K_1(\cdot, t_i)$, $i = 1, 2, \dots, n$. Clearly,

$$\sigma_0 \;=\; \gamma_0 \;=\; \lambda_1 \;=\; \|S_1\|.$$

For at least two-dimensional operators, $\dim(S_1(F_1)) \geq 2$, conditions on tractability and strong tractability are known, see [7]. We now recall them.

Theorem 3 *Consider the problem $\{S_d, F_d, G_d\}$ with $\dim(S_1(F_1)) \geq 2$ in the class of standard information. Then*

(i) the problem is tractable iff it is strongly tractable.

(ii) the problem is strongly tractable iff

$$\sigma_0 = \|S_1\| \;<\; 1, \quad \textit{and} \quad \sigma_n \;=\; O(n^{-k})$$

for some positive k.

Thus when $\dim(S_1(F_1)) \geq 2$ the situation is essentially the same as for the other classes of information. The case of linear functionals, $\dim(S_1(F_1)) = 1$, is much more complicated and treated in the next subsection.

5.2 Linear Functionals

In this section we assume that S_1 is a linear functional, $S_1 f = \langle f, h \rangle_{F_1}$. Then $S_d f = \left\langle f, h^d \right\rangle_{F_d}$ is also a linear functional and $G_d = \mathbb{R}$.

Tractability and strong tractability depend, in particular, on the sequence $\sigma_n = e(n, \Lambda_1^{\mathrm{std}})$ for the univariate case $d = 1$. In the next subsection we discuss the behavior of the sequence $\{\sigma_n\}$, and then we switch to the multivariate case with $d \geq 2$.

5.2.1 Univariate Case, $d = 1$

We have

$$\sigma_n \;=\; \inf_{a_i \in \mathbb{R},\, t_i \in D} \left\| h \;-\; \sum_{i=1}^n a_i\, K_1(\cdot, t_i) \right\|_{F_1}.$$

Indeed, this easily follows from the fact that for the approximation $U_1(f) = \sum_{i=1}^n a_i f(t_i)$ we have

$$|S_1 f - U_1(f)| \;=\; \left| \left\langle f, h - \sum_{i=1}^n a_i\, K_1(\cdot, t_i) \right\rangle_{F_1} \right| \;\leq\; \|f\|_{F_1} \left\| h - \sum_{i=1}^n a_i\, K_1(\cdot, t_i) \right\|_{F_1}$$

so that the error of U_1 is given by

$$e(U_1) \;=\; \|h - \sum_{i=1}^n a_i K(\cdot, t_i)\|_{F_1}.$$

Clearly

$$\sigma_0 \;=\; \|h\|_{F_1}.$$

We now consider the minimal error σ_1. It is obvious that for all S_1 for which $h = aK_1(\cdot, t)$ for some $t \in D$, we have $\sigma_1 = 0$. As we shall see, for some spaces F_1 of arbitrary dimension it may happen that $\sigma_1 = 0$ for *all* S_1. To show such an example we first derive the formula for σ_1 which will also be needed for further estimates.

Theorem 4 *(i) For any space F_1 we have*

$$\sigma_1 \;=\; \sqrt{\sigma_0^2 \;-\; \sup_{t \in D} \frac{h^2(t)}{K_1(t, t)}}. \tag{7}$$

Moreover, if $\sigma_0 > 0$ then $\sigma_1 < \sigma_0$.

(ii) For any positive integer k or for $k = +\infty$, there exists a Hilbert space F_1 of dimension k for which $\sigma_1 = 0$ for all linear functionals S_1.

Proof: We first show the formula for σ_1. We have, for arbitrary $t \in D$,

$$\|h \,-\, a\,K_1(\cdot,t)\|_{F_1}^2 \;=\; \sigma_0^2 \,-\, 2a\,h(t) \,+\, a^2 K_1(t,t).$$

Minimizing with respect to a we get $a = h(t)/K_1(t,t)$, and so

$$\inf_a \sup_{f\in F_1,\ \|f\|_{F_1}\le 1} \left|S_1 f \,-\, a\,f(t)\right|^2 \;=\; \sigma_0^2 \,-\, \frac{h^2(t)}{K_1(t,t)},$$

with the convention that $0/0 \;=\; 0$.[7] Minimizing with respect to t, we get

$$\sigma_1^2 \;=\; \sigma_0^2 \,-\, \sup_{t\in D} \frac{h^2(t)}{K_1(t,t)}$$

which yields (7).

Observe that σ_1 cannot be equal to σ_0 for positive σ_0. Indeed, $\sigma_1 \;=\; \sigma_0$ implies that $h(t) = 0$, $\forall\, t \in D$. Hence, $h = 0$ which contradicts $\sigma_0 = \|h\|_{F_1} > 0$.

We now turn to (ii). The dimension of F_1 is to be k, hence we are looking for $F_1 = \mathrm{span}(e_1, e_2, \ldots, e_k)$ for some functions $e_i : D \to \mathbb{R}$. We set $D = [-1,1]$. Let

$$\vec{e}\,(t) \;=\; [e_1(t), e_2(t), \ldots, e_k(t)], \quad \forall\, t \in [-1,1].$$

We choose the functions e_i such that $\vec{e}\,([-1,1])$ is dense in $[-1,1]^k$. If $k = +\infty$ we use the l_2 norm, and we additionally assume that

$$\sum_{i=1}^{\infty} e_i^2(t) \;<\; +\infty, \quad \forall\, t \in [-1,+1].$$

Clearly such functions exist since we do not impose any regularity assumptions on e_i. We may define the function $\vec{e}$ as follows. Let r_i be an ordered sequence of all rationals from $[-1,1]$, and let $\vec{p}_{i,k}$ be an ordered sequence of all rational vectors from $[-1,+1]^k$. For $k = +\infty$, we use the diagonal ordering of successive components such that each $\vec{p}_{i,\infty}$ has finitely many nonzero components. Define $\vec{e}(r_i) \;=\; \vec{p}_{i,k}$ and $\vec{e}(t) \;=\; 0$, say, otherwise. For $k = +\infty$, we see that $\sum_{i=1}^{\infty} e_i^2(t)$ equals zero for irrational t, and equals $\|\vec{p}_{\,j,\infty}\|_2 < +\infty$ for a rational $t = r_j$.

[7] Observe that $K_1(t,t) = 0$ implies $K_1(t,t) = \|K_1(t,\cdot)\|_{F_1}^2 = 0$, so in turn $K_1(t,\cdot) = 0$, and $f(t) = 0$ for all $f \in F_1$. Hence, $K_1(t,t) = 0$ yields that $h(t) \;=\; 0$ and that the error is σ_0. This is consistent with our convention $0/0 = 0$.

It is easy to check that these functions e_i are linearly independent. So we define $F_1 = \mathrm{span}(e_1, e_2, \ldots, e_k)$, with inner product such that the functions e_i are orthonormal. The reproducing kernel K_1 is then given by

$$K_1(x,t) \;=\; \sum_{i=1}^{k} e_i(x)\, e_i(t), \quad \forall\, x,t \in [-1,+1].$$

Indeed, $K_1(\cdot,t)$ belongs to F_1 since $\sum_{i=1}^k e_i^2(t) < \infty$, and $\langle f, K_1(\cdot,t)\rangle_{F_1} = f(t)$. We now show that $\sigma_1 = 0$ for an arbitrary linear functional $S_1 f = \langle f, h\rangle_{F_1}$ with $h = \sum_{i=1}^k \alpha_i e_i \in F_1$. Indeed, we have $\sigma_0^2 = \|h\|^2 = \sum_{i=1}^k \alpha_i^2 < +\infty$. If $\sigma_0 = 0$ then $\sigma_1 = 0$. Now assume $\sigma_0 > 0$. Then we have

$$\sigma_0^{-1}[\alpha_1, \alpha_2, \ldots, \alpha_k] \;\in\; [-1,+1]^k \backslash \{\vec{0}\}.$$

Let $\vec{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_k]$. Since $\vec{e}\,([-1,+1])$ is dense in $[-1,+1]^k$, for any positive $\eta < 1$ there exists $t \in [-1,+1]$ such that

$$\|\, \vec{e}\,(t) \;-\; \sigma_0^{-1}\vec{\alpha}\,\|_2 \;\le\; \eta.$$

This means that for small η we get $\|\vec{e}\,(t)\|_2 > 0$, and the vectors $\vec{e}\,(t)$ and $\vec{\alpha}$ are almost parallel. We have

$$\frac{h^2(t)}{K_1(t,t)} \;=\; \frac{(\sum_{i=1}^k \alpha_i e_i(t))^2}{\sum_{i=1}^k e_i^2(t)}.$$

Observe that

$$\sum_{i=1}^k \alpha_i e_i(t) \;=\; \sigma_0 \left(\sum_{i=1}^k e_i(t)^2 \;+\; \sum_{i=1}^k \left(\sigma_0^{-1}\alpha_i - e_i(t)\right)\, e_i(t) \right).$$

Therefore

$$\left| \sum_{i=1}^k \alpha_i e_i(t) \right| \;\ge\; \sigma_0 \sum_{i=1}^k e_i(t)^2 \left(1 \;-\; \|\vec{e}\,(t) \;-\; \sigma_0^{-1}\vec{\alpha}\,\|_2 / \|\vec{e}\,(t)\,\|_2\right),$$

and

$$\frac{h^2(t)}{K_1(t,t)} \;\ge\; \sigma_0^2\, \|\vec{e}\,(t)\|_2^2\, \left(1 \;-\; \eta/\|\vec{e}\,(t)\|_2\right)^2.$$

Letting η go to zero, we get $\|\vec{e}\,(t)\|_2 \to 1$ and $\sup_{t\in D} h^2(t)/K_1(t,t) = \sigma_0^2$. Hence, $\sigma_1 = 0$ due to (7). This completes the proof. $\square$

Remark 1 The space F_1 in the proof of Theorem 4 (ii) consists of very irregular functions. We now show that F_1 can be chosen as a subclass of the class $C([0,1])$ of continuous functions. The construction of such F_1 is as follows, see also [2].

The interval $[0,1]$ is a Peano set, i.e., there exists a continuous mapping

$$g \;=\; [g_1, g_2, \ldots] \;:\; [0,1] \to [-1,1]^{\mathbb{N}}$$

which is onto, see, e.g., [3]. Such a mapping is called a Peano map or a Peano curve. Here, g_i is the ith component of g and is a continuous function.

For a given integer k or $k = +\infty$, define

$$F_1 \;=\; \left\{ f : [0,1] \to \mathbb{R} \,|\, f = \sum_{i=1}^{k} f_i g_i \;\text{ for which }\; \sum_{i=1}^{k} i^2 f_i^2 < +\infty \right\}$$

with the inner product

$$\langle f, h \rangle_{F_1} \;=\; \sum_{i=1}^{k} i^2 f_i h_i$$

for $h = \sum_{i=1}^{k} h_i g_i \in F_1$.

Observe that $f(t) = \sum_{i=1}^{k} f_i g_i(t)$ is well defined since $|g_i(t)| \le 1$ and

$$|f(t)| \;\le\; \sum_{i=1}^{k} |f_i| \;\le\; \left(\sum_{i=1}^{k} i^2 f_i^2 \right)^{1/2} \left(\sum_{i=1}^{k} i^{-2} \right)^{1/2} \;\le\; \|f\|_{F_1} \, \pi/\sqrt{6}.$$

This also implies that f is a continuous function; hence $F_1 \subset C([0,1])$. It is easy to check that F_1 is complete so F_1 is a Hilbert space.

We now show that for any linear functional $S_1 \in F_1^*$ and any positive ε, there exist a nonnegative number β and $x \in [0,1]$ such that

$$|S_1 f \;-\; \beta\, f(x)| \;\le\; \varepsilon, \qquad \forall\, f \in F_1, \; \|f\|_{F_1} \le 1. \tag{8}$$

That is, S_1 can be recovered with arbitrarily small error by using at most one function value; hence $\sigma_1 = 0$.

Indeed, $S_1 f = \langle f, h \rangle_{F_1}$ for some $h = \sum_{i=1}^{k} h_i g_i \in F_1$. The series $\sum_{i=1}^{k} i^2 h_i^2$ is convergent, so there exists $m = m(\varepsilon)$ such that

$$\sum_{i=m+1}^{k} i^2 h_i^2 \;\le\; \varepsilon^2.$$

Let

$$\beta \;=\; \max_{i=1,2,\ldots m} |S_1 g_i|.$$

Observe that $S_1 g_i = i^2 h_i$ and since $|ih_i| \le \|h\|_{F_1}$ then $\beta \le m\|h\|_{F_1}$.

If $\beta > 0$ then

$$u \;=\; \beta^{-1}\,[S_1 g_1, S_1 g_2, \ldots, S_1 g_m] \;\in [-1,1]^m.$$

Since g is surjective, there exists $x \in [0,1]$ such that $g(x) = [u, 0, 0, \ldots]$. That is, $g_i(x) = \beta^{-1} S_1 g_i$ for $i = 1, 2, \ldots m$, and $g_i(x) = 0$ for $i > m$.

For $\|f\|_{F_1} \le 1$ we thus have

$$\begin{aligned} S_1 f &= \sum_{i=1}^{m} f_i S_1 g_i \;+\; \sum_{i=m+1}^{k} f_i S_1 g_i \\ &= \beta \sum_{i=1}^{m} f_i g_i(x) \;+\; \sum_{i=m+1}^{k} i^2 f_i h_i \\ &= \beta \sum_{i=1}^{k} f_i g_i(x) \;+\; \sum_{i=m+1}^{k} i^2 f_i h_i \\ &= \beta\, f(x) \;+\; \sum_{i=m+1}^{k} i^2 f_i h_i. \end{aligned}$$

Hence,

$$|S_1 f - \beta f(x)| \;\le\; \sum_{i=m+1}^{k} i^2 |f_i h_i| \;\le\; \|f\|_{F_1} \left(\sum_{i=m+1}^{k} i^2 h_i^2 \right)^{1/2} \;\le\; \varepsilon,$$

as claimed in (8). Obviously, for a finite k, we can set $\varepsilon = 0$ in (8). □

Although there exist spaces for which $\sigma_1 = 0$ for all linear functionals, for typical spaces and linear functionals we have that $\sigma_n > 0$ for all n. We now recall conditions under which σ_n goes to zero at least as quickly as as $n^{-1/2}$, see [8].

If $F_1 \subset L_2(D)$ and

(i)

$$|\,\langle f, h\rangle_{F_1}\,| \;\le\; C_1 \|f\|_{L_2}, \quad \forall\, f \in F_1,$$

(ii)

$$\mathrm{supess}_{t\in D}\, K_1(t,t) \;=\; C_2 \;<\; +\infty$$

then

$$\begin{aligned}\sigma_n &\le \frac{C_1\sqrt{C_2\,\lambda(D)}}{\sqrt{n}},\\ \mathrm{comp}(\varepsilon,\Lambda_1^{\mathrm{std}}) &\le C_1^2\,C_2\,\lambda(D)\varepsilon^{-2},\end{aligned}$$

where $\lambda(D)$ denotes the Lebesgue measure of the set D.

In general, if (i) or (ii) does not hold then σ_n may go arbitrarily slowly to zero, or equivalently, $\mathrm{comp}(\varepsilon,\Lambda_1^{\mathrm{std}})$ may go arbitrarily quickly to infinity as ε approaches zero. More precisely, as in Section 3, for any convex decreasing function $g:[0,+\infty]\to\mathbb{R}_+$, there exists a linear functional S_1 for which

$$\sigma_n^2 = g(n).$$

We now provide two such examples which will also play an additional role of illustrating further estimates.

Example 1: Nonseparable Space

We present a nonseparable Hilbert space F_1 with a bounded reproducing kernel which does not satisfy the assumption (i) and satisfies the assumption (ii), and for which σ_n may go arbitrarily slowly to zero.

Define F_1 as the space of functions defined on $D=[0,1]$ with the reproducing kernel

$$K_1(t,t) = 1, \quad \text{and} \quad K_1(x,t) = 0 \text{ for } x\neq t.$$

Here, F_1 is the Hilbert space of functions f such that $f=\sum_{i=1}^\infty a_iK_1(\cdot,t_i)$ for some distinct t_i from $[0,1]$, and $\|f\|_{F_1}^2=\sum_{i=1}^\infty a_i^2<+\infty$. The inner product is $\langle f,g\rangle_{F_1}=\Sigma_{i,j}a_ib_j\delta(t_i,s_j)$. Hence we have $f(t_i)=a_i$ and $f(t)=0$ for t distinct from all t_i , so that each function f from F_1 vanishes almost everywhere. Thus, (i) does not hold, and (ii) holds with $C_2=1$.

Note that $K(\cdot,x)$ and $K(\cdot,t)$ are orthonormal for $x\neq t$. Hence, F_1 has an uncountable orthonormal system, and therefore is *not* separable.

Consider now an arbitrary linear functional $S_1f=\langle f,h\rangle_{F_1}$ with $h=\sum_{i=1}^\infty \alpha_iK_1(\cdot,t_i^*)$, where $\sum_{i=1}^\infty\alpha_i^2=1$ and $|\alpha_1|\ge|\alpha_2|\ge\cdots$. Clearly, the information

$$N(f)=[f(t_1),f(t_2),\dots,f(t_n)]$$

should consist only of sample points t_i from the set $\{t_1^*,t_2^*,\dots\}$. This corresponds to Fourier information with $\eta_i=K_1(\cdot,t_i^*)$. As in Section 3 we thus have

$$\sigma_n = e(n,\Lambda_1^{\mathrm{std}}) = e(n,\Lambda_1^{\mathrm{Fou}}) = \sqrt{1-\sum_{i=1}^n\alpha_i^2}.$$

We can define the coefficients α_i by (6) such that $\sigma_n^2 = g(n)$ for any convex decreasing function g. Hence, we can have arbitrarily slow convergence, or equivalently, arbitrarily bad complexity. $\square$

Example 2: Unbounded Kernel

We present a separable Hilbert space F_1 with an unbounded reproducing kernel which does not satisfy the assumption (ii) and satisfies the assumption (i), and for which σ_n may go arbitrarily slowly to zero. This is done by a simple modification of the space from Example 1.

Define F_1 as the space of functions $f : [0,1] \to \mathbb{R}$ which are constant over the intervals $(1/(i+1), 1/i]$ for $i = 1, 2, \ldots$. That is,

$$f(x) \;=\; \sum_{i=1}^{\infty} f(1/i)\, \chi_{(1/(i+1),1/i]}(x),$$

where $\chi_{(a,b]}$ is the characteristic (indicator) function of the set $(a, b]$.

We assume that $\sum_{i=1}^{\infty} f^2(1/i) < +\infty$, and define the inner product of F_1 as

$$\langle f, h\rangle_{F_1} \;=\; \sum_{i=1}^{\infty} f(1/i)\, h(1/i)\, i^{-1}(i+1)^{-1}.$$

Observe that

$$\int_0^1 f(x)h(x)\,dx \;=\; \sum_{i=1}^{\infty} \int_{1/(i+1)}^{1/i} f(x)h(x)\,dx \;=\; \sum_{i=1}^{\infty} f(1/i)\, h(1/i)\, i^{-1}(i+1)^{-1}.$$

Thus, $\langle f, h\rangle_{F_1} = \langle f, h\rangle_{L_2}$. This shows that $F_1 \subset L_2([0,1])$ and $\|f\|_{F_1} = \|f\|_{L_2}$. Hence, (i) holds with $C_1 = \|h\|_{F_1}$.

We now show that F_1 is a reproducing kernel Hilbert space and find the reproducing kernel K_1. For any $t \in (1/(i+1), 1/i]$ we should have

$$f(t) \;=\; f(1/i) \;=\; \langle f, K_1(\cdot, 1/i)\rangle_{F_1} \;=\; \sum_{j=1}^{\infty} f(1/j) K_1(1/j, 1/i) j^{-1}(j+1)^{-1}.$$

This is satisfied for all f if

$$K_1(1/j, 1/i) \;=\; i(i+1)\, \delta_{i,j}.$$

Since $K_1(\cdot, t)$ should be piecewise constant we finally have

$$K_1(x,t) \;=\; i(i+1) \quad \text{if } T(x) \;=\; T(t) \;= i \text{ for some } i,$$

and $K_1(x,t) = 0$ otherwise. Here, $T(x) = k$ iff $x \in (1/(k+1), 1/k]$. Since

$$K_1(1/i, 1/i) = i(i+1), \forall\, i,$$

K_1 is unbounded, and (ii) does not hold.

Let $S_1 f = \langle f, h \rangle_{F_1}$ with $h = \sum_{i=1}^{\infty} \alpha_i \chi_{(1/(i+1),1/i]}$ and $\|h\|_{F_1}^2 = \sum_{i=1}^{\infty} \alpha_i^2 i^{-1}(i+1)^{-1} = 1$. Consider the approximation $U_1(f) = \sum_{i=1}^{n} a_i f(t_i)$. Since f is piecewise constant we may assume that $t_i = 1/j_i$ for some integers j_i. Since $K(\cdot, 1/i)$ and $K(\cdot, 1/j)$ are orthogonal for distinct i and j, it is easy to check that $a_i = \alpha_{j_i} j_i^{-1}(j_i+1)^{-1}$ minimizes the error. Then the square of the error is

$$1 - \sum_{i=1}^{n} \alpha_{j_i}^2 j_i^{-1}(j_i+1)^{-1}.$$

The n best sample points correspond to the n largest numbers of the sequence $\alpha_i^2 i^{-1}(i+1)^{-1}$. Assume that

$$\frac{\alpha_{j_1}^2}{j_1(j_1+1)} \geq \frac{\alpha_{j_2}^2}{j_2(j_2+1)} \geq \cdots \geq 0.$$

Then

$$\sigma_n = \sqrt{\sum_{i=n+1}^{\infty} \alpha_{j_i}^2 j_i^{-1}(j_i+1)^{-1}}.$$

As in Section 3, we can define the coefficients α_i such that $\sigma_n^2 = g(n)$ for any convex decreasing function g. Hence, we can have arbitrarily slow convergence, or equivalently, arbitrarily bad complexity. $\square$

5.2.2 Multivariate Case, $d \geq 2$

We study multivariate linear functionals S_d. We first find the formula for $e(1, \Lambda_d^{\rm std})$ and check that $\sigma_1 = 0$ yields the trivial multivariate problems.

Lemma 1 *We have*

$$e(1, \Lambda_d^{\rm std}) = \sqrt{\sigma_0^{2d} - (\sigma_0^2 - \sigma_1^2)^d}. \tag{9}$$

Hence, $\sigma_1 = 0$ implies $e(1, \Lambda_d^{\rm std}) = 0, \forall\, d$, and $\mathrm{comp}(\varepsilon, \Lambda_d^{\rm std}) = 1$, $\forall\, \varepsilon > 0$, $d = 1, 2, \ldots$. *This means that the problem is strongly tractable with strong exponent zero.*

Proof: To prove the formula for $e(1, \Lambda_d^{\rm std})$ we proceed similarly to the case $d = 1$ in (i) of Theorem 4. That is, we approximate $S_d f = \left\langle f, h^d \right\rangle_{F_d}$ by $af(t)$ for $t = [t_1, t_2, \ldots, t_d]$, where

$t_i \in D$. We conclude that the best a is given by $a = h^d(t)/K_d(t,t)$ and

$$e^2(1,\Lambda_d^{\mathrm{std}}) \;=\; \|h^d\|_{F_d}^2 \;-\; \sup_{t\in D^d} \frac{h^{2d}(t)}{K_d(t,t)}.$$

Obviously, $\|h^d\|_{F_d} = \|h\|_{F_1}^d = \sigma_0^d$. Observe that $h^d(t)/K_d(t,t) = \prod_{i=1}^d h(t_i)/K_1(t_i,t_i)$, and therefore

$$\sup_{t\in D^d} \frac{h^{2d}(t)}{K_d(t,t)} \;=\; \left(\sup_{t\in D} \frac{h^2(t)}{K_1(t,t)}\right)^d \;=\; \left(\sigma_0^2 - \sigma_1^2\right)^d$$

using the formula (7) for σ_1. This completes the proof. □

From now on, we assume that $\sigma_1 > 0$. We study tractability issues for multivariate linear functionals S_d. It is clear (since the multivariate case cannot be easier than the univariate case) that a necessary condition for tractability is that σ_n goes to zero as a polynomial in n^{-1}. Tractability also depends on the norm of S_1. The following theorem is proven in [7].

Theorem 5 *Consider the problem $\{S_d, F_d, \mathbb{R}\}$ with $\sigma_1 > 0$ in the class of standard information.*

Assume that $\sigma_0 = \|S_1\| < 1$. Then

(i) the problem is tractable iff it is strongly tractable.

(ii) the problem is strongly tractable iff $\sigma_n = O(n^{-k})$ for some positive k.

Assume that $\sigma_0 = \|S_1\| \geq 1$. Then $\sigma_n > 0$, $\forall\, n$, implies that the problem is not strongly tractable.

Unlike the corresponding result for the previous classes of information, Theorem 5 does not cover all cases for linear functionals. In particular, Theorem 5 does not rule out the possibility that the problem is tractable for $\sigma_0 \geq 1$. As we shall see, tractability may indeed happen for some spaces F_1 and all linear functionals S_1, or for some linear functionals S_1 in a given space F_1. On the other hand, there exists a space F_1 for which tractability will never happen. Hence, the situation is much more complicated than for the other classes of information.

Even when the problem is strongly tractable, Theorem 5 does not supply bounds on the strong exponent. Some bounds on the strong exponent may be found in [7]. These bounds tend to infinity as $\|S_1\|$ tends to 1.

The unresolved case is when $\|S_1\| \geq 1$. Here we consider the normalized case $\sigma_0 = \|S_1\| = 1$. We now present several estimates for the sequence $e(n,\Lambda_d^{\mathrm{std}})$ in terms of σ_1.

Theorem 6 *Assume $\sigma_0 = 1$ and $\sigma_1 > 0$. Let*

$$\tau \,=\, 1 \,-\, \sigma_1^2 \,\in\, (0,1).$$

Then

$$\begin{aligned} e(d,\Lambda_d^{\rm std}) &\ \ge\ \sigma_1^d \,>\, 0, & (10)\\ e(n,\Lambda_{nd}^{\rm std}) &\ \ge\ \Big(1 \,-\, \tau^d\Big)^{n/2}, & (11)\\ \lim_{d\to+\infty} e(n,\Lambda_d^{\rm std}) &\ =\ 1, \quad \forall\, n, & (12)\\ \lim_{d\to+\infty} e(\lceil d^p\rceil,\Lambda_d^{\rm std}) &\ =\ 1, \quad \forall\, p\in[0,1). & (13) \end{aligned}$$

Proof: We recall, see (4), that

$$e(n,\Lambda_d^{\rm std}) \,=\, \inf_{t_i\in D^d,\, i=1,2,\dots,n}\ \sup_{f\in F_d,\, \|f\|_{F_d}\le 1,\, f(t_i)=0,\, i=1,2,\dots,n} \Big\langle f, h^d\Big\rangle_{F_d}. \tag{14}$$

In particular,

$$\sigma_1 \,=\, \inf_{t\in D}\ \sup_{f\in F_1,\, \|f\|_{F_1}\le 1,\, f(t)=0} \langle f, h\rangle_{F_1}.$$

Let $\eta\in(0,\sigma_1)$. Then for every $t\in D$ there exists $f_t\in F_1$, $\|f_t\|_{F_1}=1$, such that $f_t(t)=0$ and $\langle f_t, h\rangle_{F_1} \ge \sigma_1-\eta$.

To prove (10), take $n=d$ and arbitrary points $t_1,t_2,\dots,t_d\in D^d$. Let $t_{i,i}\in D$ denote the ith component of the point t_i. Define the function

$$f(x) \,=\, f_{t_{1,1}}(x_1) f_{t_{2,2}}(x_2)\,\cdots\, f_{t_{d,d}}(x_d), \quad \forall\, x=(x_1,\cdots,x_d)\in D^d.$$

Then $f\in F_d$, $\|f\|_{F_d}=1$, and $f(t_i)=0$ for $i=1,2,\dots,d$. Furthermore,

$$\Big\langle f, h^d\Big\rangle_{F_d} \,=\, \prod_{i=1}^d \Big\langle f_{t_{i,i}}, h\Big\rangle_{F_1} \,\ge\, (\sigma_1-\eta)^d.$$

Since this holds for arbitrary t_i, from (14) we have $e(d,\Lambda_d^{\rm std}) \ge (\sigma_1-\eta)^d$. Letting η go to zero we obtain (10).

To prove (11) we proceed similarly. This time let $\eta\in(0, e(1,\Lambda_d^{\rm std}))$. From (14) with $n=1$, for any $t\in D^d$ there exists $f_t\in F_d$, $\|f\|_{F_d}=1$ such that $f_t(t)=0$ and

$$\Big\langle f_t, h^d\Big\rangle_{F_d} \,\ge\, e(1,\Lambda_d^{\rm std})-\eta \,=\, (1-\tau^d)^{1/2}-\eta,$$

where we used Lemma 1.

Take arbitrary points $t_1, t_2, \dots, t_n \in D^{nd}$. Let $t_{i,d} \in D^d$ denote the components from $(i-1)d+1$ to id of the point t_i. For $x = [x_1, x_2, \dots, x_n] \in D^{nd}$ with $x_j \in D^d$ for $j = 1, 2 \dots, n$ define the function

$$f(x) \;=\; f_{t_{1,d}}(x_1)\, f_{t_{2,d}}(x_2) \,\cdots\, f_{t_{n,d}}(x_n), \quad \forall\, x \in D^{nd}.$$

Then $f \in F_{nd}$, $\|f\|_{F_{nd}} = 1$, $f(t_i) = 0$ for $i = 1, 2, \dots, n$ and

$$\left\langle f, h^{nd} \right\rangle_{F_{nd}} \;=\; \prod_{i=1}^{n} \left\langle f_{t_{i,d}}, h^d \right\rangle_{F_d} \;\geq\; \left((1-\tau^d)^{1/2} - \eta \right)^n .$$

Letting η go to zero we obtain (11).

The last estimates (12) and (13) follow easily from (11). Indeed, $e(n, \Lambda_d^{\rm std}) \leq e(0, \Lambda_d^{\rm std}) = 1$ and by letting d go to infinity in (11) we get (12). Finally, using $d = \Theta(\lceil d^p \rceil\, d^{1-p})$ and (11) we have for large d,

$$e(\lceil d^p \rceil, \Lambda_d^{\rm std}) \;\geq\; \left(1 - \tau^{c\, d^{1-p}} \right)^{\lceil d^p \rceil /2}$$

for some positive c. Since $1-p$ is positive and $\tau^c < 1$, the logarithm of the right hand side of the last inequality is of order $d^p (\tau^c)^{d^{1-p}}$ and goes to zero as d approaches infinity. Thus, the right hand side goes to one. This completes the proof. $\square$

Observe that the last estimate of Theorem 6 means, in particular, that the problem is *not* strongly tractable. This strengthens the second part of Theorem 5 where this is proven under the stronger assumption that all σ_n are positive. We summarize this, together with a tractability condition that also follows from the last part of Theorem 6, in the following corollary.

Corollary 1 *If $\sigma_0 \geq 1$ and $\sigma_1 > 0$ then the problem $\{S_d, F_d, \mathbb{R}\}$ in the class of standard information is not strongly tractable. If the problem is tractable then its exponent with respect to d is at least one.*

Theorem 6 says, in particular, that $e(d, \Lambda_d^{\rm std})$ is positive. We now prove that, in general, this estimate cannot be improved, in that it can fail if d is replaced by $d+1$. Furthermore, we also show that the last estimate of Theorem 6 is somewhat sharp in the sense that it can fail for arbitrary $p > 1$.

Theorem 7 *There exists a Hilbert space F_1 for which*
(i) all multivariate linear tensor product functionals are tractable;
(ii) there exist linear functionals S_1 with $\sigma_1 > 0$ and for all such problems S_d we have

$$e(d, \Lambda_d^{\mathrm{std}}) \;>\; 0 \quad \textit{and} \quad e(d+1, \Lambda_d^{\mathrm{std}}) \;=\; 0.$$

Therefore $\mathrm{comp}(\varepsilon, \Lambda_d^{\mathrm{std}}) \;\le\; d+1,\ \forall\, \varepsilon \ge 0,$ *and the exponent with respect to ε^{-1} is zero whereas the exponent with respect to d is one whenever $\sigma_1 > 0$.*

Proof: We construct a space F_1 as a two-dimensional space, $F_1 = \mathrm{span}(e_1, e_2)$, where e_1 and e_2 are two linearly independent functions defined on $D = [0, 1]$. We choose an inner product in such a way that e_i are orthonormal.

Take an arbitrary linear functional $S_1 f = \langle f, h \rangle_{F_1}$ with $h = \alpha_1 e_1 + \alpha_2 e_2$. Without loss of generality[8], assume that $\|h\|_{F_1}^2 = \alpha_1^2 + \alpha_2^2 = 1$. We first check for which α_i's we have $\sigma_1 = 0$. We compute $\sigma_1 = \sqrt{1-\tau}$ given by Theorem 6. The reproducing kernel of F_1 is given by

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

Let

$$g(t) \;=\; \frac{\alpha_1\, e_1(t) \;+\; \alpha_2\, e_2(t)}{\sqrt{e_1^2(t) \;+\; e_2^2(t)}}, \quad \forall\, t \in [0, 1], \quad (\text{here } 0/0 = 0).$$

Then from (7)

$$\sigma_1 \;=\; \sqrt{1 \;-\; \sup_{t \in [0,1]} g^2(t)}$$

and $\sigma_1 = 0$ iff $\sup_{t\in[0,1]} g^2(t) = 1$. This, in turn, holds (by application of the Cauchy-Schwarz inequality for l_2) if there exists $t \in [0, 1]$ such that

$$\alpha_1 e_2(t) \;=\; \alpha_2 e_1(t). \tag{15}$$

Let $Z_2 = \{t \in [0, 1] : e_2(t) = 0\}$ denote the roots of e_2. Consider the function

$$r := e_1/e_2 \;:\; [0, 1] \setminus Z_2 \;\to \mathbb{R}.$$

Then (15) holds for some $t \in [0, 1]$ for arbitrary α_1, α_2 iff $r([0, 1] \setminus Z_2) = \mathbb{R}$. In this case, for all linear functionals of F_1 we have $\sigma_1 = 0$.

[8] For $\|h\|_{F_1} = 1$ we construct an approximation U_d which uses $d + 1$ function values and which recovers S_d exactly. For a general h it is enough to multiply U_d by $\|h\|_{F_1}^d$.

From now on, we assume that the functions e_1 and e_2 are chosen such that $\overline{r([0,1] \setminus Z_2)}$ is a proper subset of $\mathbb{R}$. This implies that there exist linear functionals for which $\sigma_1 > 0$. They are characterized by the condition

$$\alpha_1/\alpha_2 \notin \overline{r([0,1] \setminus Z_2)}.$$

For such functionals we know from (10) that $e(d, \Lambda_d^{\rm std}) > 0$. To prove that $e(d+1, \Lambda_d^{\rm std}) = 0$ we need to assume that

$$r\left([0,1] \setminus Z_2\right) \quad \text{has infinitely many elements.} \tag{16}$$

Obviously there exist functions e_1 and e_2 satisfying all these assumptions. For instance, one can take $e_1(t) = t$ and $e_2(t) = t^2 + 1$.

For $d \geq 2$, we have $S_d f = \left\langle f, h^d \right\rangle_{F_d}$ with

$$h^d(x) \;= h^d(x_1, x_2, \dots, x_d) = \prod_{j=1}^{d} \left(\alpha_1 e_1(x_j) + \alpha_2 e_2(x_j)\right).$$

We approximate $S_d f$ by computing

$$U_d(f) \;=\; \sum_{i=1}^{d+1} a_i\, f(t_i, t_i, \dots, t_i), \quad \forall\, f \in F_d, \tag{17}$$

for some $a_i \in \mathbb{R}$ and $t_i \in [0,1]$. We stress that U_d uses the $d+1$ function values at the points whose all components are equal. The error of U_d is $e(U_d) = \|g_d\|_{F_d}$ with

$$g_d \;=\; h^d \;-\; \sum_{i=1}^{d+1} a_i\, K_d(\cdot, [t_i, t_i, \dots, t_i]).$$

That is, we have

$$g_d \;=\; \prod_{j=1}^{d} \left(\alpha_1 e_{1,j} + \alpha_2 e_{2,j}\right) \;-\; \sum_{i=1}^{d+1} a_i \prod_{j=1}^{d} \left(e_1(t_i)\, e_{1,j} \;+\; e_2(t_i)\, e_{2,j}\right),$$

where $e_{i,j}(x) = e_i(x_j)$.

Define the set

$$J_k \;=\; \left\{ \vec{j} \;=\; [j_1, j_2, \dots, j_d] \;:\; j_i \in \{1,2\}, \text{ and the number of } i \text{ with } j_i = 1 \text{ is } k \right\},$$

for $k = 0, \ldots, d$. The cardinality of the set J_k is $\begin{pmatrix} d \\ k \end{pmatrix}$. We now decompose the first term in g_d as

$$\prod_{j=1}^{d} (\alpha_1 e_{1,j} + \alpha_2 e_{2,j}) \;=\; \sum_{k=0}^{d} \alpha_1^k \, \alpha_2^{d-k} \, e_k^*,$$

where

$$e_k^* \;=\; \sum_{\vec{j} \in J_k} e_{j_1,1} e_{j_2,2} \cdots e_{j_d,d}.$$

Similarly we have

$$\prod_{j=1}^{d} (e_1(t_i)\, e_{1,j} \;+\; e_2(t_i)\, e_{2,j}) \;=\; \sum_{k=0}^{d} e_1^k(t_i)\, e_2^{d-k}(t_i)\, e_k^*.$$

Substituting these expressions into the above expression for g_d we obtain

$$g_d \;=\; \sum_{k=0}^{d} \left(\alpha_1^k \alpha_2^{d-k} \;-\; \sum_{i=1}^{d+1} a_i \, e_1^k(t_i) e_2^{d-k}(t_i) \right) \, e_k^*.$$

Hence $e(U_d) = 0$ iff $\|g_d\|_{F_d} = 0$, which in turn holds, because the $e_0^*, e_1^*, \ldots, e_d^*$ are linearly independent, iff

$$\sum_{i=1}^{d+1} a_i \, e_1^k(t_i) e_2^{d-k}(t_i) \;=\; \alpha_1^k \alpha_2^{d-k}, \quad \text{for } k = 0, 1, \ldots, d.$$

We have a system of $d+1$ linear equations and $d+1$ unknown coefficients a_i. We can find a_i for arbitrary α_i's iff the matrix

$$M \;=\; \Big(e_1^k(t_i) e_2^{d-k}(t_i) \Big) = (m_{k,i}), \quad k = 0, 1, \ldots, d, \; i = 1, 2, \ldots, d+1,$$

is nonsingular.

Take now points t_i for which $e_2(t_i)$ are nonzero and $q_i = r(t_i)$ are distinct for all $i = 1, 2, \ldots, d+1$. Due to (16) such points exist.

We claim that for these points t_i the matrix M is nonsingular. Indeed, let $W = \operatorname{diag}\Big(e_2^{-d}(t_1), e_2^{-d}(t_2), \ldots, e_2^{-d}(t_{d+1})\Big)$ be a diagonal matrix. By our assumptions it is nonsingular. Moreover, $M\,W = (a_{k,i})$ is a Vandermonde matrix with $a_{k,i} = q_i^k$. Since the q_i are distinct, the matrix $M\,W$ is nonsingular, and therefore so is M. This completes the proof. $\square$

Remark 2 We stress that the points t_i in the proof of Theorem 7, part (ii) do not depend on the functionals S_d. More precisely, in the space F_d used in the proof of Theorem 7, let

$$N_j(f) \,=\, [f(t_1,\ldots,t_1), f(t_2,\ldots,t_2),\ldots,f(t_j,\ldots,t_j)]$$

be the information, with numbers t_i for which $e_2(t_i)$ are all nonzero and $e_1(t_i)/e_2(t_i)$ are distinct for all i. Then for any linear functional S_d, we have $r(N_{d+1}) \,=\, 0$, i.e., the minimal error with $j = d+1$ is zero.

In fact, for an arbitrary functional S_d and any choice of $t_1,\cdots,t_{d+1}$, as above, we showed that there exist numbers $a_i = a_i(S_d)$, $i = 1, 2\ldots, d+1$, such that

$$S_d f \,= U_d(f) = \sum_{i=1}^{d+1} a_i\, f(t_i,\ldots,t_i), \quad \forall\, f \in F_d.$$

Remark 3 The proof of Theorem 7 presents a two-dimensional univariate space F_1 for which all linear functionals are tractable. It is possible to generalize the proof of Theorem 7 for spaces F_1 of dimension $p \geq 2$. Namely, assume that $F_1 \,=\, \mathrm{span}(e_1, e_2, \ldots, e_p)$ for orthonormal e_i defined on D. For given points $t_i \in D$ consider the $n \times n$ matrix

$$M \,=\, \Big(e_1^{k_1}(t_i)\, e_2^{k_2}(t_i) \;\cdots\; e_p^{k_p}(t_i) \Big)$$

for nonnegative k_j such that $k_1 + k_2 + \ldots + k_p = d$, and $i = 1, 2, \ldots, n \,=\, \begin{pmatrix} d+p-1 \\ p-1 \end{pmatrix}$.

We prove that if there exist points $t_1,\cdots,t_n$ such that M is nonsingular then

$$e(n, \Lambda_d^{\mathrm{std}}) \,=\, 0 \tag{18}$$

for all multivariate linear tensor product functionals. In this case, the problem is tractable and the exponent with respect to ε^{-1} is zero whereas the exponent with respect to d is at most $p-1$.

Indeed, we have

$$h^d \,=\, \prod_{j=1}^{d} \left(\alpha_1 e_{1,j} + \alpha_2 e_{2,j} + \cdots + \alpha_p e_{p,j}\right).$$

To decompose the last expression, let

$$A_{p,d} \,=\, \left\{ \vec{k} \,= [k_1, k_2, \ldots, k_p] : \text{ for nonnegative integers } k_i \text{ and } k_1 + k_2 + \cdots + k_p = d \right\}.$$

The cardinality of the set $A_{p,d}$ is $n = \begin{pmatrix} d+p-1 \\ p-1 \end{pmatrix}$. For each $\vec{k} \in A_{p,d}$ define the set

$J_{\vec{k}} \;=\; \left\{ \vec{j} = [j_1, j_2, \dots, j_d] : j_i \in \{1, 2, \dots, p\}, \text{ and the number of } j_i = m \text{ is } k_m,\ m \in [1, p] \right\}.$
Then

$$h^d \;=\; \sum_{\vec{k} \in A_{p,d}} \alpha_1^{k_1} \alpha_2^{k_2} \cdots \alpha_p^{k_p}\, e^*_{\vec{k}},$$

where

$$e^*_{\vec{k}} \;=\; \sum_{\vec{j} \in J_{\vec{k}}} e_{j_1,1} e_{j_2,2} \cdots e_{j_d,d}.$$

Consider U_d given by (17) with the number of function values $n = \begin{pmatrix} d+p-1 \\ p-1 \end{pmatrix}$. As before we can show that the error $e(U_d) = \|g_d\|_{F_d}$ with

$$g_d \;=\; \sum_{\vec{k} \in A_{p,d}} \left(\alpha_1^{k_1} \alpha_2^{k_2} \cdots \alpha_p^{k_p} \;-\; \sum_{i=1}^{n} a_i\, e_1^{k_1}(t_i) e_2^{k_2}(t_i) \cdots e_p^{k_p}(t_i) \right) e^*_{\vec{k}}.$$

To guarantee that $\|g_d\|_{F_d} = 0$ we require that a_i's satisfy the system of linear equations

$$\sum_{i=1}^{n} a_i\, e_1^{k_1}(t_i) e_2^{k_2}(t_i) \cdots e_p^{k_p}(t_i) \;=\; \alpha_1^{k_1} \alpha_2^{k_2} \cdots \alpha_p^{k_p}, \quad \forall\, \vec{k} \in A_{p,d}.$$

If the matrix M of this system is nonsingular, we can find a_i for arbitrary α_i. This completes the proof of (18).

It is natural to ask for which points t_i the matrix M is nonsingular. An example is provided for $D = [0, +\infty)$ and $e_i(t) = t^{\sqrt{q_i}}$, where q_i is the ith prime number, with $q_1 = 1$. Then

$$e_1^{k_1}(t_i)\, e_2^{k_2}(t_i) \;\cdots\; e_p^{k_p}(t_i) \;=\; t_i^{k_1 + k_2\sqrt{q_2} + \cdots + k_p\sqrt{q_p}}.$$

Clearly, the exponents $u_{\vec{k}} \;=\; k_1 + k_2\sqrt{q_2} + \cdots + k_p\sqrt{q_p}$ are different for different vectors $\vec{k} = [k_1, k_2, \dots, k_p]$.

We use induction on n to check nonsingularity of $M = \left(t_i^{u_{\vec{k}}} \right)$. The inductive hypothesis is that for $1 \le m < n$ the $m \times m$ submatrices of M that involve only $t_1, \dots, t_m$ can all be made nonsingular by appropriate choice of $t_1, \dots, t_m$. If the result holds for submatrices of size $m = \nu - 1$ then for each submatrix M_ν of size ν we find, by expansion of the determinant along the appropriate row, that

$$\det(M_\nu) \;=\; a\, t_\nu^{\beta} \;+\; o(t_\nu^{\beta}), \quad \text{as } t_\nu \to +\infty$$

for some nonzero a and β. Hence, we can take a large t_ν for which each $\det(M_\nu)$ is nonzero. From this it follows that choices of points always exist for which M is non-singular. $\square$

The preceding theorem says that in some spaces all multivariate linear tensor product functionals are tractable. We now show that the opposite can also happen.

Theorem 8 *There exists a Hilbert space for which all multivariate linear tensor product functionals with $\sigma_0 \geq 1$ and $\sigma_1 > 0$ are intractable.*

Proof: Take the Hilbert space F_1 from Example 1. That is, F_1 is a nonseparable space of functions defined on $[0,1]$ with the reproducing kernel $K_1(t,t) = 1$ and $K_1(t,x) = 0$ for $x \neq t$.

Consider now an arbitrary linear functional $S_1 f = \langle f, h \rangle_{F_1}$ with $h \in F_1$. Then $h = \sum_{i=1}^{\infty} \alpha_i K_1(\cdot, t_i)$ for some $\alpha_i \in \mathbb{R}$ and distinct $t_i \in [0,1]$, with $\sigma_0^2 = \sum_{i=1}^{\infty} \alpha_i^2 < +\infty$. We know by assumption that $\sigma_0 \geq 1$. As in Example 1, we can show that $\sigma_1 = \sqrt{\sigma_0^2 - \max_i |\alpha_i|^2}$. Hence, $\sigma_1 > 0$ iff at least two α_i are nonzero.

We showed in Example 1 that standard information for this space is equivalent to Fourier information. We thus have $\gamma_0 = \sigma_0$ and $\gamma_1 = \sigma_1$, and Theorem 8 follows from Theorem 2. $\square$

Theorems 7 and 8 state that, in general, tractability of linear functionals depends on the Hilbert space. For some spaces all nontrivial linear functionals are intractable whereas for other spaces all linear functionals are tractable with the exponents zero and at most one for ε^{-1} and d, respectively.

The spaces of Theorems 7 and 8 are very special. We believe that in "typical" Hilbert spaces some linear functionals are tractable and some others are not. The next theorem presents conditions under which we can find tractable and intractable linear functionals in a given space.

Theorem 9 *Let F_1 be a Hilbert space of real-valued functions on a domain D.*

(i) For two distinct t_1 and t_2 from D, let e_1 and e_2 be orthonormal elements from

$$\mathit{span}\,(K_1(\cdot, t_1), K_1(\cdot, t_2))\,.$$

If the function e_1/e_2 takes infinitely many values then all linear functionals $S_1 f = \langle f, h \rangle_{F_1}$ with $h \in \mathit{span}(e_1, e_2)$ are tractable with exponents zero and at most one, since $e(d{+}1, \Lambda_d^{\mathrm{std}}) = 0$ and $\mathrm{comp}(\varepsilon, \Lambda_d^{\mathrm{std}}) \leq d + 1$.

(ii) If there exist two orthonormal elements e_1 and e_2 from the space F_1 which have disjoint supports then all linear functionals $S_1 f = \langle f, h \rangle_{F_1}$ with $h = \alpha\, e_1 + \sqrt{1 - \alpha^2}\, e_2$ for $\alpha \in (0,1)$ are intractable.

Proof: To prove the first part we may use exactly the same construction as in Theorem 7, with span(e_1, e_2) now playing the role of F_1. The second part is proven as in [7] p. 53. $\square$

We believe that for "typical" spaces F_1 the assumptions of (i) and (ii) are satisfied. This holds, for example, for Sobolev spaces $F_1 = W^r([0,1])$. Hence, the classes of tractable linear functionals and intractable linear functionals are both, in general, non-empty. The trouble is that for the fixed problem $S_1 f = \langle f, h \rangle_{F_1}$ (like integration or weighted integration) we do not know whether the problem is tractable or intractable. Clearly, there remains much work to be done.

Acknowledgments. The first of us (E.N.) was supported by a Heisenberg fellowship of the German DFG. The second of us (I.H.S.) gratefully acknowledges the support of the Australian Research Council. The third of us (H. W.) was supported in part by the National Science Foundation and the Air Force Office of Scientific Research.

We thank K. Ritter, G. W. Wasilkowski and A. G. Werschulz for valuable comments.

References

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

[2] E. Novak, Quadrature and widths, *J. Approx. Th.*, 47, p. 195-202, 1986.

[3] Z. Semadeni, Banach Spaces of Continuous Functions, Vol. 1, Warsaw, 1971.

[4] I. H. Sloan and H. Woźniakowski, An intractability result for multiple integration, to appear in Math. Comp.

[5] S. A. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, *Dokl. Akad. Nauk SSSR*, p. 240-243, 1964.

[6] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, Information-based Complexity, *Academic Press*, New York, 1988.

[7] G. W. Wasilkowski, and H. Woźniakowski, Explicit Cost Bounds of Algorithms for Multivariate Tensor Product Problems, *J. of Complexity*, 11, 1-56, 1995.

[8] H. Woźniakowski, Tractability and strong tractability of linear multivariate problems, *J. Complexity*, 10, 96-128, 1994.

[9] H. Woźniakowski, Tractability and strong tractability of multivariate tensor product problems, in Proceedings of 2nd Gauss Symposium, München, 1993, *J. Computing and Information*, 4, 1-19, 1994.

Addresses:

Erich Novak, Universität Erlangen, Mathematisches Institut, Bismarckstr. 1 1/2, D-91054 Erlangen, Germany

`novak@mi.uni-erlangen.de`

Ian H. Sloan, University of New South Wales, School of Mathematics, Sydney NSW 2052, Australia

`sloan@maths.unsw.edu.au`

Henryk Woźniakowski, Columbia University, 450 Computer Science Building, New York, NY 10027, USA

`henryk@cs.columbia.edu`