JOURNAL OF COMPLEXITY **15**, 402–447 (1999)
Article ID jcom.1999.0512, available online at http://www.idealibrary.com on IDEAL®

Weighted Tensor Product Algorithms for Linear Multivariate Problems[1]

G. W. Wasilkowski

Department of Computer Science, University of Kentucky, Lexington, Kentucky 40506
E-mail: greg@cs.uky.edu

and

H. Woźniakowski

E-mail: henryk@cs.columbia.edu

Received October 5, 1998

We study the ε-approximation of linear multivariate problems defined over weighted tensor product Hilbert spaces of functions f of d variables. A class of weighted tensor product (WTP) algorithms is defined which depends on a number of parameters. Two classes of permissible information are studied. Λ^{all} consists of all linear functionals while Λ^{std} consists of evaluations of f or its derivatives. We show that these multivariate problems are sometimes tractable even with a worst-case assurance. We study problem tractability by investigating when a WTP algorithm is a polynomial-time algorithm, that is, when the minimal number of information evaluations is a polynomial in $1/\varepsilon$ and d. For Λ^{all} we construct an optimal WTP algorithm and provide a necessary and sufficient condition for tractability in terms of the sequence of weights and the sequence of singular values for $d=1$. For Λ^{std} we obtain a weaker result by constructing a WTP algorithm which is optimal only for some weight sequences. © 1999 Academic Press

[1] The authors were partially supported by the National Science Foundation under Grants CCR-9729971 and CCR-9731858, respectively.

0885-064X/99 $30.00

1. INTRODUCTION

Recently, multivariate problems of high dimension d have been successfully solved by deterministic algorithms. A typical example is multivariate integration for financial applications with $d=360$ which is solved by quasi-Monte Carlo algorithms; see, for example, [2, 5, 7, 8, 12]. The success of these algorithms presents a challenge to provide a theoretical explanation why some multivariate problems can be so efficiently solved for large d.

A possible explanation was provided in [10] by identifying weighted classes of functions for which the error of certain quasi-Monte Carlo algorithms does not depend on d. This was done only for multivariate integration and for specific weighted tensor product Hilbert spaces. Furthermore, the proofs were non-constructive.

The purpose of this paper is to study general linear multivariate problems defined over weighted tensor product Hilbert spaces and check which ones can be efficiently solved for large d, and, if so, to provide constructive algorithms for their approximation. By an efficient solution we mean that there exists a polynomial-time algorithm or equivalently that the problem is *tractable* in the worst case setting.

There are a number of papers dealing with tractability of general linear multivariate problems; see, e.g., [14, 16] and papers cited there. A linear multivariate problem is specified by a linear operator whose domain is a space of funtions of d variables. We want to compute an approximation of the linear operator to within ε in the worst case setting. Such an approximation is somputed by evaluating function values or, more generally, values of arbitrary linear functionals. *Tractability* means that the minimal number of evaluations needed to compute an ε-approximation is bounded by a polynomial in $1/\varepsilon$ and d. *Strong* tractability means that such a bound does not depend on d, i.e., it is a polynomial only in $1/\varepsilon$. The minimal degrees of such polynomials are called the d-exponent and the ε-exponent of tractability or strong tractability.

The domain of a linear operator is often assumed to be a tensor product class of functions of d variables such that each variable is equally important. For some applications, it seems more appropriate to assume that the significance of the jth variable changes with j. This can be modeled by *weighted* tensor product rather than standard tensor product spaces. For a weighted tensor product space, we have a sequence of nonnegative weights $\gamma_{d,j}$ where $\gamma_{d,j}$ moderates the behavior of functions for the jth variable. Roughly speaking, a small weight $\gamma_{d,j}$ means that functions depend only slightly on the jth variable. For the extreme case, $\gamma_{d,j}=0$, functions are independent of the jth variable. Relatively less is known for weighted tensor product problems; see, e.g., [10, 17]. This paper is a continuation of the study of tractability of linear multivariate problems defined over such weighted spaces $\mathscr{F}_d$.

The formal definition of weighted tensor product spaces is given in the next section. Here we only consider an example of such spaces from [10]. Let $\mathscr{F}_d$ be, roughly, the space of functions $f\colon [0,1]^d \to \mathbb{R}$ for which $(\prod_{j=1}^d \gamma_{d,j}^{-1/2}\, \partial/\partial x_j) f(x)$ is square-integrable. The weights $\gamma_{d,j}$ are positive and for any of their values, $\mathscr{F}_d$ is a weighted tensor product Hilbert space whose reproducing kernel is $\mathscr{K}_d(x,t) = \prod_{j=1}^d (1+\gamma_{d,j} \min\{x_j, t_j\})$. For such $\mathscr{F}_d$, the integration problem, i.e., approximating the integral $\mathscr{S}_d(f) = \int_{[0,1]^d} f(x)\,dx$, is considered in [10]. More specifically, tractability of Quasi-Monte Carlo algorithms is studied. In particular, it is proven that for weights independent of d, i.e., $\gamma_{d,j} = \gamma_j$, strong tractability holds iff $\sum_{j=1}^\infty \gamma_j < \infty$. Then the ε-exponent of strong tractability belongs to $[1,2]$. The proofs in [10] are *non-constructive*, i.e., we only know about the existence of such stronly polynomial-time Quasi-Monte Carlo algorithms.

In this paper we address similar questions for a general class of weighted tensor product spaces $\mathscr{F}_d$ of functions of d variables and for general linear tensor product operators $\mathscr{S}_d = \bigotimes_{j=1}^d \mathscr{S}$. More specifically, we assume that $\mathscr{F}_d$ is a reproducing kernel Hilbert space whose kernel is the product of one dimensional kernels, and $\mathscr{S}$ is a linear operator defined over a reproducing kernel Hilbert space of functions of one variable.

We define a class of algorithms which are called *weighted tensor product*, or shortly *WTP*, algorithms. This class depends on a number of parameters. The values of these parameters are very crucial and the efficiency of WTP algorithms depends on them. WTP algorithms are a generalization of the algorithm proposed by Smolyak [11] for unweighted tensor product problems. In Section 3 we discuss similarities and differences between WTP and Smolyak's algorithms in more detail.

We study tractability and strong tractability of $\{\mathscr{S}_d\}$ by checking when there exist WTP algorithms that are polynomial, i.e., when the number of evaluations used by certain WTP algorithms needed to compute an ε-approximation is polynomial in $1/\varepsilon$ and d. Thus, our approach is *constructive*. The efficiency of WTP algorithms depends greatly on the weight sequence $\gamma_{d,j}$ and on the class of permissible information operations.

Consider first the case of unrestricted information Λ^{all} for which we can use arbitrary linear functionals as information evaluations. In this case, the problem of approximating linear functionals $\mathscr{S}_d$ is trivial since we can now compute $\mathscr{S}_d(f)$. Obviously, $\mathscr{S}_d$ is a linear functional iff the same holds for $\mathscr{S}$. Therefore we assume that the rank $\mathscr{S}$ is at least 2. We then choose the parameters of the WTP algorithm in such a way that it is optimal in the class of all algorithms using functionals from Λ^{all}. Hence, tractability or strong tractability of $\{\mathscr{S}_d\}$ is equivalent to whether this optimal WTP algorithm is polynomial or strongly polynomial.

We provide a necessary and sufficient condition for strong tractability. It is expressed in terms of the sequence of weights $\gamma_{d,j}$ and the sequence of singular values of the operator $\mathscr{S}$. We obtain the ε-exponent of strong tractability, see Theorem 1. Roughly speaking, strong tractability holds if both sequences go to zero polynomially fast, and the faster their convergence to zero, the smaller the ε-exponent of strong tractability. In particular, if the sequence of weights goes to zero at least as fast as the sequence of singular values, then the number of evaluations for arbitrary d is of the same order as for $d=1$.

We also obtain relations between tractability and strong tractability. For general weights, these concepts are different. If, however, the weights $\gamma_{d,j}$ do not essentially depend on d, then tractability is equivalent to strong tractability.

Consider now the case of restricted information Λ^{std} for which we can only use function or derivative evaluations. Since approximating linear functionals is no longer trivial, we now allow $\mathscr{S}$ to be a linear operator of arbitrary rank. For example, if we set $\mathscr{S}(f)=\int_0^1 f(x)\,dx$ then $\mathscr{S}_d(f)=\int_{[0,1]^d} f(x_1, ..., x_d)\,dx_1 \cdots dx_d$ is a multivariate integration problem.

For Λ^{std} we choose a particular WTP algorithm. We are not sure if this choice is optimal. We prove that this WTP algorithm is polynomial if the sequence of singular values goes polynomially to zero (which is necessary for tractability of $\{\mathscr{S}_d\}$) and if the sequence of weights converges sufficiently fast to zero. Since we are not sure whether the parameters of this WTP algorithm have been chosen optimally, we believe that the assumption on the behavior of the weight sequence is too severe. We illustrate this point by the integration example from [10]. As we already indicated for $\gamma_{d,j}=\gamma_j$ the condition $\sum_{j=1}^{\infty}\gamma_j<\infty$ implies that the integration problem is strongly tractable with the ε-exponent in $[1,2]$. If we assume that $\sum_{j=1}^{\infty}\gamma_j^{1/2}<\infty$ then the WTP algorithm is polynomial and its ε-exponent is at most 2. If we additionally assume that $\sum_{j=1}^{\infty}\gamma_j^{1/3}<\infty$ then its ε-exponent is 1. Hence, for $\sum_{j=1}^{\infty}\gamma_j^{1/3}<\infty$ we have a constructive algorithm which solves the integration problem with minimal ε-exponent. However, it is open whether the WTP algorithm minimizes the ε-exponent for $\sum_{j=1}^{\infty}\gamma_j^q<\infty$ with $q>1/3$.

The paper is organized as follows. Section 2 provides basic definitions of weighted tensor product spaces, multivariate problems and their tractability and strong tractability as well as polynomial-time algorithms. Section 3 provides the definition and basic properties of WTP algorithms. The results for unrestricted class Λ^{all} of information are in Section 4, and results for restricted class Λ^{std} are in Section 5. Specific applications in Section 6 conclude the paper.

2. FORMULATION OF THE PROBLEM

In this section, we define a weighted tensor product problem for a class of functions of d variables. We also define information and algorithms as well as their costs and errors. Tractability concepts will be also defined in this section. We begin with the scalar case, $d=1$.

2.1. *Univariate Case*

Let $\mathscr{D} \subset \mathbb{R}$ be a closed (not necessarily finite) interval such that $0 \in \mathscr{D}$. Let $\mathscr{F}$ be a real separable reproducing kernel Hilbert space of real functions f defined over $\mathscr{D}$ which vanish at zero, $f(0)=0$, $\forall f \in \mathscr{F}$. The reproducing kernel of $\mathscr{F}$ will be denoted by $\mathscr{K}$. For basic properties of reproducing kernel Hilbert spaces we refer the reader to [1]. Here we remind the reader that

$$\mathscr{K}: \mathscr{D} \times \mathscr{D} \to \mathbb{R}$$

is a symmetric and nonnegative definite function. That is, $\mathscr{K}(x, y) = \mathscr{K}(y, x)$ and the matrix $(\mathscr{K}(x_i, x_j))_{i, j=1, \ldots, n}$ is nonnegative definite for all n and arbitrary choices of n points x_i from $\mathscr{D}$. The space $\mathscr{F}$ is the closure of the space of linear combinations $\sum_{i=1}^n \alpha_i \mathscr{K}(\cdot, x_i)$ and

$$\mathscr{K}(\cdot, x) \in \mathscr{F}, \forall x \in \mathscr{D}, \qquad f(x) = \langle f, \mathscr{K}(\cdot, x) \rangle_{\mathscr{F}}, \quad \forall f \in \mathscr{F}, \quad \forall x \in \mathscr{D}.$$

From the assumption that all functions from $\mathscr{F}$ vanish at zero we have

$$\mathscr{K}(\cdot, 0) \equiv 0.$$

Let $h^*: \mathscr{D} \to \mathbb{R}$ be a fixed function such that $h^*(0)=1$. Let $\gamma \in (0, 1]$. We define a weighted space $\mathscr{F}^\gamma$ as a real separable reproducing kernel Hilbert space whose reproducing kernel $\mathscr{K}_\gamma$ is given by

$$\mathscr{K}_\gamma(x, y) = h^*(x)\, h^*(y) + \gamma \mathscr{K}(x, y).$$

Such a weighted space exists since, obviously, $\mathscr{K}_\gamma$ is also a symmetric and positive definite function.

The space $\mathscr{F}^\gamma$ is the algebraic sum of $\mathscr{F}$ and $\text{span}\{h^*\}$, where the latter is the space of multiples of h^*. That is, any $f_\gamma \in \mathscr{F}^\gamma$ equals $\alpha h^* + f$ with $\alpha \in \mathbb{R}$ and $f \in \mathscr{F}$, and $f_\gamma(0) = \alpha h^*(0) = \alpha$. The inner products of $\mathscr{F}^\gamma$ and $\mathscr{F}$ are related by the formula

$$\langle g_1, g_2 \rangle_{\mathscr{F}^\gamma} = g_1(0)\, g_2(0) + \gamma^{-1} \langle f_1, f_2 \rangle_{\mathscr{F}},$$

where $g_i = g_i(0)\, h^* + f_i$ with $f_i \in \mathcal{F}$. For $f_\gamma \in \mathcal{F}^\gamma$ we have $f_\gamma - f_\gamma(0) h^* \in \mathcal{F}$ and

$$\|f_\gamma\|^2_{\mathcal{F}^\gamma} = f_\gamma^2(0) + \gamma^{-1}\, \|f_\gamma - f_\gamma(0) h^*\|^2_{\mathcal{F}}.$$

Observe that $\mathcal{K}_\gamma(\cdot, 0) = h^*$, and therefore $\langle h^*, f_\gamma \rangle_{\mathcal{F}^\gamma} = f_\gamma(0)$, $\forall f_\gamma \in \mathcal{F}^\gamma$. In particular, $\langle h^*, f \rangle_{\mathcal{F}^\gamma} = 0$, $\forall f \in \mathcal{F}$, i.e., h^* is orthogonal to $\mathcal{F}$.

The spaces $\mathcal{F}^{\gamma_1}$ and $\mathcal{F}^{\gamma_2}$ are algebraically the same, $\mathcal{F}^{\gamma_1} = \mathcal{F}^{\gamma_2}$, but they have different norms for distinct γ_1, γ_2. Their norms are equivalent. That is, for $\gamma_1 \leqslant \gamma_2$ and $f \in \mathcal{F}^{\gamma_1} = \mathcal{F}^{\gamma_2}$ we have

$$\|f\|_{\mathcal{F}^{\gamma_2}} \leqslant \|f\|_{\mathcal{F}^{\gamma_1}} \leqslant \left(\frac{\gamma_2}{\gamma_1}\right)^{1/2} \|f\|_{\mathcal{F}^{\gamma_2}}.$$

We now take $\gamma = 1$ and consider the space $\mathcal{F}^1$. By a *solution* operator we mean a continuous linear operator

$$\mathcal{S}\colon \mathcal{F}^1 \to \mathcal{G},$$

where $\mathcal{G}$ is a real Hilbert space.

Let $\mathcal{S}^*\colon \mathcal{G} \to \mathcal{F}^1$ denote the adjoint operator of $\mathcal{S}$, and let $\mathcal{W} = \mathcal{S}^*\mathcal{S}\colon \mathcal{F}^1 \to \mathcal{F}^1$. Clearly, $\|\mathcal{S}f\|_{\mathcal{G}} = \langle \mathcal{W}f, f \rangle^{1/2}_{\mathcal{F}^1}$. We assume that the solution opertor $\mathcal{S}$ is compact. This implies that $\mathcal{W}$ is also compact and there exists an orthonomal base of elements $\{\eta_i\}$ from $\mathcal{F}^1$ such that

$$\mathcal{W}\eta_i = \lambda_i \eta_i, \qquad \langle \eta_i, \eta_j \rangle_{\mathcal{F}^1} = \delta_{i,j}. \tag{1}$$

Here, $i, j = 1, 2, \ldots$. If $\dim(\mathcal{G}) < \infty$ we formally set $\lambda_i = 0$ for $i > \dim(\mathcal{G})$. The eigenvalues are ordered

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0, \qquad \text{and} \qquad \lim_{n \to \infty} \lambda_n = 0.$$

We have

$$\mathcal{W}f = \sum_{i=1}^{\infty} \lambda_i \langle f, \eta_i \rangle_{\mathcal{F}^1} \eta_i, \qquad \forall f \in \mathcal{F}^1,$$

and $\lambda_i^{1/2}$'s are singular values of $\mathcal{S}$. Clearly,

$$\|\mathcal{S}\|^2_{\mathcal{F}^1 \to \mathcal{G}} = \|\mathcal{W}\|_{\mathcal{F}^1 \to \mathcal{F}^1} = \lambda_1.$$

The solution operator $\mathscr{S}$ is also well defined for the space $\mathscr{F}^{\gamma}$ for any $\gamma \in (0, 1]$. We show that

$$\max\{\|\mathscr{S}h^*\|_{\mathscr{G}}, \gamma^{1/2}\|\mathscr{S}\|_{\mathscr{F}\to\mathscr{G}}\} \leqslant \|\mathscr{S}\|_{\mathscr{F}^{\gamma}\to\mathscr{G}} \leqslant (\|\mathscr{S}h^*\|_{\mathscr{G}}^2 + \gamma\|\mathscr{S}\|_{\mathscr{F}\to\mathscr{G}}^2)^{1/2}. \tag{2}$$

Moreover, when $\mathscr{S}$ is a functional, then the right-hand side inequality in (2) becomes an equality. Indeed, for $f \in \mathscr{F}^{\gamma}$ we have

$$\begin{aligned} \|\mathscr{S}f\|_{\mathscr{G}} &\leqslant |f(0)|\,\|\mathscr{S}h^*\|_{\mathscr{G}} + \|\mathscr{S}(f - f(0)h^*)\|_{\mathscr{G}} \\ &\leqslant |f(0)|\,\|\mathscr{S}h^*\|_{\mathscr{G}} + \gamma^{1/2}\|\mathscr{S}\|_{\mathscr{F}\to\mathscr{G}}\,\gamma^{-1/2}\|f - f(0)h^*\|_{\mathscr{F}} \\ &\leqslant (f^2(0) + \gamma^{-1}\|f - f(0)h^*\|_{\mathscr{F}}^2)^{1/2}(\|\mathscr{S}h^*\|_{\mathscr{G}}^2 + \gamma\|\mathscr{S}\|_{\mathscr{F}\to\mathscr{G}}^2)^{1/2} \\ &= \|f\|_{\mathscr{F}^{\gamma}}(\|\mathscr{S}h^*\|_{\mathscr{G}}^2 + \gamma\|\mathscr{S}\|_{\mathscr{F}\to\mathscr{G}}^2)^{1/2}, \end{aligned}$$

which proves the right hand side of (2).

To prove the left hand side, note that $\|h^*\|_{\mathscr{F}^{\gamma}} = \|h^*\|_{\mathscr{F}^1} = h^*(0) = 1$. Hence, $\|\mathscr{S}\|_{\mathscr{F}^{\gamma}\to\mathscr{G}} \geqslant \|\mathscr{S}h^*\|_{\mathscr{G}}$. Take now $f \in \mathscr{F}$ with $\|f\|_{\mathscr{F}} = 1$ and $\|\mathscr{S}f\|_{\mathscr{G}} = \|\mathscr{S}\|_{\mathscr{F}\to\mathscr{G}}$. Since $1 = \|f\|_{\mathscr{F}} = \gamma^{1/2}\|f\|_{\mathscr{F}^{\gamma}}$ then $\|\mathscr{S}f\|_{\mathscr{G}} = \|\mathscr{S}\|_{\mathscr{F}\to\mathscr{G}} = \gamma^{1/2}\|\mathscr{S}\|_{\mathscr{F}\to\mathscr{G}}\,\|f\|_{\mathscr{F}^{\gamma}}$. This shows that $\|\mathscr{S}\|_{\mathscr{F}^{\gamma}\to\mathscr{G}} \geqslant \gamma^{1/2}\|\mathscr{S}\|_{\mathscr{F}\to\mathscr{G}}$, and completes the proof of (2).

The adjoint $\mathscr{S}_{\gamma}^*\colon \mathscr{F}^{\gamma} \to \mathscr{G}$ now depends on γ since the inner product of $\mathscr{F}^{\gamma}$ does depend on γ. Let $\mathscr{W}_{\gamma} = \mathscr{S}_{\gamma}^*\mathscr{S}\colon \mathscr{F}^{\gamma} \to \mathscr{F}^{\gamma}$. It is easy to check that

$$\mathscr{W}_{\gamma}f = (1-\gamma)\langle f, \mathscr{W}h^*\rangle_{\mathscr{F}^1}\,h^* + \gamma\mathscr{W}f, \qquad \forall f \in \mathscr{F}^{\gamma}.$$

Let $\{\eta_{i,\gamma}, \lambda_{i,\gamma}\}$ be the orthonormal sequence of eigenpairs of $\mathscr{W}_{\gamma}$

$$\mathscr{W}_{\gamma}\eta_{i,\gamma} = \lambda_{i,\gamma}\eta_{i,\gamma}, \qquad \langle \eta_{i,\gamma}, \eta_{j,\gamma}\rangle_{\mathscr{F}^{\gamma}} = \delta_{i,j},$$

with ordered eigenvalues

$$\lambda_{1,\gamma} \geqslant \lambda_{2,\gamma} \geqslant \cdots \geqslant \lambda_{n,\gamma} \geqslant 0.$$

Since $\lambda_{1,\gamma} = \|\mathscr{W}_{\gamma}\|_{\mathscr{F}^{\gamma}\to\mathscr{F}^{\gamma}} = \|\mathscr{S}\|_{\mathscr{F}^{\gamma}\to\mathscr{G}}^2$, we have due to (2)

$$\|\mathscr{S}h^*\|_{\mathscr{G}}^2 \leqslant \lambda_{1,\gamma} \leqslant \|\mathscr{S}h^*\|_{\mathscr{G}}^2 + \gamma\lambda_1. \tag{3}$$

Observe that $\mathscr{W}_{\gamma}$ differs from $\gamma\mathscr{W}$ by a rank one operator $(1-\gamma)\langle f, \mathscr{W}h^*\rangle_{\mathscr{F}^1}h^*$ whose eigenvalue is $(1-\gamma)\langle h^*, \mathscr{W}h^*\rangle_{\mathscr{F}^1} = (1-\gamma)\|\mathscr{S}h^*\|_{\mathscr{G}}^2 \geqslant 0$. Weyl's monotonicity theorem, see, e.g., [6], yields

$$\gamma\lambda_i \leqslant \lambda_{i,\gamma} \leqslant \gamma\lambda_{i-1}, \qquad \forall i = 2, 3, \ldots. \tag{4}$$

We also have

$$\mathscr{W}_\gamma f = \sum_{i=1}^{\infty} \lambda_{i,\gamma} \langle f, \eta_{i,\gamma} \rangle_{\mathscr{F}^\gamma} \eta_{i,\gamma}, \qquad \forall f \in \mathscr{F}^\gamma.$$

The weighted case simplifies when we assume that $\mathscr{W} h^* = \mathscr{S}^* \mathscr{S} h^* = \beta^2 h^*$ for some nonnegative β. Then

$$\mathscr{W}_\gamma f = \beta^2 f(0)\, h^* + \gamma \mathscr{W}(f - f(0)\, h^*)$$

and $\mathscr{W}\mathscr{F} \subset \mathscr{F}$. The eigenpairs of $\mathscr{W}_\gamma$ are now $\{\eta_{i,\gamma}\} = \{\eta_i\}$ and $\{\lambda_{i,\gamma}\} = \{\beta^2, \gamma\lambda_i\}$. If $\gamma\lambda_1 \leqslant \beta^2$ then

$$\lambda_{1,\gamma} = \beta^2, \qquad \lambda_{i,\gamma} = \gamma\lambda_{i-1} \qquad \text{for} \quad i = 2, 3, \dots.$$

So far we assumed that γ is positive. It is also possible to take $\gamma = 0$. Then the space $\mathscr{F}^0 = \mathrm{span}\{h^*\}$ is a one dimensional space and $\langle f, g \rangle_{\mathscr{F}^0} = f(0)\, g(0)$ for $f, g \in \mathscr{F}^0$.

We illustrate these concepts by the following example.

Example. Let $\mathscr{D} = [0, 1]$ and $\mathscr{K}(x, y) = \min\{x, y\}$. Then $\mathscr{F}$ is the space of absolutely continuous functions f with $f(0) = 0$ and L_2-integrable f'. Of course,

$$\langle f, g \rangle_{\mathscr{F}} = \int_0^1 f'(x)\, g'(x)\, dx.$$

Let $h^* \equiv 1$. Then $\mathscr{F}^\gamma = W_2^1([0, 1])$ is the classical Sobolev space with inner product given by

$$\langle f, g \rangle_{\mathscr{F}^\gamma} = f(0)\, g(0) + \gamma^{-1} \int_0^1 f'(x)\, g'(x)\, dx$$

For $\gamma = 0$ we have $f' \equiv 0$ for all $f \in \mathscr{F}^0$, and we adopt the convention that $0/0 = 0$ in the formula above, as well as in the rest of this paper.

The *integration* solution operator is given by $\mathscr{S}(f) = \mathrm{INT}(f) = \int_{\mathscr{D}} f(x)\, dx$ with $\mathscr{G} = \mathbb{R}$. Then $\mathscr{S} h^* = 1$ and $\|\mathscr{S} h^*\|_{\mathscr{G}} = 1$. It is easy to check that the norm of INT over $\mathscr{F}^\gamma$ is given by

$$\lambda_{1,\gamma} = \|\mathrm{INT}\|^2_{\mathscr{F}^\gamma \to \mathscr{G}} = 1 + \gamma/3.$$

Since $\mathcal{S}$ is a functional then all $\lambda_{i,\gamma}=0$ for $i\geqslant 2$.

The *approximation* solution operator is given by $\mathcal{S}(f)=\mathrm{APP}(f)=f$ with $\mathcal{G}=\mathcal{L}_2([0,1])$. As before, $\mathcal{S}h^*=1$ and $\|\mathcal{S}h^*\|_{\mathcal{G}}=1$. For the approximation operator, the operator $\mathcal{W}_\gamma=\mathcal{S}_\gamma^*\mathcal{S}\colon \mathcal{F}^\gamma\to\mathcal{F}^\gamma$ is given by

$$\mathcal{W}_\gamma f(x)=\int_0^1 K_\gamma(x,y)\,f(y)\,dy=\int_0^1 f(y)\,dy+y\int_0^1 \min\{x,y\}\,f(y)\,dy.$$

Let (f,λ) be an eigenpair of $\mathcal{W}_\gamma$, $(\mathcal{W}_\gamma f)(x)=\lambda f(x)$ for $x\in[0,1]$. Then setting $x=0$ we obtain

$$\int_0^1 f(y)\,dy=\lambda f(0).$$

After differentiating with respect to x we obtain

$$f'(1)=0,$$
$$-\gamma f(x)=\lambda f''(x).$$

By substituting $f(x)=\cos(\alpha x-\alpha)$ we find out that the orthonormal sequence of eigenpairs of $\mathcal{W}_\gamma$ is equal to

$$\eta_{i,\gamma}(x)=a_{i,\gamma}\cos(\alpha_{i,\gamma}x-\alpha_{i,\gamma}),\qquad \lambda_{i,\gamma}=y/\alpha_{i,\gamma}^2,\quad i=1,2,...,$$

where $\alpha_{i,\gamma}\in((i-1)\pi,i\pi)$ is the unique solution of

$$\cot x=x/\gamma\qquad\text{for}\quad x\in((i-1)\pi,i\pi),$$

and

$$a_{i,\gamma}=\|\cos(\alpha_{i,\gamma}\cdot-\alpha_{i,\gamma})\|_{\mathcal{F}^\gamma}^{-1}=\left(\cos^2(\alpha_{i,\gamma})+\frac{\alpha_{i,\gamma}}{2\gamma}(\alpha_{i,\gamma}-0{,}5\sin(2\alpha_{i,\gamma}))\right)^{-1/2}.$$

2.2. *Weighted Multivariate Case*

The weighted multivariate problems studied in this paper are defined as tensor products of univariate problems over weighted spaces. For $d=1,2,...,$ assume that the weights

$$1\geqslant\gamma_{d,1}\geqslant\gamma_{d,2}\geqslant\cdots\geqslant\gamma_{d,d}\geqslant 0$$

are given.

The weighted real separable Hilbert space $\mathscr{F}_d$ of multivariate functions $f\colon \mathscr{D}^d \to \mathbb{R}$ is defined as the tensor product of $\mathscr{F}^{\gamma_{d,k}}$ spaces,

$$\mathscr{F}_d = \bigotimes_{k=1}^{d} \mathscr{F}^{\gamma_{d,k}}. \tag{5}$$

We remind the reader of the tensor product operations. The tensor product $f = f_1 \otimes \cdots \otimes f_d = \bigotimes_{k=1}^d f_k$ for numbers f_k is just the product $\prod_{k=1}^d f_k$. When the f_k are scalar functions, f is a function of d variables given by $f(t_1, ..., t_d) = \prod_{k=1}^d f_k(t_k)$.

By $\mathscr{F}_d = \bigotimes_{k=1}^d \mathscr{F}^{\gamma_{d,k}}$ we mean a Hilbert space spanned by $\bigotimes_{k=1}^d f_k$ with $f_k \in \mathscr{F}^{\gamma_{d,k}}$, and the inner product in $\mathscr{F}_d$ is given by $\langle \bigotimes_{k=1}^d f_k, \bigotimes_{k=1}^d h_k \rangle_{\mathscr{F}_d} = \prod_{k=1}^d \langle f_k, h_k \rangle_{\mathscr{F}^{\gamma_{d,k}}}$ for $f_k,\ h_k \in \mathscr{F}^{\gamma_{d,k}}$. Note that, in particular, $\{ \bigotimes_{k=1}^d \eta_{k, j_k} \}_{j_1, ..., j_k}$ is an orthogonal system in $\mathscr{F}_d$ if $\{\eta_{k,j}\}_j$ is an orthogonal system in $\mathscr{F}^{\gamma_{d,k}}$ for every $k = 1, ..., d$.

The reproducing kernel $\mathscr{K}_d$ of $\mathscr{F}_d$ is given by

$$\mathscr{K}_d(x, y) = \prod_{k=1}^{d} \mathscr{K}_{\gamma_{d,k}}(x_k, y_k) = \prod_{k=1}^{d} (h^*(x_k)\, h^*(y_k) + \gamma_{d,k} \mathscr{K}(x_k, y_k)) \tag{6}$$

for all $x = [x_1, ..., x_d],\ y = [y_1, ..., y_d] \in \mathscr{D}^d$.

We are ready to define weighted multivariate solution operators $\mathscr{S}_d$. In Subsection 2.1, we defined $\mathscr{S}\colon \mathscr{F}^{\gamma} \to \mathscr{G}$. Let $\mathscr{G}_d$ be a tensor product of $\mathscr{G}$,

$$\mathscr{G}_d = \bigotimes_{k=1}^{d} \mathscr{G}.$$

The *weighted multivariate solution* operator $\mathscr{S}_d : \mathscr{F}_d \to \mathscr{G}_d$ is defined as a tensor product of $\mathscr{S}$,

$$\mathscr{S}_d = \bigotimes_{k=1}^{d} \mathscr{S}. \tag{7}$$

Let us recall that for linear operators T_k, $T = \bigotimes_{k=1}^d T_k$ is a linear operator such that $T(\bigotimes_{k=1}^d f_k) = \bigotimes_{k=1}^d T_k(f_k)$. Then we also have that $\|T\|_{\mathscr{F}_d \to \mathscr{G}_d} = \prod_{k=1}^d \|T_k\|_{\mathscr{F}^{\gamma_{d,k}} \to \mathscr{G}}$.

We illustrate the multivariate case for the example from Subsection 2.1.

Example (continued). We now have $\mathscr{D}^d = [0, 1]^d$ and

$$\mathscr{K}_d(x, y) = \prod_{k=1}^{d} (1 + \gamma_{d,k} \min\{x_k, y_k\}).$$

The space $\mathscr{F}_d = W_2^{(1,1,\dots,1)}([0,1]^d)$ is now the tensor product of the classical Sobolev spaces $W_2^1([0,1])$ of functions f defined over $[0,1]^d$ with the following norm, see [10],

$$\|f\|^2_{\mathscr{F}_d} = f^2(0) + \sum_{u \neq \varnothing} \int_{[0,1]^{|u|}} \left(\prod_{k\in u} \frac{\partial}{\partial x_k} f_u(x) \right)^2 \prod_{k \in u} \frac{dx_k}{\gamma_{d,k}},$$

where the summation is with respect to all nonempty subsets $u \subseteq \{1, \dots, d\}$, $|u|$ denotes its cardinality, and $f_u(x) = f(y)$ where $y_k = x_k$ for $k \in u$ and $y_k = 0$ for $k \notin u$.

For the integration problem we have $\mathscr{G}_d = \mathbb{R}$ and $\mathrm{INT}_d(f) = \int_{[0,1]^d} f(x)\, dx$. For the approximation problem we have $\mathscr{G}_d = \mathscr{L}_2([0,1]^d)$ and $\mathrm{APP}_d(f) = f$.

2.3. *Tractability and Polynomial-Time Algorithms*

The problem studied in this paper is to approximate the solution elements $\mathscr{S}_d f$ by linear algorithms A that use finitely many linear functionals as information about f. That is,

$$A(f) = \sum_{i=1}^{n} a_i L_i(f) \tag{8}$$

for elements $a_i \in \mathscr{G}_d$ and continuous linear functionals $L_i\colon \mathscr{F}_d \to \mathbb{R}$. The number n of functionals used in (8) is called the cost of A, $\mathrm{cost}(A) = n$.[2]

The linear functionals L_i are from a given class Λ. In this paper we consider two classes. The first is $\Lambda^{\mathrm{all}} = \mathscr{F}_d^*$ consisting of *all* continuous linear functionals, and the second is Λ^{std} consisting of function and derivative evaluations. That is, $L \in \Lambda^{\mathrm{std}}$ iff there exists a sample point $x \in \mathscr{D}^d$ and an integer j such that $L(f) = f^{(j)}(x)$, $\forall f \in \mathscr{F}_d$.

We study the *worst case* setting in which the error of A is defined as

$$e(A, \mathscr{S}_d) = \sup_{\|f\|_{\mathscr{F}_d} \leqslant 1} \|\mathscr{S}_d f - A(f)\|_{\mathscr{G}_d}, \tag{9}$$

i.e., it is the (operator) norm $\|\mathscr{S}_d - A\|_{\mathscr{F}_d \to \mathscr{G}_d}$.

[2] As it is well known, the restriction to linear algorithms and nonadaptive information is without loss of generality for linear problems defined over Hilbert spaces in the worst case setting. Then, assuming that precomputation is allowed, the cost of computing $A(f)$ is equal to the computation of n functionals $L_i(f)$ as well as performing n multiplications and $n-1$ additions in the space $\mathscr{G}$. For simplicity, we take n as the cost of A; see, e.g., [13] for more details.

For $n=0$ we set $A=0$ and $e(0, \mathcal{S}_d)=\|\mathcal{S}_d\|_{\mathcal{F}_d\to\mathcal{G}_d}$. Hence, $\|\mathcal{S}_d\|_{\mathcal{F}_d\to\mathcal{G}_d}$ is the initial error for approximating $\mathcal{S}_d$ without sampling the function.

We would like to reduce the initial error by a factor ε, where $\varepsilon\in(0,1)$. That is, we seek an algorithm A for which

$$e(A, \mathcal{S}_d)\leqslant\varepsilon\,\|\mathcal{S}_d\|_{\mathcal{F}_d\to\mathcal{G}_d}. \tag{10}$$

If (10) holds then we say that $A(f)$ is an *ε-approximation*. Let $\mathrm{comp}(\varepsilon, d)$ be the worst case complexity (minimal cost) of computing an ε-approximation. That is,[3]

$$\mathrm{comp}(\varepsilon, d)=\min\{n\colon \exists A \text{ with } e(A, \mathcal{S}_d)\leqslant\varepsilon\,\|\mathcal{S}_d\|_{\mathcal{F}_d\to\mathcal{G}_d} \text{ and } \mathrm{cost}(A)\leqslant n\}. \tag{11}$$

As in [15] we say that the problem $\{\mathcal{S}_d\}$ is *tractable* iff there exist nonnegative numbers C, q and p such that

$$\mathrm{comp}(\varepsilon, d)\leqslant Cd^q\varepsilon^{-p}, \qquad \forall d\geqslant 1, \quad \forall\varepsilon\in(0,1]. \tag{12}$$

The infima of q and p for which (12) holds are called the *d-exponent* and *ε-exponent* of tractability for $\{\mathcal{S}_d\}$. If $q=0$ in (12) then we say that the problem $\{\mathcal{S}_d\}$ is *strongly tractable*.

Consider now a family of algorithms $\{A_{\varepsilon,d}\}$. We say that the family $\{A_{\varepsilon,d}\}$ is a *polynomial-time* algorithm iff there exist nonnegative C, q and p such that

$$e(A_{\varepsilon,d}, \mathcal{S}_d)\leqslant\varepsilon\,\|\mathcal{S}_d\|_{\mathcal{F}_d\to\mathcal{G}_d} \quad\text{and}\quad \mathrm{cost}(A_{\varepsilon,d})\leqslant Cd^q\varepsilon^{-p},\ \forall d\geqslant 1,\ \forall\varepsilon\in(0,1]. \tag{13}$$

The infima of q and p for which (13) holds are called the *d-exponent* and *ε-exponent* of the polynomial-time algorithm $\{A_{\varepsilon,d}\}$. If $q=0$ in (13) then we say that the sequence $\{A_{\varepsilon,d}\}$ is a *strongly polynomial-time* algorithm.

Obviously, a necessary condition for tractability is that $\mathcal{S}_d$ is compact. That is why we already assumed that $\mathcal{S}\colon\mathcal{F}^1\to\mathcal{G}$ is compact since this is equivalent to the compactness of $\mathcal{S}_d$ on $\mathcal{F}_d$ for any choice of parameters $\gamma_{1,d}, \ldots, \gamma_{d,d}$.

[3] Sometimes $\mathrm{comp}(\varepsilon, d)$ is defined as the minimal n for which the condition $e(A, \mathcal{S}_d)\leqslant\varepsilon$ holds. Our condition $e(A, \mathcal{S}_d)\leqslant\varepsilon\,\|\mathcal{S}_d\|_{\mathcal{F}_d\to\mathcal{G}_d}$ can be viewed as the normalization of the operator $\mathcal{S}_d$. That is, for $\mathcal{S}'_d:=\|\mathcal{S}_d\|^{-1}_{\mathcal{F}_d\to\mathcal{G}_d}\mathcal{S}_d$ we have $\|\mathcal{S}'_d\|_{\mathcal{F}_d\to\mathcal{G}_d}=1$ and the two conditions coincide.

3. WEIGHTED TENSOR PRODUCT ALGORITHMS

In this section, we define a class of *weighted tensor product* algorithms (or *WTP* algorithms, for short) for approximation of the weighted multivariate solution operator $\mathcal{S}_d$. These algorithms depend on a number of parameters. Our main issue will be to find out when there exist WTP algorithms that are polynomial, or even strongly polynomial. As we shall see, this will depend, in particular, on the sequence of weights $\gamma_{d,k}$.

As already mentioned in the introduction, WTP algorithms are related to the algorithm proposed by Smolyak [11] for (unweighted) tensor product problems. In Remark 1 we discuss similarities and differences between them.

We now define WTP algorithms. As Smolyak's algorithm, they are based on a sequence of algorithms for the scalar cases. Hence, we start with algorithms for the spaces $\mathcal{F}^{\gamma}$.

Let $\{U_{n,\gamma}\}$ be a sequence of algorithms of the form (8) for approximation of the operator $\mathcal{S}\colon \mathcal{F}^{\gamma} \to \mathcal{G}$. That is,

$$U_{n,\gamma}(f) = \sum_{i=1}^{n} a_{n,\gamma,i} L_{n,\gamma,i}(f) \tag{14}$$

for some elements $a_{n,\gamma,i}$ of $\mathcal{G}$ and some continuous linear functionals $L_{n,\gamma,i}$ from the class Λ. For $n=0$, we set $U_{0,\gamma}=0$. Observe that the cost of $U_{n,\gamma}$ is n.

We remark that since the spaces $\mathcal{F}^{\gamma}$ differ from the space $\mathcal{F}$ only by a one dimensional space spanned by h^*, it is usually enough to have algorithms for approximating the operator $\mathcal{S}$ over the space $\mathcal{F}$ and then properly extend these algorithms to the spaces $\mathcal{F}^{\gamma}$. Indeed, if $\{B_n\}$ is such a sequence for the space $\mathcal{F}$ then we may set

$$U_{1,\gamma}(f) = f(0)\, \mathcal{S}h^*,$$

and

$$U_{n,\gamma}(f) = f(0)\, \mathcal{S}h^* + B_{n-1}(f - f(0)h^*) \qquad \text{for} \quad n \geqslant 2. \tag{15}$$

In this case $U_{n,\gamma}$ does not depend on γ. Observe that $U_{n,\gamma}$ is well defined since $f - f(0)h^* \in \mathcal{F}$ and $B_{n-1}(f - f(0)h^*)$ makes sense. The information used by $U_{n,\gamma}$ consists of one function value $f(0)$ and the information used by B_{n-1}. Since the functional L, $L(f) = f(0)$, belongs to both classes Λ^{std} and Λ^{all}, the information used by $U_{n,\gamma}$ belongs to the same class Λ as the information used by B_{n-1}. The number of information evaluations used by $U_{n,\gamma}$ is at most n.

We assume that the errors of $U_{n,\gamma}$ given by (14) converge to zero. That is,

$$\lim_{n\to\infty} \|\mathcal{S}-U_{n,\gamma}\|_{\mathcal{F}^\gamma\to\mathcal{G}}=0, \tag{16}$$

For $\{U_{n,\gamma}\}$ given by (15), this is equivalent to assuming that

$$\lim_{n\to\infty} \|\mathcal{S}-B_n\|_{\mathcal{F}\to\mathcal{G}}=0 \tag{17}$$

since

$$\|\mathcal{S}-U_{n,\gamma}\|_{\mathcal{F}^\gamma\to\mathcal{G}}=\sqrt{\gamma}\,\|\mathcal{S}-B_{n-1}\|_{\mathcal{F}\to\mathcal{G}} \qquad \text{for} \quad n\geqslant 2.$$

For each weight γ, we assume that we have an increasing sequence of integers

$$m_{0,\gamma}=0<m_{1,\gamma}=1<m_{2,\gamma}<\cdots<m_{i,\gamma}, \tag{18}$$

and define

$$\Delta_{i,\gamma}(f)=U_{m_{i,\gamma},\gamma}(f)-U_{m_{i-1,\gamma},\gamma}(f) \qquad \text{for} \quad i\geqslant 1. \tag{19}$$

Observe that $\Sigma_{i=1}^{j}\Delta_{i,\gamma}=U_{m_{j,\gamma},\gamma}$ and $U_{m_{j,\gamma},\gamma}(f)$ converges to $\mathcal{S}(f)$ for every $f\in\mathcal{F}^\gamma$.

We are ready to consider the multivariate case. Let $\mathbb{N}_+^d$ be the set of vectors $\mathbf{i}=[i_1,\dots,i_d]$ with positive integer coefficients i_k. To stress their role, we shall refer to them as multi-indices. By $|\mathbf{i}|$ we mean $\sum_{k=1}^d i_k$.

Let $\{P_{n,d}\}$ be a sequence of subsets of $\mathbb{N}_+^d$ such that $P_{n,d}$ consists of n multi-indices, $P_{n,d}\subset P_{n+1,d}$ and $\bigcup_n P_{n,d}=\mathbb{N}_+^d$. Each set $P_{n,d}$ may depend on all weights $\gamma_{d,k}$ for $k=1,2,\dots,d$.

By the *weighted tensor product* (*WTP*) algorithm we mean the sequence of algorithms

$$A^*_{n,d}(f)=\sum_{\mathbf{i}\in P_{n,d}}\left(\bigotimes_{k=1}^{d}\Delta_{i_k,\gamma_{d,k}}\right)(f). \tag{20}$$

The WTP algorithm depends on a number of parameters. First of all, it depends on the sequence of weights $\gamma_{d,k}$, the sequence of cardinalities $m_{i,\gamma_{d,k}}$, as well as the sequence of sets $P_{n,d}$. It also depends on one-dimensional algorithms $U_{n,\gamma_{d,k}}$. By varying these parameters we obtain the class of WTP algorithms.

Since $\lim_{j\to\infty} \sum_{i=1}^{j} \Delta_{m_{i,\gamma}}(f) = \mathcal{S}f$ for every $f \in \mathcal{F}^{\gamma}$, we have

$$\mathcal{S}_d f = \sum_{\mathbf{i}\in\mathbb{N}_+^d} \left(\bigotimes_{k=1}^{d} \Delta_{i_k,\gamma_{d,k}} \right)(f), \qquad \forall f \in \mathcal{F}_d. \tag{21}$$

This yields

$$\begin{aligned}\|\mathcal{S}_d f - A^*_{n,d}(f)\|_{\mathcal{G}_d} &= \left\| \sum_{\mathbf{i}\in\mathbb{N}_+^d\setminus P_{n,d}} \bigotimes_{k=1}^{d} \Delta_{i_k,\gamma_{d,k}}(f) \right\|_{\mathcal{G}_d} \\ &\leqslant \sum_{\mathbf{i}\in\mathbb{N}_+^d\setminus P_{n,d}} \left\| \bigotimes_{k=1}^{d} \Delta_{i_k,\gamma_{d,k}}(f) \right\|_{\mathcal{G}_d}.\end{aligned}$$

Therefore the error of $A^*_{n,d}$ is bounded by

$$\begin{aligned}e(A^*_{n,d}, \mathcal{S}_d) &\leqslant \sum_{\mathbf{i}\in\mathbb{N}_+^d\setminus P_{n,d}} \left\| \bigotimes_{k=1}^{d} \Delta_{i_k,\gamma_{d,k}} \right\|_{\mathcal{F}_d\to\mathcal{G}_d} \\ &= \sum_{\mathbf{i}\in\mathbb{N}_+^d\setminus P_{n,d}} \prod_{k=1}^{d} \|\Delta_{i_k,\gamma_{d,k}}\|_{\mathcal{F}^{\gamma_{k,d}}\to\mathcal{G}}.\end{aligned} \tag{22}$$

This formula suggests that a good choice for $P_{n,d}$ is the set of n multi-indices $\mathbf{i}$ which correspond to the n largest norms of $\bigotimes_{k=1}^{d} \Delta_{i_k,\gamma_{d,k}}$.

We now discuss the special case when $\mathcal{S}$ is a continuos linear functional. Of course, then $\mathcal{G} = \mathcal{G}_d = \mathbb{R}$ and $\mathcal{S}_d$ is also a continuous linear functional. Since the case of $\Lambda = \Lambda^{\text{all}}$ is now trivial, we consider only $\Lambda = \Lambda^{\text{std}}$. When $\mathcal{S}_d$ is a functional, some properties of the WTP algorithm can be strengthened under the following additional assumptions. We assume that the algorithms $U_{n,\gamma}$ are defined by (15), i.e., in terms of algorithms B_n. Assume also that the information used by the algorithms B_n is *nested*, i.e., the evaluation points used by the information for B_n are contained in the set of evaluation points used by the information for B_{n+1} for all $n \geqslant 1$. Finally, assume that the algorithms B_n are *central*, i.e., they minimize the error for all information values. For more explanation and illustration of these assumptions we refer to [14]. The following result can easily be derived using a proof technique from Lemma 6 in [14]. Namely, under the three assumptions mentioned above, the worst case error of $A^*_{n,d}$ for approximating a linear functional $\mathcal{S}_d$ satisfies the equality

$$\begin{aligned}e^2(A^*_{n,d}, \mathcal{S}_d) &= \sum_{\mathbf{i}\in\mathbb{N}_+^d\setminus P_{n,d}} \prod_{k=1}^{d} \|\Delta_{i_k,\gamma_{d,k}}\|^2_{\mathcal{F}^{\gamma_{d,k}}\to\mathcal{G}} \\ &= \|\mathcal{S}_d\|^2_{\mathcal{F}_d\to\mathcal{G}} - \sum_{\mathbf{i}\in P_{n,d}} \prod_{k=1}^{d} \|\Delta_{i_k,\gamma_{d,k}}\|^2_{\mathcal{F}^{\gamma_{d,k}}\to\mathcal{G}}.\end{aligned} \tag{23}$$

We end this section by the following remark.

Remark 1. In [11], Smolyak proposed an algorithm for approximating tensor product problems. Specifically, he considered *unweighted* problems, i.e., $\gamma_{d,k} \equiv \gamma$. The main emphasis was to obtain asymptotic results for a fixed d. Early tests of Smolyak's algorithm were usually done only for $d = 2$.

Smolyak's algorithm uses scalar algorithms U_i with $m_i = \Theta(2^i)$ function evaluations. Moreover, given an integer $q \geqslant d$, the algorithm equals the sum of tensor products $\bigotimes_{k=1} \Delta_{i_k}$ with $|\mathbf{i}| \leqslant q$. That is, the set $P_{n,d}$ is

$$P_{n,d} = \{\mathbf{i}\colon |\mathbf{i}| \leqslant q\} \tag{24}$$

and its cardinality $n = \binom{q}{d}$. Hence, $n = n(q)$ varies only through the changes of q. For large d, the sequence $n(q)$ has huge gaps, and the Smolyak algorithm is only useful for a few initial $q = d, d+1, \ldots$. For example, consider multivariate integration with $d = 360$ which models collateralized mortgage obligations in finance; see [5, 7, 8]. Then $n(360) = 1$, $n(361) = 361$, $n(362) = 65{,}341$, $n(363) = 7{,}906{,}261$ and $n(364) = 719{,}469{,}751$, and probably only four initial values of q can be computationally realized. This undesirable property makes the practicality of the Smolyak algorithm questionable for large d; see [2] where this point is also mentioned.

Furthermore, as shown in [14], even though we may use optimal information for $d = 1$, the information used for $d \geqslant 2$ in the Smolyak algorithm is not optimal. However, the loss due to the use of nonoptimal information is not very significant.

Modifications of the Smolyak algorithm have been proposed in the literature. For instance, in [3] the cardinality m_i is independent on the coordinates, i.e., $m_{i,k}$ is used for the kth coordinate, and the choice of $m_{i,k}$ is based on error estimates. This decreases the gaps in the cardinalities of the sets $P_{n,d}$.

The Smolyak algorithm is also known in the literature as Boolean interpolation, discrete blending algorithm, hyperbolic cross points, and sparse grids algorithm; see, e.g., [2, 14].

The WTP algorithm proposed in this paper not only deals with weighted tensor products. It also uses a different and much more flexible selection of multi-indices $\mathbf{i}$ in the $P_{n,d}$ sets. In particular, as we shall see in Section 4, the cardinality of the set $P_{n,d}$ may be equal to n for arbitrary d, so that there is no gap in the cardinalities of the algorithms $A^*_{n,d}$. Furthermore, optimality of information is preserved if the parameters of the WTP algorithm are carefully chosen. Hence, the WTP algorithm can be more efficient than the Smolyak algorithm even for the unweighted case.

4. CLASS Λ^{all}

In this section, we study the WTP algorithm for the class $\Lambda^{\text{all}} = \mathscr{F}_d^*$ of all continuous linear functionals. We show that with a proper selection of the parameters, the WTP algorithm is optimal and, hence, is strongly polynomial (or polynomial) iff the problem $\{\mathscr{S}_d\}$ is strongly tractable (or tractable). We also find conditions on tractability and strong tractability of $\{\mathscr{S}_d\}$.

We specify the form of the WTP algorithm by assuming that the algorithms $U_{n,\gamma}$ of (14) have the form

$$U_{n,\gamma}(f) = \sum_{i=1}^{n} \langle f, \eta_{i,\gamma} \rangle_{\mathscr{F}^\gamma} \mathscr{S}\eta_{i,\gamma}, \qquad \forall f \in \mathscr{F}^\gamma,$$

where, as in Subsection 2.1, $\{\eta_{i,\gamma}, \lambda_{i,\gamma}\}$ is the orthonormal sequence of the eigenpairs of $\mathscr{W}_\gamma = \mathscr{S}_\gamma^* \mathscr{S}: \mathscr{F}^\gamma \to \mathscr{F}^\gamma$.

The sequence $m_{i,\gamma}$ of (18) is now defined as

$$m_{i,\gamma} = i \qquad \text{for} \quad i = 0, 1, ..., \ \text{and for all } \gamma.$$

This means that (19) becomes

$$\Delta_{i,\gamma}(f) = \langle f, \eta_{i,\gamma} \rangle_{\mathscr{F}^\gamma} \mathscr{S}\eta_{i,\gamma}, \qquad \forall f \in \mathscr{F}^\gamma,$$

whose square norm is $\|\Delta_{i,\gamma}\|^2_{\mathscr{F}^\gamma \to \mathscr{G}} = \lambda_{i,\gamma}$. Observe that the elements $\Delta_{i,\gamma}(f)$ and $\Delta_{j,\gamma}(f)$ are now orthogonal for $i \neq j$.

For the vectors $\mathbf{i} = [i_1, i_2, ..., i_d]$, $\boldsymbol{\gamma} = [\gamma_{d,1}, \gamma_{d,2}, ..., \gamma_{d,d}]$, and $x = [x_1, x_2, ..., x_d]$, denote

$$\begin{aligned} n_{\mathbf{i},\boldsymbol{\gamma}}(x) &= \eta_{i_1,\gamma_{d,1}}(x_1)\, \eta_{i_2,\gamma_{d,2}}(x_2) \cdots \eta_{i_d,\gamma_{d,d}}(x_d), \\ \lambda_{\mathbf{i},\boldsymbol{\gamma}} &= \lambda_{i_1,\gamma_{d,1}} \lambda_{i_2,\gamma_{d,2}} \cdots \lambda_{i_d,\gamma_{d,d}}. \end{aligned}$$

Then

$$\bigotimes_{k=1}^{d} \Delta_{i_k,\gamma_{d,k}}(f) = \langle f, \eta_{\mathbf{i},\boldsymbol{\gamma}} \rangle_{\mathscr{F}_d} \mathscr{S}_d \eta_{\mathbf{i},\boldsymbol{\gamma}}$$

and

$$\left\langle \bigotimes_{k=1}^{d} \Delta_{i_k,\gamma_{d,k}}(f), \bigotimes_{k=1}^{d} \Delta_{j_k,\gamma_{d,k}}(f) \right\rangle_{\mathscr{G}_d} = \lambda_{\mathbf{i},\boldsymbol{\gamma}} \langle f, \eta_{\mathbf{i},\boldsymbol{\gamma}} \rangle^2_{\mathscr{F}_d} \delta_{\mathbf{i},\mathbf{j}}$$

with $\delta_{\mathbf{i},\mathbf{j}}$ being zero or one depending on whether $\mathbf{i}$ and $\mathbf{j}$ are different or not. This implies that

$$\left\| \bigotimes_{k=1}^{d} \Delta_{i_k, \gamma_{d,k}} \right\|^2_{\mathscr{F}_d \to \mathscr{G}_d} = \lambda_{\mathbf{i},\boldsymbol{\gamma}}.$$

Clearly, $\{\eta_{\mathbf{i},\boldsymbol{\gamma}}, \lambda_{\mathbf{i},\boldsymbol{\gamma}}\}$ is the sequence of the orthonormal eigenpairs of the operator $\mathscr{W}_d = \mathscr{S}_d^* \mathscr{S}_d \colon \mathscr{F}_d \to \mathscr{F}_d$, and $\lambda_{\mathbf{i},\boldsymbol{\gamma}}^{1/2}$'s are singular values of $\mathscr{S}_d$.

We order the numbers $\lambda_{\mathbf{i},\boldsymbol{\gamma}}$. That is, we define a sequence of multi-indices $\mathbf{i}_j = \mathbf{i}_j(\boldsymbol{\gamma})$ such that $\mathbf{i}_j \neq \mathbf{i}_k$ for $j \neq k$, $\bigcup_{j=1}^{\infty} \{\mathbf{i}_j\} = \mathbb{N}_+^d$ and

$$\lambda_{\mathbf{i}_1,\boldsymbol{\gamma}} \geqslant \lambda_{\mathbf{i}_2,\boldsymbol{\gamma}} \geqslant \cdots \geqslant \lambda_{\mathbf{i}_j,\boldsymbol{\gamma}} \geqslant 0.$$

We take for $P_{n,d}$ the sets of n multi-indices $\mathbf{i}_j$ which correspond to the n largest eigenvalues $\lambda_{\mathbf{i}_j,\boldsymbol{\gamma}}$. That is,

$$P_{n,d} = \{\mathbf{i}_1, \mathbf{i}_2, ..., \mathbf{i}_n\}.$$

This completes the construction of a WTP algorithm, and we have

$$A_{n,d}^*(f) = \sum_{j=1}^{n} \langle f, \eta_{\mathbf{i}_j,\boldsymbol{\gamma}} \rangle_{\mathscr{F}_d} \mathscr{S}_d \eta_{\mathbf{i}_j,\boldsymbol{\gamma}}. \tag{25}$$

This WTP algorithm is the projection of $\mathscr{S}_d$ to the n dimensional subspace spanned by eigenelements of the operators $\mathscr{W}_d$ which corresponds to the n largest eigenvalues of $\mathscr{W}_d$. Clearly, the cost of $A_{n,d}^*$ is n.

The error between $\mathscr{S}_d f$ and $A_{n,d}^*(f)$ is equal to

$$\begin{aligned} \|\mathscr{S}_d f - A_{n,d}^*(f)\|_{\mathscr{G}_d}^2 &= \left\langle \sum_{j=n+1}^{\infty} \langle f, \eta_{\mathbf{i}_j,\boldsymbol{\gamma}} \rangle_{\mathscr{F}_d} \mathscr{W}_d \eta_{\mathbf{i}_j,\boldsymbol{\gamma}}, \sum_{j=n+1}^{\infty} \langle f, \eta_{\mathbf{i}_j,\boldsymbol{\gamma}} \rangle_{\mathscr{F}_d} \eta_{\mathbf{i}_j,\boldsymbol{\gamma}} \right\rangle_{\mathscr{F}_d} \\ &= \sum_{i=n+1}^{\infty} \lambda_{\mathbf{i}_j,\boldsymbol{\gamma}} \langle f, \eta_{\mathbf{i}_j,\boldsymbol{\gamma}} \rangle_{\mathscr{F}_d}^2. \end{aligned}$$

This yields

$$e^2(A_{n,d}^*, \mathscr{S}_d) = \lambda_{\mathbf{i}_{n+1},\boldsymbol{\gamma}}.$$

That is, the square of the error of $A_{n,d}^*$ is the $(n+1)$st largest eigenvalue of $\mathscr{W}_d$. It is known, see, e.g., [13], that this is the smallest possible worst case error of any algorithm that uses n continuous linear functionals. This proves the following proposition.

PROPOSITION 1. *The WTP algorithm $A^*_{n,d}$ defined by (25) is optimal among all algorithms that use n continuous linear functionals.*

If we want to compute an ε-approximation by the WTP algorithm, we set

$$n = n(\varepsilon, d) = \min\{k: \sqrt{\lambda_{\mathbf{i}_{k+1}}} \leqslant \varepsilon \, \|S\|_{\mathscr{F}_d \to \mathscr{G}_d}\}.$$

Then

$$\mathscr{A}^*_{\varepsilon, d} = A^*_{n(\varepsilon, d), d} \tag{26}$$

has worst case error at most ε. Optimality of $A_{n,d}$ implies that

$$\mathrm{comp}(\varepsilon, d) = \mathrm{cost}(\mathscr{A}^*_{\varepsilon, d}) = n(\varepsilon, d).$$

We now check when WTP is a polynomial-time algorithm. Due to its optimality, this is equivalent to checking when the problem $\{\mathscr{S}_d\}$ is tractable. As we shall see, tractability will depend on the sequences of weights $\{\gamma_{d,k}\}$ and eigenvalues $\{\lambda_i\}$ of the operator $\mathscr{W} = \mathscr{S}^*\mathscr{S}: \mathscr{F}^1 \to \mathscr{F}^1$.

We need to define the *sum-exponent* of an arbitrary sequence $\{\psi_{d,k}\}$ with nonnegative numbers $\psi_{d,k}$ (for $d = 1, 2, ...,$ and $k = 1, 2, ..., d$) such that $\psi_{d,1} \geqslant \psi_{d,2} \geqslant \cdots \geqslant \psi_{d,d} \geqslant 0$. We say that p_ψ is the sum-exponent of $\{\psi_{d,k}\}$ iff

$$p_\psi = \inf\left\{\alpha \geqslant 0 : \sup_d \sum_{k=1}^{d} \psi^\alpha_{d,k} < \infty\right\} \tag{27}$$

with the convention that $\inf \varnothing = \infty$.

Observe that for $\xi_{d,k} = \Theta(k^{-\beta})$ with positive β and with constants in the Theta notation independent of d, we have $p_\xi = \beta^{-1}$. On the other hand, for any sequence $\{\xi_{d,k}\}$ with $p_\xi > 0$ we have the following. For any $\alpha > p_\xi$ there exists a nonnegative $M = M(\alpha)$ such that $\Sigma^d_{k=1} \xi^\alpha_{d,k} \leqslant M$, $\forall d$. Since $\xi_{d,k}$ are ordered, $k\xi^\alpha_{d,k} \leqslant \sum_{j=1}^k \xi^\alpha_{d,j} \leqslant M$ and $\xi_{d,k} \leqslant M^{1/\alpha} k^{-1/\alpha}$. Hence, $\xi_{d,k}$ goes to zero as $k^{-1/\alpha}$ and α can be arbitrarily close to p_ξ. This shows that the sum-exponent measures how fast $\{\xi_{d,k}\}$ goes to zero as a function of k and p_ξ is positive only if $\{\xi_{d,k}\}$ goes to zero polynomially in k^{-1}. In this case, p_ξ is the inverse of the largest degree polynomial for which this happens.

Recall that λ_i given by (1) are ordered eigenvalues of the operator $\mathscr{W} = \mathscr{S}^*\mathscr{S}$. If $\mathscr{S}$ is a continuous linear functional then, obviously, $\lambda_i = 0$, $\forall i \geqslant 2$. In this case $\mathscr{S}_d$ is also a continuous linear functionals and the WTP algorithm has zero error for all $n \geqslant 1$. Hence, the problem is trivially strongly tractable. In what follows, we therefore assume that the rank of $\mathscr{S}$ is at least two. This is the same as assuming that

$$\lambda_2 > 0. \tag{28}$$

Observe that (28) implies that $\gamma_{d,1} > 0$. Indeed, $\gamma_{d,1} = 0$ yields that all $\gamma_{d,j} = 0$ and the space $\mathscr{F}_d = \operatorname{span}\{\mathbf{h}^*\}$ is one-dimensional, where $\mathbf{h}^*(x_1, x_2, ..., x_d) = \prod_{k=1}^d h^*(x_k)$. Then all $\mathscr{S}_d$ are linear functionals and $\lambda_2 = 0$.

We are ready to present necessary and sufficient conditions on strong tractability of the problem $\{\mathscr{S}_d\}$.

THEOREM 1. *Let p_γ and p_λ be, respectively, the sum-exponents of $\{\gamma_{d,k}\}$ and of $\{\lambda_{d,k}\}$ with $\lambda_{d,k} = \lambda_k$. Let (28) hold and let*

$$\beta = \|\mathscr{S} h^*\|_{\mathscr{G}}. \tag{29}$$

If $\beta = 0$ then the problem $\{\mathscr{S}_d\}$ is not *tractable.*

If $\beta > 0$ then the poblem $\{\mathscr{S}_d\}$ is strongly tractable iff $p_\gamma < \infty$ and $p_\lambda < \infty$.

If the problem $\{\mathscr{S}_d\}$ is strongly tractable then the ε-exponent $p^* = p^*(\{\mathscr{S}_d\}, \Lambda^{\text{all}})$ of strong tractability is

$$p^* = 2 \max\{p_\gamma, p_\lambda\}.$$

The WTP algorithm $\mathscr{A}^*_{\varepsilon,d}$ is then strongly polynomial and

$$\operatorname{comp}(\varepsilon, d) = \operatorname{cost}(\mathscr{A}^*_{\varepsilon,d}) \leqslant C_p \varepsilon^{-p}, \qquad \forall p > p^*,$$

with

$$C_p = \alpha_p \prod_{j=2}^{\infty} \left(1 + \frac{\gamma_{d,j}^{p/2}}{\beta^p} \sum_{k=1}^{\infty} \lambda_k^{p/2}\right) < \infty \tag{30}$$

and

$$\alpha_p = \sup_{\varepsilon \in (0,1]} \varepsilon^p \min\{k \geqslant 2 : \lambda_{k-1} \leqslant \varepsilon^2 \beta^2\} < \infty. \tag{31}$$

Proof. Assume first that $\beta = 0$. Then the eigenvalues of $\mathscr{W}_\gamma$ are $\lambda_{i,\gamma} = \gamma \lambda_i$ for $i \geqslant 1$, and the eigenvalues of $\mathscr{W}_d$ are

$$\lambda_{\mathbf{i},\boldsymbol{\gamma}} = \gamma_{d,1} \gamma_{d,2} \cdots \gamma_{d,d} \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_d}$$

and $\|\mathscr{S}_d\|^2_{\mathscr{F}_d \to \mathscr{G}_d} = \gamma_{d,1}\gamma_{d,2}\cdots\gamma_{d,d}\lambda_1^d$. Observe that

$$\mathrm{comp}(\varepsilon, d) = |\{\mathbf{i} = [i_1, ..., i_d] : \lambda_{\mathbf{i},\boldsymbol{\gamma}} > \varepsilon^2 \, \|\mathscr{S}_d\|^2_{\mathscr{F}_d \to \mathscr{G}_d}\}|. \tag{32}$$

Let $\lambda_i^* = \lambda_i/\lambda_1$. Then we can rewrite (32) as

$$\mathrm{comp}(\varepsilon, d) = |\{\mathbf{i} = [i_1, ..., i_d] : \lambda_{i_1}^* \lambda_{i_2}^* \cdots \lambda_{i_d}^* > \varepsilon^2\}|.$$

In Theorem 3.1 of [16] it is proven that $\lambda_1^* = 1$ and $\lambda_2^* > 0$ imply that $\{\mathscr{S}_d\}$ is not tractable.

Assume then that $\beta > 0$. For this case, the proof is similar to the proof of Theorem 1 in [17]. Take first $d = 1$. Then $\mathscr{S}_1 = \mathscr{S} : \mathscr{F}^{\gamma_{d,1}} \to \mathscr{G}$. The worst case complexity is now given by

$$\mathrm{comp}(\varepsilon, 1) = |\{k \colon \lambda_{k,\gamma_{d,1}} > \varepsilon^2 \lambda_{1,\gamma_{d,1}}\}|$$

and due to (3) we have $\beta^2 \leqslant \lambda_{1,\gamma_{d,1}} \leqslant \beta^2 + \gamma_{d,1}\lambda_1 \leqslant \beta^2 + \lambda_1$ with a positive $\gamma_{d,1}$.

Due to (4), we know that

$$\gamma_{d,1}\lambda_k \leqslant \lambda_{k,\gamma_{d,1}} \leqslant \gamma_{d,1}\lambda_{k-1}, \qquad \forall k \geqslant 2.$$

Hence, $\mathrm{comp}(\varepsilon, 1) = O(\varepsilon^{-p})$ iff $\lambda_{k,\gamma_{d,1}} = O(k^{-2/p})$ iff $\lambda_k = O(k^{-2/p})$ iff $p_\lambda \leqslant p/2$. This proves that $p_\lambda < \infty$ is a necessary condition for tractability and that the ε-exponent is at least $2p_\lambda$.

Consider the case $d \geqslant 1$. Take vectors $\mathbf{i}$ with $(d-1)$ components equal to 1 and the remaining component equal to 2. That is, for some k we have $i_k = 2$ and $i_j = 1$ for all $j \neq k$. Since $\lambda_{1,\gamma} \in [\beta^2, \beta^2 + \gamma\lambda_1]$ and $\lambda_{i,\gamma} \geqslant \gamma\lambda_i$ for all $i \geqslant 2$ due to (3) and (4), we have

$$\lambda_{\mathbf{i},\boldsymbol{\gamma}} \geqslant \gamma_{d,k}\lambda_2 \prod_{j \neq k} \lambda_{1,\gamma_{d,j}}.$$

Since λ_2 is positive and $\|\mathscr{S}_d\|^2_{\mathscr{F}_d \to \mathscr{G}_d} = \prod_{j=1}^d \lambda_{1,\gamma_{d,j}}$, (32) yields

$$\mathrm{comp}(\varepsilon, d) \geqslant |\{k \colon \gamma_{d,k} > \varepsilon^2(\beta^2 + \gamma_{d,k}\lambda_1)/\lambda_2\}|.$$

For $\varepsilon^2 < \lambda_2/\lambda_1$, we have

$$\mathrm{comp}(\varepsilon, d) \geqslant |\{k \colon \gamma_{d,k} > \varepsilon^2\beta^2/(\lambda_2 - \varepsilon^2\lambda_1)\}|.$$

Hence, $\mathrm{comp}(\varepsilon, d) \leqslant C\varepsilon^{-p}$ for all d and all $\varepsilon \in (0, 1]$ holds iff $\gamma_{d,k} = O(k^{-2/p})$ iff $p_\gamma \leqslant p/2$. This proves that $p_\gamma < \infty$ is a necessary condition for strong tractability and that the ε-exponent is at least $2p_\gamma$.

Assume then that both p_γ and p_λ are finite. Take $p > 2\max\{p_\gamma, p_\lambda\}$. Since $\lambda_k = O(k^{-2/p})$, there exists a constant α_p such that

$$\min\{k \geqslant 2 : \lambda_{k-1} \leqslant \varepsilon^2\beta^2\} \leqslant \alpha_p \varepsilon^{-p}, \qquad \forall \varepsilon \in (0, 1].$$

This proves (31). For $d = 1$ we have

$$\mathrm{comp}(\varepsilon, d) = \min\{k - 1 : \lambda_{k, \gamma_{d,1}} \leqslant \varepsilon^2 \lambda_{1, \gamma_{d,1}}\}.$$

Since $\lambda_{1,\gamma_{d,1}} \geqslant \beta^2$ and $\lambda_{k,\gamma_{d,1}} \leqslant \gamma_{d,1}\lambda_{k-1} \leqslant \lambda_{k-1}$, we have

$$\mathrm{comp}(\varepsilon, 1) \leqslant \min\{k \geqslant 2 : \lambda_{k-1} \leqslant \varepsilon^2\beta^2\} \leqslant \alpha_p \varepsilon^{-p}.$$

We now show that C_p of (30) is finite. First of all, notice that $p/2 > p_\lambda$ implies that the series $\alpha = \sum_{k=1}^\infty \lambda_k^{p/2}$ is finite. Hence, it is enough to prove that the product $\prod_{j=2}^\infty (1 + \gamma_{d,j}^{p/2} \alpha/\beta^p)$ is finite. This product is finite iff $\Sigma_{j=2}^\infty \gamma_{d,j}^{p/2}$ is finite. This holds since $p/2 > p_\gamma$. Hence, C_p is finite.

For $j \geqslant 1$, define

$$\begin{aligned}\mathrm{comp}(\varepsilon, j) = |\{\mathbf{i} = [i_1, i_2, ..., i_j] : \lambda_{i_1, \gamma_{d,1}} \lambda_{i_2, \gamma_{d,2}} \cdots \lambda_{i_j, \gamma_{d,j}} \\ > \varepsilon^2 \lambda_{1,\gamma_{d,1}} \lambda_{1,\gamma_{d,2}} \cdots \lambda_{1,\gamma_{d,j}}\}|\end{aligned}$$

as the worst case complexity of computing an ε-approximation to

$$\mathcal{S}_j = \bigotimes_{k=1}^{j} \mathcal{S} : \bigotimes_{k=1}^{j} \mathcal{F}^{\gamma_{d,j}} \to \mathcal{G}_j = \bigotimes_{k=1}^{j} \mathcal{G}.$$

We are interested in estimating $\mathrm{comp}(\varepsilon, d)$.

Assume inductively that $\mathrm{comp}(\varepsilon, j) \leqslant \beta_j \varepsilon^{-p}$ for some β_j. For $j = 1$ we have $\mathrm{comp}(\varepsilon, 1) \leqslant \beta_1 \varepsilon^{-p}$ with $\beta_1 = \alpha_p$. Observe that $\mathrm{comp}(\varepsilon, j)$ can be rewritten as

$$\begin{aligned}\mathrm{comp}(\varepsilon, j) &= \sum_{k=1}^\infty \left| \left\{ \mathbf{i} = [i_1, i_2, ..., i_{j-1}] : \frac{\lambda_{i_1,\gamma_{d,1}}}{\lambda_{1,\gamma_{d,1}}} \cdots \frac{\lambda_{i_{j-1},\gamma_{d,j-1}}}{\lambda_{1,\gamma_{d,j-1}}} > \varepsilon^2 \frac{\lambda_{1,\gamma_{d,j}}}{\lambda_{k,\gamma_{d,j}}} \right\} \right| \\ &= \sum_{k=1}^\infty \mathrm{comp}(\varepsilon(\lambda_{1,\gamma_{d,j}}/\lambda_{k,\gamma_{d,j}})^{1/2}, j-1) \leqslant \beta_{j-1}\varepsilon^{-p} \sum_{k=1}^\infty \left(\frac{\lambda_{k,\gamma_{d,j}}}{\lambda_{1,\gamma_{d,j}}}\right)^{p/2} \\ &= \beta_{j-1}\varepsilon^{-p}\left(1 + \sum_{k=2}^\infty \left(\frac{\lambda_{k,\gamma_{d,j}}}{\lambda_{1,\gamma_{d,j}}}\right)^{p/2}\right).\end{aligned}$$

Due to (3) and (4) we obtain

$$\mathrm{comp}(\varepsilon, j) \leqslant \beta_{j-1} \varepsilon^{-p} \left(1 + \frac{\gamma_{d,j}^{p/2}}{\beta^p} \sum_{k=1}^{\infty} \lambda_k^{p/2}\right).$$

As already remarked, the last series is finite since $p/2 > p_\lambda$. Thus we can set

$$\begin{aligned}\beta_j &= \beta_{j-1}\left(1 + \frac{\gamma_{d,j}^{p/2}}{\beta^p} \sum_{k=1}^{\infty} \lambda_k^{p/2}\right)\\ &= \beta_1 \prod_{l=2}^{j} \left(1 + \frac{\gamma_{d,l}^{p/2}}{\beta^p} \sum_{k=1}^{\infty} \lambda_k^{p/2}\right) \leqslant C_p < \infty.\end{aligned}$$

This proves that

$$\mathrm{comp}(\varepsilon, d) \leqslant C_p \varepsilon^{-p}.$$

Hence, the problem $\{\mathscr{S}_d\}$ is strongly tractable and since p can be arbitrarily close to $p^* = 2\max\{p_\gamma, p_\lambda\}$, the ε-exponent of strong tractability is at most p^*. From the first part of the proof we know that the ε-exponent is at least p^*. Therefore p^* is the ε-exponent. This completes the proof. ∎

The first part of Theorem 1 states that $\mathscr{S}h^* = 0$ implies intractability for all $\mathscr{S}$ of rank 2. In this case, the weights $\gamma_{d,j}$ do not play any role since they are multipliers in the sequence of singular values of $\mathscr{S}_d$ and they cancel when we consider the reduction of the initial error. The problem then becomes "unweighted". We add that tractability issues for unweighted problems with $\mathscr{S}$ of rank 1 in the class Λ^{std} are quite complicated and rich in possibility, see [4, 14].

The second part of Theorem 1 states necessary and sufficient conditions on strong tractability and provides the ε-exponent of strong tractability. Strong tractability is equivalent to the fact that the sum-exponents of the two sequences are finite. The first sequence is the sequence of weights. Hence, the weights must go polynomially to zero and the speed of their convergence affects the ε-exponent of strong tractability. The second sequence is the sequence of one dimensional singular values of the operator. Clearly, they must go polynomially to zero to even guarantee that the one dimensional problem is tractable.

The ε-exponent of strong tractability is $p^* = 2\max\{p_\gamma, p_\lambda\}$. Observe that $2p_\lambda$ measures the speed of convergence for the one dimensional case, and $p^* \geqslant 2p_\lambda$ simply states that there is no way to beat the complexity for the one dimensional case. If $p_\gamma \leqslant p_\lambda$ then the affect of arbitrary dimension is negligible and the complexity for arbitrary dimension d behaves essentially as for $d=1$. If, however, $p_\gamma > p_\lambda$ then we still have strong tractability but the multivariate complexity is larger than the one dimensional complexity.

For example, consider $\lambda_k = \Theta(k^{-2r})$ which corresponds to many practical problems defined over spaces with r times differentiable functions. Then $2p_\lambda = 1/r$ is the usual exponent of the one dimensional complexity. Consider now the weights $\gamma_{d,k} = \Theta(k^{-2m})$ with the constants in the Theta notation independent of d and k. Then $2p_\gamma = 1/m$ and $p^* = \max\{r^{-1}, m^{-1}\}$. Hence, for $m \geqslant r$ the effect of dimension is negligible.

We also remark that the ε-exponent p^* may be equal to 0. This happens if both weights and singular values tend to zero faster than polynomial. For instance, this is the case for exponential weights and singular values, i.e., sequences of the form ρ^k with $\rho \in [0, 1)$.

We now discuss tractability of the problem $\{\mathscr{S}_d\}$. As already mentioned in the proof of Theorem 1, $p_\lambda < \infty$ is also a necessary condition for tractability. However, the problem $\{\mathscr{S}_d\}$ may be tractable for $p_\gamma = \infty$. It may happen even for an operator $\mathscr{S}$ of rank two. Indeed, assume for simplicity that $\mathscr{S}^*\mathscr{S}h^* = h^*$, i.e., $\beta = 1$, and that the eigenvalues of $\mathscr{W}$ in (1) are $\lambda_1 = \lambda_2 = 1$ and $\lambda_i = 0$ for all $i \geqslant 3$. For a positive q, define the following sequence of weights

$$\gamma_{d,k} = \begin{cases} 1 & \text{if} \quad k = 1, 2, ..., \lceil q \log_2 d \rceil, \\ 0 & \text{otherwise.} \end{cases}$$

Observe that $p_\gamma = \infty$ and therefore strong tractability does not hold. It is easy to see that the operator $\mathscr{S}_d$ has exactly $2^{\lceil q \log_2 d \rceil}$ eigenvalues equal to one, and the rest of them equal to zero. Therefore for all $\varepsilon < 1$ we have

$$\operatorname{comp}(\varepsilon, d) = 2^{\lceil q \log_2 d \rceil} = \alpha d^q \qquad \text{for some} \quad \alpha \in (1, 2].$$

Hence, the problem is tractable with the d-exponent equal to q, and the ε-exponent equal to zero.

The essence of the example above is that the weights $\gamma_{d,k}$ depend essentially on d and k. As we shall see in the next theorem, for weights depending essentially only on k a quite different result holds. In what follows, we assume that there exist two positive numbers β_1 and β_2, and a sequence η_k with $\eta_1 \geqslant \eta_2 \geqslant \cdots \geqslant 0$, such that

$$\beta_1 \eta_k \leqslant \gamma_{d,k} \leqslant \beta_2 \eta_k, \qquad \text{for} \quad d = 1, 2, ..., \quad k = 1, 2, ..., d. \tag{33}$$

THEOREM 2. *Let the sequence* $\{\gamma_{d,k}\}$ *satisfy* (33), *and let* $\lambda_2 > 0$. *Then the problem* $\{\mathscr{S}_d\}$ *is tractable iff it is strongly tractable.*

Proof. It is enough to prove that tractability implies strong tractability. This means that we only need to show that $p_\gamma < \infty$. Due to (33), we have $p_\gamma = p_\eta$. Assume that $\mathrm{comp}(\varepsilon, d) \leqslant Cd^q \varepsilon^{-p}$ for some nonnegative C, p and q, where $\mathrm{comp}(\varepsilon, d)$ is given by (32).

Take $d > q+1$. Choose vectors $\mathbf{i}$ with $(d-q-1)$ components equal to 1 and with $q+1$ components equal to 2. We have $\binom{d}{q+1} = \Theta(d^{q+1})$ such vectors. Let $\beta = \|\mathscr{S}h^*\|_{\mathscr{G}}$. Then (3), (4) and (33) yield

$$\frac{\lambda_{i_1, \gamma_{d,1}}}{\lambda_{1, \gamma_{d,1}}} \cdots \frac{\lambda_{i_d, \gamma_{d,d}}}{\lambda_{1, \gamma_{d,d}}} \leqslant \left(\frac{\gamma_{d,d}\lambda_2}{\beta^2 + \lambda_1}\right)^{q+1} \geqslant \left(\frac{\beta_1 \lambda_2}{\beta^2 + \lambda_1}\right)^{q+1} \eta_d^{q+1}.$$

Take $\varepsilon = 0.5(\beta_1 \lambda_2 \eta_d / (\beta^2 + \lambda_1))^{(q+1)/2}$. Then all vectors $\mathbf{i}$ belong to the set in (32). Hence,

$$\Theta(d^{q+1}) = \binom{d}{q+1} \leqslant \mathrm{comp}(\varepsilon, d) \leqslant Cd^q \varepsilon^{-p}.$$

Substituting ε we get

$$\eta_d = O(d^{-2/(p(q+1))}).$$

This means that $p_\eta < \infty$, and completes the proof. ∎

5. ARBITRARY CLASS Λ

In this section we study a WTP algorithm whose information about f is not necessarily given by the values of inner-products with the eigenelements of $\mathscr{W}_\gamma$. We also assume that the algorithms $U_{n,\gamma}$ which are used in the WTP algorithm are given by (15). Hence, the information used by the WTP algorithm is the same as the information used by the algorithms B_n which may be from the class Λ^{all} or Λ^{std}. Of course, the more interesting case is when B_n uses information from the class Λ^{std} since then the results of the previous section do not apply. The analysis of the WTP algorithm will now be different than in the previous section, that is, it will be not based on the spectrum of $\mathscr{W}_\gamma$.

We assume that we know a sequence $\{B_n\}$, $B_0 = 0$, of algorithms such that B_n is of the form (14) and satisfies (17). We stress that the cost of B_n is n. We also assume that there exists a sequence

$$m_0 = 0 < m_1 = 1 < \cdots < m_k < \cdots \tag{34}$$

such that

$$\|B_{m_i-1} - B_{m_{i-1}-1}\|_{\mathscr{F}\to\mathscr{G}} \leqslant CD^{i-1}\, \|Sh^*\|_{\mathscr{G}}, \qquad \forall i \geqslant 2, \tag{35}$$

and

$$m_i D^{(i-1)p} \leqslant 1, \qquad \forall i \geqslant 2, \tag{36}$$

for some $D \in (0, 1)$ and some nonnegative C and p. To satisfy (35) for nontrivial problems, we need to assume that $\|\mathscr{S}h^*\|_{\mathscr{G}} > 0$. This condition is necessary for tractability for all operators $\mathscr{S}$ of rank at least 2, see Theorem 1. As will be explained at the end of Section 6, the conditions (35) and (36) hold with $p = 1/r$ whenever the *nth minimal error* $e(n) = e(n; \mathscr{S}, \mathscr{F})$ is bounded by

$$e(n; \mathscr{S}, \mathscr{F}) = O(n^{-r}).$$

Recall that by the definition,

$$e(n; \mathscr{S}, \mathscr{F}) = \inf\{e(A, S)\colon A \text{ uses } n \text{ function evaluations}\}.$$

Let $U_{i,\gamma}$ be given by (15) and $\Delta_{i,\gamma}$ by (19). Since neither $U_{i,\gamma}$ nor $\Delta_{i,\gamma}$ depend on γ, we shall denote them in this section by U_i and Δ_i, respectively. Then $\Delta_1(f) = U_1(f) = f(0)\mathscr{S}h^*$, its cost is 1, and (2) yields

$$\|\Delta_1\|_{\mathscr{F}^\gamma\to\mathscr{G}} = \|\mathscr{S}h^*\|_{\mathscr{G}} \leqslant \|\mathscr{S}\|_{\mathscr{F}^\gamma\to\mathscr{G}}. \tag{37}$$

For $i \geqslant 2$, we have $\Delta_i(f) = (B_{m_i-1} - B_{m_{i-1}-1})(f - f(0)h^*)$. Since $\|f - f(0)h^*\|_{\mathscr{F}^\gamma} = \sqrt{\|f\|^2_{\mathscr{F}^\gamma} - f^2(0)} \leqslant \|f\|_{\mathscr{F}^\gamma}$, we obtain

$$\|\Delta_i\|_{\mathscr{F}^\gamma\to\mathscr{G}} = \sqrt{\gamma}\, \|B_{m_i-1} - B_{m_{i-1}-1}\|_{\mathscr{F}\to\mathscr{G}} \leqslant \sqrt{\gamma}\, CD^{i-1}\, \|\mathscr{S}h^*\|_{\mathscr{G}} \tag{38}$$

due to (35). Once more (2) yields

$$\|\Delta_i\|_{\mathscr{F}^\gamma\to\mathscr{G}} \leqslant \sqrt{\gamma}\, CD^{i-1}\, \|\mathscr{S}\|_{\mathscr{F}^\gamma\to\mathscr{G}}, \qquad \forall i \geqslant 2. \tag{39}$$

Note that the information used by Δ_i is $f(0)$ and the information used by B_{m_i-1} and $B_{m_{i-1}-1}$. For $i = 2$, the cost of Δ_2 is at most m_2 since $B_{m_1-1} = B_0 = 0$. For $i \geqslant 3$, the cost of Δ_i is at most $m_i + m_{i-1} - 1 \leqslant 2m_i$. For nested information, the algorithm B_{m_i-1} reuses the information used by $B_{m_{i-1}-1}$, and therefore the cost of Δ_i is at most m_i.

As in (21) we have

$$\mathcal{S}_d = \sum_{\mathbf{i} \in \mathbb{N}_+^d} \bigotimes_{k=1}^{d} \Delta_{i_k},$$

where

$$\left\| \bigotimes_{k=1}^{d} \Delta_{i_k} \right\|_{\mathcal{F}_d \to \mathcal{G}_d} = \prod_{k=1}^{d} \|\Delta_{i_k}\|_{\mathcal{F}^{\gamma_{d,k}} \to \mathcal{G}} \leqslant b(d, \mathbf{i}) \|\mathcal{S}_d\|_{\mathcal{F}_d \to \mathcal{G}_d}$$

and

$$b(d, \mathbf{i}) = \prod_{k=1}^{d} b(k, i_k) \qquad \text{with} \quad b(k, i_k) = \begin{cases} 1 & \text{if} \quad i_k = 1, \\ \sqrt{\gamma_{d,k}}\, C D^{i_k - 1} & \text{if} \quad i_k \geqslant 2. \end{cases} \tag{40}$$

To finish the construction of the WTP algorithm we need to define the sets $P(n, d)$ of multi-indices $\mathbf{i}$. A good idea would be to select $\mathbf{i}$'s which correspond to the largest numbers $b(d, \mathbf{i})$. It turns out that we can do better by selecting the largest "weighted" $b(d, \mathbf{i})$. This is done as follows. Let $\xi_{d,1} = 1$, and let $\{\xi_{d,k}\}_{k \geqslant 2}$ be a sequence of positive numbers. Define

$$\xi(d, \mathbf{i}) = \prod_{k=1}^{d} \xi_{d,k}^{1 - \delta_{1, i_k}} \tag{41}$$

with $\delta_{1,k}$ being the Kronecker delta. That is, if $i_k = 1$ the corresponding component of $\xi(d, \mathbf{i})$ is 1 whereas for $i_k \geqslant 2$ we have $\xi_{d,k}$. Hence, if all i_k are at least 2 then $\xi(d, \mathbf{i}) = \prod_{k=1}^{d} \xi_{d,k}$.

Consider the sequence $\{b(d, \mathbf{i})/\xi(d, \mathbf{i})\}$. Note that

$$\frac{b(d, \mathbf{i})}{\xi(d, \mathbf{i})} = D^{|\mathbf{i}| - d} \prod_{k=1}^{d} \left(\frac{C \gamma_{d,k}^{1/2}}{\xi_{d,k}} \right)^{1 - \delta_{1, i_k}}.$$

From this it follows that $b(d, \mathbf{i})/\xi(d, \mathbf{i})$ goes to zeros as $|\mathbf{i}|$ goes to infinity. Therefore we can order the elements of $\{b(d, \mathbf{i})/\xi(d, \mathbf{i})\}$. That is, we define the sequence $\mathbf{i}_j$ such that

$$\frac{b(d, \mathbf{i}_1)}{\xi(d, \mathbf{i}_1)} \geqslant \frac{b(d, \mathbf{i}_2)}{\xi(d, \mathbf{i}_2)} \geqslant \cdots \geqslant \frac{b(d, \mathbf{i}_k)}{\xi(d, \mathbf{i}_k)}, \qquad \forall k.$$

The set $P(n, d)$ is now given as

$$P(n, d) = \{\mathbf{i}_1, \mathbf{i}_2, ..., \mathbf{i}_n\},$$

and the corresponding WTP algorithm takes the form

$$A_{n,d} = \sum_{\mathbf{i} \in P(n,d)} \bigotimes_{k=1}^{d} \Delta_{i_k}. \tag{42}$$

To guarantee that the WTP algorithm computes an ε-approximation, $\varepsilon \in (0, 1)$, we need to set

$$n(\varepsilon, d) = |\{\mathbf{i}\colon b(d, \mathbf{i}) > \xi(d, \mathbf{i})(\varepsilon/C_1(d, \eta))^{1/(1-\eta)}\}|, \tag{43}$$

where $\eta \in (0, 1)$ and

$$C_1(d, \eta) = \frac{1}{(1-D^s)^{1/s}} \prod_{k=2}^{d} \left(1 + (C^2 \gamma_{d,k})^{s\eta/2}\, \xi_{d,k}^{s(1-\eta)}\, \frac{D^{s\eta}}{1-D^{s\eta}}\right)^{1/s}$$

with

$$s = \begin{cases} 2 & \text{if } \mathscr{S} \text{ is a functional and } B_n \text{ are central} \\ & \text{algorithms that use nested information,} \\ 1 & \text{otherwise.} \end{cases} \tag{44}$$

Observe that $n(\varepsilon, d) \geqslant 1$. Indeed, for $\mathbf{1} = [1, 1, ..., 1]$ we have $b(d, \mathbf{1}) = \xi(d, \mathbf{1}) = 1$ and $C_1(d, \eta) \geqslant 1$. Therefore the multi-index $\mathbf{1}$ belongs to the set of (43) and $n(\varepsilon, d) \geqslant 1$.

Note that

$$P(n(\varepsilon, d), d) = \{\mathbf{i}\colon b(d, \mathbf{i}) > \xi(d, \mathbf{i})(\varepsilon/C_1(d, \eta))^{1/(1-\eta)}\}$$

and we can define

$$\mathscr{A}^*_{\varepsilon,d} = A^*_{n(\varepsilon,d),d} = \sum_{\mathbf{i} \in P(n(\varepsilon,d),d)} \bigotimes_{k=1}^{d} \Delta_{i_k}. \tag{45}$$

We are ready to estimate the error and cost of $\mathscr{A}^*_{\varepsilon,d}$.

THEOREM 3. *Let* $\eta \in (0, 1)$. *Then the WTP algorithm* $\mathscr{A}^*_{\varepsilon, d}$ *defined by* (45) *reduces the initial error by* ε,

$$e(\mathscr{A}^*_{\varepsilon, d}, \mathscr{S}_d) \leqslant \varepsilon \, \|\mathscr{S}_d\|_{\mathscr{F}_d \to \mathscr{G}_d}, \tag{46}$$

and its cost is bounded by

$$\mathrm{cost}(\mathscr{A}^*_{\varepsilon, d}) \leqslant C(d, \varepsilon) \left(\frac{1}{\varepsilon}\right)^{p/(1-n)}, \tag{47}$$

where

$$C(d, \varepsilon) = \frac{\begin{pmatrix} \alpha \prod_{k=2}^{d} (1 + \alpha(C^p \gamma_{d,k}^{p/2} / \xi_{d,k}^{p}) \, g(k, \varepsilon)) \\ \times (1 + C^{s\eta} \gamma_{d,k}^{s\eta/2} \xi_{d,k}^{s(1-\eta)} (D^{s\eta}/(1 - D^{s\eta})))^{p/(s(1-\eta))} \end{pmatrix}}{(1 - D^p)(1 - D^s)^{p/(s(1-\eta))}},$$

$$g(k, \varepsilon) = \left\lfloor \frac{\begin{pmatrix} \ln(C \gamma_{d,k}^{1/2} / (\xi_{d,k}(1 - D^s))^{1/(s(1-\eta))}) \prod_{i=2}^{k} (1 + C^{s\eta} \gamma_{d,i}^{s\eta/2} \xi_{d,i}^{s(1-\eta)} \\ \times (D^{s\eta}/(1 - D^{s\eta})))^{1/s(1-\eta)} \, (1/\varepsilon)^{1/(1-\eta)} \end{pmatrix}}{\ln D^{-1}} \right\rfloor_+,$$

with $\alpha = 1$ *if nested information is used by* $\{B_n\}$, *and* $\alpha = 2$, *otherwise.* (*By* $\lfloor x \rfloor_+$ *we mean* $\max\{0, x\}$.)

Proof. We first estimate the error. For a technical reason, we also need to consider the case $\varepsilon \geqslant 1$. For such ε, we formally set $\mathscr{A}^*_{\varepsilon, d} = 0$. Then $\|\mathscr{S}_d - \mathscr{A}^*_{\varepsilon, d}\|_{\mathscr{F}_d \to \mathscr{G}_d} / \|\mathscr{S}_d\|_{\mathscr{F}_d \to \mathscr{G}_d} = 1$. From this, and (22), (23), as well as (40), the worst case error of $\mathscr{A}^*_{\varepsilon, d}$ is bounded by

$$\frac{e^s(\mathscr{A}^*_{\varepsilon, d}, \mathscr{S}_d)}{\|\mathscr{S}_d\|^s_{\mathscr{F}_d \to \mathscr{G}_d}} \leqslant e^s(d, \varepsilon) := \begin{cases} \sum_{\mathbf{i}:\, b(d, \mathbf{i}) \leqslant \xi(d, \mathbf{i}) \, \alpha(d, \varepsilon)} b^s(d, \mathbf{i}) & \text{if} \quad \varepsilon < 1, \\ 1 & \text{if} \quad \varepsilon \geqslant 1, \end{cases}$$

where

$$\alpha(d, \varepsilon) = (\varepsilon / C_1(d, \eta))^{1/(1-\eta)}.$$

For $j = 1, 2, ..., d$, we also define

$$\bar{e}^s(j, \varepsilon) := \begin{cases} \sum_{\mathbf{i}:\, \bar{b}(j, \mathbf{i}) \leqslant \bar{\xi}(j, \mathbf{i}) \, \bar{\alpha}(j, \varepsilon)} \bar{b}^s(j, \mathbf{i}) & \text{if} \quad \varepsilon < 1 \\ 1 & \text{if} \quad \varepsilon \geqslant 1, \end{cases}$$

where

$$\bar{b}(j, \mathbf{i}) = \prod_{k=1}^{j} b(k, i_k),$$

$$\bar{\xi}(j, \mathbf{i}) = \prod_{k=1}^{j} \xi_{d,k}^{1-\delta_{1,i_k}},$$

$$\bar{\alpha}(j, \varepsilon) = \left(\frac{\varepsilon(1-D^s)^{1/s}}{\prod_{k=2}^{j}(1+C^{s\eta}\gamma_{d,k}^{s\eta/2}\xi_{d,k}^{s(1-\eta)}D^{s\eta}/(1-D^{s\eta})^{1/s}} \right)^{1/(1-\eta)}.$$

Clearly, $\bar{e}^s(d, \varepsilon) = e^s(d, s)$.

We show by induction on j that $\bar{e}^s(j, \varepsilon) \leqslant \varepsilon$ for all positive ε. For $j = 1$, it is trivially true for $\varepsilon \geqslant 1$. For $\varepsilon < 1$, denote

$$k^* = |\{ i\colon \bar{b}(1, i) > \bar{\alpha}(1, \varepsilon)\}|.$$

Observe that $k^* \geqslant 1$. Then we have

$$\bar{e}^s(1, \varepsilon) = \sum_{i=k^*+1}^{\infty} C^s\gamma_{d,1}^{s/2} D^{s(i-1)} = C^s\gamma_{d,1}^{s/2} \frac{D^{k^*s}}{1-D^s} \leqslant \left(\frac{\bar{\alpha}(1, \varepsilon)}{(1-D^s)^{1/s}} \right)^s \leqslant \varepsilon^s.$$

Hence, $\bar{e}(1, \varepsilon) \leqslant \varepsilon$, $\forall \varepsilon > 0$. For $j \geqslant 2$ assume inductively that $\bar{e}(j-1, \varepsilon) \leqslant \varepsilon$, $\forall \varepsilon > 0$. Note that we have

$$\bar{b}(j, \mathbf{i}) = \bar{b}(j-1, \mathbf{i})\; b(j, i_j),$$

$$\bar{\xi}(j, \mathbf{i}) = \bar{\xi}(j-1, \mathbf{i})\; \xi_{d,j}^{1-\delta_{1,i_j}}.$$

Then for $\varepsilon < 1$,

$$\bar{e}^s(j, \varepsilon) = \sum_{\ell=1}^{\infty} \sum_{\mathbf{i}\colon \bar{b}(j-1,\mathbf{i}) \leqslant \bar{\xi}(j-1,\mathbf{i})\, \bar{\beta}_\ell(j,\varepsilon)} b^s(j-1, \mathbf{i})\, b^s(j, \ell),$$

where

$$\bar{\beta}_\ell(j, \varepsilon) = \frac{\xi_{d,j}^{1-\delta_{1,\ell}}\bar{\alpha}(j, \varepsilon)}{b(j, \ell)} = \bar{\alpha}\left(j-1, \left(\frac{\xi_{d,j}^{1-\delta_{1,\ell}}}{b(j, \ell)} \right)^{1-\eta} a\varepsilon \right)$$

with $a = (1 + C^{s\eta} \gamma_{d,j}^{s\eta/2} \xi_{d,j}^{s(1-\eta)} D^{s\eta}/(1-D^{s\eta}))^{-1/s}$. Therefore

$$\begin{aligned}
\bar{e}^s(j,\varepsilon) &= \bar{e}^s(j-1, a\varepsilon) + \sum_{\ell=2}^{\infty} C^s \gamma_{d,j}^{s/2} D^{s(\ell-1)} \bar{e}^s \left(j-1, \left(\frac{\xi_{d,j}}{b(j,\ell)} \right)^{1-\eta} a\varepsilon \right) \\
&\leqslant a^s \varepsilon^s \left(1 + C^s \gamma_{d,j}^{s/2} \sum_{\ell=2}^{\infty} D^{s(\ell-1)} \left(\frac{\xi_{d,j}}{b(j,\ell)} \right)^{s(1-\eta)} \right) \\
&= a^s \varepsilon^s \left(1 + C^{s\eta} \gamma_{d,j}^{s\eta/2} \xi_{d,j}^{s(1-n)} \sum_{\ell-2}^{\infty} D^{s\eta(\ell-1)} \right) \\
&= a^s \varepsilon^s \left(1 + C^{s\eta} \gamma_{d,j}^{s\eta/2} \xi_{d,j}^{s(1-\eta)} \frac{D^{s\eta}}{1-D^{s\eta}} \right) = \varepsilon^s.
\end{aligned}$$

This proves (46).

We now analyze the cost of $\mathscr{A}^*_{\varepsilon,d}$. Let $\bar{m}_1 = 1$, and for $i \geqslant 2$, let $\bar{m}_i = m_i$ for nested information, and $\bar{m}_i = 2m_i$ otherwise. Then $\bar{m}_i$ is an upper bound on the cost of Δ_i. The cost of $\bigotimes_{k=1}^d \Delta_{i_k}$ is thus bounded by $\prod_{k=1}^d \bar{m}_{i_k}$. Therefore

$$\operatorname{cost}(\mathscr{A}^*_{\varepsilon,d}) \leqslant c(d,\varepsilon) := \begin{cases} \sum_{\mathbf{i}:\, b(d,\mathbf{i}) > \xi(d,\mathbf{i})\, \alpha(d,\varepsilon)} \prod_{k=1}^d \bar{m}_{i_k}, & \text{if} \quad \varepsilon < 1, \\ 0 & \text{if} \quad \varepsilon \geqslant 1, \end{cases}$$

where $\alpha(d,\varepsilon)$ is defined as before.

For $j = 1, 2, ..., d$, we define

$$\bar{c}(j,\varepsilon) = \begin{cases} \sum_{\mathbf{i}:\, \bar{b}(j,\mathbf{i}) > \bar{\xi}(j,\mathbf{i})\, \bar{\alpha}(j,\varepsilon)} \prod_{k=1}^j \bar{m}_{i_k}, & \text{if} \quad \varepsilon < 1, \\ 0 & \text{if} \quad \varepsilon \geqslant 1, \end{cases}$$

where $\bar{b}$, $\bar{\xi}$, and $\bar{\alpha}$ are defined as before. Clearly, $\bar{c}(d,\varepsilon) = c(d,\varepsilon)$.

For $d = 1$, we use k^* defined before and for $\varepsilon < 1$ we have $\bar{c}(1,\varepsilon) = \sum_{i=1}^{k^*} \bar{m}_i$. Due to (36),

$$\bar{c}(1,\varepsilon) \leqslant 1 + \alpha \sum_{i=2}^{k^*} D^{-(i-1)p} \leqslant \alpha \frac{D^{-k^*p} - 1}{D^{-p} - 1} \leqslant \alpha \frac{D^p}{1 - D^p} D^{-k^*p}.$$

Since $D^{(k^*-1)} > \bar{\alpha}(1,\varepsilon)$, we conclude

$$\bar{c}(1,\varepsilon) \leqslant \frac{\alpha}{1-D^p} \bar{\alpha}(1-\varepsilon)^{-p} =: C_1 \left(\frac{1}{\varepsilon} \right)^{p/(1-\eta)}$$

with

$$C_1 = \frac{\alpha}{(1-D^p)(1-D^s)^{p/(s(1-\eta))}}.$$

Hence, $\bar{c}(1,\varepsilon) \leqslant C_1 \varepsilon^{-p/(1-\eta)}$, $\forall \varepsilon > 0$.

For $j \geqslant 2$, assume inductively that $\bar{c}(j-1,\varepsilon) \leqslant C_{j-1}\varepsilon^{-p/(1-\eta)}$ for some C_{j-1} and all $\varepsilon > 0$. Then

$$\begin{aligned}
\bar{c}(j,\varepsilon) &= \sum_{\ell=1}^{\infty} \sum_{\mathbf{i}:\, \bar{b}(j-1,\mathbf{i}) > \bar{\xi}(j-1,\mathbf{i})\, \bar{\beta}_\ell(j,\varepsilon)} \bar{m}_\ell \left(\prod_{k=1}^{j-1} \bar{m}_{i_k}\right), \\
&= \bar{c}(j-1, a\varepsilon) + \sum_{\ell=2}^{g^*(j,\varepsilon)} \bar{m}_\ell \bar{c}\left(j-1, \left(\frac{\xi_{d,j}}{C\gamma_{d,j}^{1/2} D^{\ell-1}}\right)^{1-\eta} a\varepsilon\right) \\
&= C_{j-1}\left(\frac{1}{a\varepsilon}\right)^{p/(1-\eta)} \left(1 + \sum_{\ell=2}^{g^*(j,\varepsilon)} \bar{m}_\ell D^{(\ell-1)p} \frac{C^p \gamma_{d,j}^{p/2}}{\xi_{d,j}^p}\right),
\end{aligned}$$

where $\bar{\beta}_\ell(j,\varepsilon)$ and a are defined as before. Here, $g^*(j,\varepsilon)$ is defined as the largest index i for which the second sum has at least one term. That is, $g^*(j,\varepsilon)$ is the largest integer solution of the equation

$$C\gamma_{d,j}^{1/2} D^{i-1} > \xi_{d,j}\alpha(j,\varepsilon)$$

and

$$g^*(j,\varepsilon) = 1 + \left\lfloor \frac{\ln(C\gamma_{d,j}^{1/2}/(\xi(d,j)\,\bar{\alpha}(j,\varepsilon)))}{\ln D^{-1}} \right\rfloor_+ .$$

Note that $g^*(j,\varepsilon) = 1 + g(j,\varepsilon)$, where g is defined in Theorem 3.

From (36) we conclude

$$\bar{c}(j,\varepsilon) \leqslant C_{j-1}\left(\frac{1}{a\varepsilon}\right)^{p/(1-\eta)} \left(1 + \alpha \frac{C^p \gamma_{d,j}^{p/2}}{\xi_{d,j}^p} g(j,\varepsilon)\right) =: C_j \varepsilon^{-p}.$$

Hence,

$$C_j = C_1 \prod_{k=2}^{j} \left(1 + \alpha \frac{C^p \gamma_{d,k}^{p/2}}{\xi_{d,k}^p} g(k,\varepsilon)\right) \left(1 + C^{s\eta}\gamma_{d,k}^{s\eta/2} \xi_{d,k}^{s(1-\eta)} \frac{D^{s\eta}}{1-D^{s\eta}}\right)^{p/(s(1-\eta))}.$$

Since $C_d = C(d, \varepsilon)$, we proved that $c(d, \varepsilon) = \bar{c}(d, \varepsilon) \leqslant C(d, \varepsilon)\varepsilon^{-p}$, as claimed. This completes the proof. ∎

Theorem 3 presents an explicit bound on the cost of the WTP algorithm that computes an ε-approximation. We first comment on the cost bound (47) for arbitrary weights $\gamma_{d,k}$ and $\xi_{d,k}$ as a function of ε. Observe that the function g depends logarithmically on ε^{-1}. This implies that $C(d, \varepsilon)$ depends on $(\ln \varepsilon^{-1})^{d-1}$. Furthermore, we can claim such a dependence for an arbitrary $\eta \in (0, 1)$. This proves the following corollary.

COROLLARY 1. *For every positive δ there exists a positive $c(d, \delta)$ such that the cost of the WTP algorithm $\mathscr{A}^*_{\varepsilon,d}$ defined by* (45) *is bounded by*

$$\operatorname{cost}(\mathscr{A}^*_{\varepsilon,d}) \leqslant c(d, \delta) \left(\frac{1}{\varepsilon}\right)^{p+\delta}. \tag{48}$$

The exponent p in Corollary 1 satisfies (36) which measures the behavior for the one dimensional case. Clearly, $p \geqslant 2p_\lambda$ with p_λ defined in the previous section as the sum-exponent of the squares of the singular values of the operator $\mathscr{S}$. Hence, Corollary 1 states that we essentially preserve the same dependence on ε^{-1} for all dimensions d, however, the factor $c(d, \delta)$ may depend on d.

We now address the dependence of the cost bound (47) on d. Obviously, this crucially depends on the weights $\gamma_{d,k}$ and $\xi_{d,k}$.

THEOREM 4. *Assume that*

$$\sup_d \left\{ \sum_{k=2}^{d} \left(\frac{\gamma_{d,k}^{1/2}}{\xi_{d,k}}\right)^{p}, \sum_{k=2}^{d} \xi_{d,k}^{s} \left(\frac{\gamma_{d,k}^{1/2}}{\xi_{d,k}}\right)^{s\eta} \right\} \leqslant M < \infty. \tag{49}$$

Then the WTP algorithm $\mathscr{A}^*_{\varepsilon,d}$ defined by (45) *is strongly polynomial and its ε-exponent $p^*(\{\mathscr{A}^*_{\varepsilon,d}\})$ is bounded by*

$$p^*(\{\mathscr{A}^*_{\varepsilon,d}\}) \leqslant \frac{p}{1-\eta}. \tag{50}$$

That is, for every positive δ there exists c_δ such that

$$\operatorname{cost}(\mathscr{A}^*_{\varepsilon,d}) \leqslant c_\delta \left(\frac{1}{\varepsilon}\right)^{p/(1-\eta)+\delta}, \qquad \forall \varepsilon \in (0, 1) \text{ and } \forall d. \tag{51}$$

Proof. Let $\alpha_{d,k} = (\gamma_{d,k}^{1/2}/\xi_{d,k})^p$ and $\beta_{d,k} = \xi_{d,k}^s(\gamma_{d,k}^{1/2}/\xi_{d,k})^{s\eta}$. In what follows, we use c_i to denote positive constants that do not depend on ε and d. We first estimate $g(k,\varepsilon)$. From Theorem 3 we have

$$g(k,\varepsilon) \leqslant c_1 \left(1 + \ln \varepsilon^{-1} + \sum_{k=2}^{d} \ln(1 + c_2 \beta_{d,k}) \right).$$

Using $\ln(1+x) \leqslant x$ and the fact that $\sum_{k=2}^d \beta_{d,k} \leqslant M$ we easily conclude that

$$g(k,\varepsilon) \leqslant c_1(c_3 + \ln \varepsilon^{-1}).$$

We now estimate $C(d,\varepsilon)$. We have

$$C(d,\varepsilon) \leqslant c_4 \prod_{k=2}^{d} (1 + c_5 \alpha_{d,k}(1 + \ln \varepsilon^{-1})) \prod_{k=2}^{d} (1 + c_6\, \beta_{d,k})^{p/(s(1-\eta))}.$$

Observe that

$$\begin{aligned} \prod_{k=2}^{d} (1 + c_6 \beta_{d,k})^{p/(s(1-\eta))} &= \exp\left(\frac{p}{s(1-\eta)} \ln(1 + c_6 \beta_{d,k}) \right) \\ &\leqslant \exp\left(c_7 \sum_{k=2}^{d} \beta_{d,k} \right) \leqslant \exp(c_7 M) = c_8. \end{aligned}$$

We now estimate the first product in the estimate of $C(d,\varepsilon)$. Without loss of generality, we may assume that $\alpha_{d,k}$ are nonincreasing, i.e., $\alpha_{d,j} \geqslant \alpha_{d,j+1}$, $\forall j$. Since $\sum_{k=2}^d \alpha_{d,k} \leqslant M$ for all d, there exists $k^* = k^*(\delta)$ such that $\sum_{k=k^*+1}^d \alpha_{d,k} \leqslant \delta/(2c_5)$. Then

$$\begin{aligned} &\prod_{k=k^*+1}^{d} (1 + c_5 \alpha_{d,k}(1 + \ln \varepsilon^{-1})) \\ &\quad = \exp\left(\sum_{k=k^*+1}^{d} \ln(1 + c_5 \alpha_{d,k}(1 + \ln \varepsilon^{-1})) \right) \\ &\quad \leqslant \exp\left(c_5 \sum_{k=k^*+1}^{d} \alpha_{d,k}(1 + \ln \varepsilon^{-1}) \right) \leqslant \varepsilon^{-\delta/2} e^{\delta/2}. \end{aligned}$$

Clearly, we also have

$$\prod_{k=1}^{k^*} (1 + c_5 \alpha_{d,k}(1 + \ln \varepsilon^{-1})) \leqslant c_9 \varepsilon^{-\delta/2}.$$

This proves that

$$C(d, \varepsilon) \leqslant c_8 c_9 \varepsilon^{-\delta}.$$

The proof is complete by applying (47).

Theorem 4 states the sufficient condition (49) under which the WTP algorithm is strongly polynomial. This condition is expressed in terms of the weight sequences $\gamma_{d,k}$ and $\xi_{d,k}$ as well as the parameter η. The sequence $\xi_{d,k}$ and the parameter η are at out disposal. It is therefore natural to define them such that the WTP algorithm is strongly polynomial and its ε-exponent is minimized.

To find out when the WTP algorithm is strongly polynomial, we recall Theorem 1 where it is proved that a necessary condition for strong tractability for all operators of rank at least 2 is that the sum-exponent p_γ is finite. Hence, we assume that $p_\gamma < \infty$. As we shall see in the next theorem, we will need to assume that p_γ is sufficiently small. The sum-exponent p_γ is defined by (27). For simplicity we assume that

$$\sup_d \sum_{k=2}^{d} \gamma_{d,k}^{p_\gamma} < \infty. \tag{52}$$

If (52) does not hold then it is enough to increase p_γ by an arbitrarily small number.

We will be using the WTP algorithm with[4]

$$\begin{aligned} \xi_{d,k} &= \gamma_{d,k}^{(1-2p_\gamma/p)/2}, \qquad k = 2, 3, \ldots, \\ \eta &= \begin{cases} 1 - \dfrac{p}{2p_\gamma}\left(1 - \dfrac{2p_\gamma}{s}\right) & \text{if} \quad p < 2p_\gamma/(1 - 2p_\gamma/s), \\ \delta & \text{otherwise,} \end{cases} \end{aligned} \tag{53}$$

where δ is an arbitrary positive number.

THEOREM 5. *Assume that*

$$p_\gamma < \frac{s}{2}. \tag{54}$$

[4] It can be proven that among sequences $\xi_{d,k} = \Theta(\gamma_{d,k}^{\beta})$ the choice $\beta = (1 - 2p_\gamma/p)/2$, as in (53), is optimal.

Then the WTP algorithm $\mathscr{A}^*_{\varepsilon, d}$ *defined by* (45) *and* (53) *is strongly polynomial and its ε-exponent* $p^*(\{\mathscr{A}^*_{\varepsilon, d}\})$ *is bounded by*

$$p^*(\{\mathscr{A}^*_{\varepsilon, d}\}) \leqslant \max\left\{\frac{p}{1-\delta}, \frac{2p_\gamma}{1-2p_\gamma/s}\right\}. \tag{55}$$

Proof. For $\xi_{d,k}$ given by (53) we have

$$\alpha_{d,k} = \left(\frac{\gamma_{d,k}^{1/2}}{\xi_{d,k}}\right)^p = \gamma_{d,k}^{p_\gamma}$$

and the series $\sum_{k=1}^d \alpha_{d,k}$ is uniformly bounded in d. Hence, the first part of (49) holds.

Assume now that $p < 2p_\gamma/(1-2p_\gamma/s)$. Then

$$\beta_{d,k} = \xi_{d,k}^s \left(\frac{\gamma_{d,k}^{1/2}}{\xi_{d,k}}\right)^{s\eta} = \Theta(\gamma_{d,k}^{s(1-2p_\gamma/p)/2 + s\eta p_\gamma/p}) = \Theta(\gamma_{d,k}^{p_\gamma}),$$

due to the definition of η in (53). Hence, the series $\sum_{k=2}^d \beta_{d,k}$ is also uniformly bounded in d. Therefore (49) holds, and Theorem 4 states that the WTP is strongly polynomial with the ε-exponent at most

$$\frac{p}{1-\eta} = \frac{2p_\gamma}{1-2p_\gamma/s},$$

as claimed in (55).

Assume that $p \geqslant 2p_\gamma/(1-2p_\gamma/s)$. Then $1-2p_\gamma/p \geqslant 2p_\gamma/s$ and

$$\beta_{d,k} = \Theta(\gamma_{d,k}^a),$$

where

$$a = s(1-2p_\gamma/p)/2 + s\eta p_\gamma/p > p_\gamma, \qquad \forall \delta > 0.$$

Therefore $\sum_{k=2}^d \beta_{d,k}$ is uniformly bounded in d, (49) holds, and Theorem 4 completes the proof. ∎

We now comment on the assumption (54) that $p_\gamma < s/2$. Recall that s is defined by (44) and $s = 2$ for the functional case, and $s = 1$ for the general case.

For the functional case, we thus assume that $p_\gamma < 1$. It seems to us that a slightly relaxed inequality $p_\gamma \leqslant 1$ is needed. Indeed, observe that in the analysis of the WTP algorithm we only use its global error properties represented by the parameters p, D and the weights $\gamma_{d,k}$. The structure of the spaces $\mathscr{F}_d$ did not play any role. Hence, in particular, we may use the weighted Sobolev spaces and multivariate integration as in [10]. In [10] it is proved that $p_\gamma \leqslant 1$ is a necessary and sufficient condition for quasi-Monte Carlo algorithms to be strongly polynomial. Although quasi-Monte Carlo algorithms are not optimal (central) algorithms it is widely believed that they are optimal modulo a multiplicative factor. If so, the condition $p_\gamma \leqslant 1$ is needed for strong tractability. It is, therefore, maybe not surprising that we need to assume that $p_\gamma < 1$ for the WTP algorithm to be strongly polynomial.

The discrepancy between the two inequalities $p_\gamma < 1$ and $p_\gamma \leqslant 1$ is probably caused by the presence of the parameter η in (43) which makes the analysis of the WTP algorithm much easier. We believe that a more refined analysis may allow to take $\eta = 0$ with the same conclusion that $p_\gamma \leqslant 1$ implies that the WTP algorithm is strongly polynomial.

For the general case, we need to assume that $p_\gamma < 1/2$. We do not know if this condition may be relaxed in general. If the condition $p_\gamma < 1/2$ cannot be relaxed then this may indicate a shortcoming of the WTP algorithm defined by (45) and (53).

We now discuss the bound (55) on the ε-exponent of the WTP algorithm. Sine δ can be arbitrarily small, the bound (55) roughly states that we can achieve the ε-exponent

$$\max\left\{p, \frac{2p_\gamma}{1-2p_\gamma/s}\right\}.$$

As already remarked, the one-dimensional exponent p must be at least $2p_\lambda$. Hence,

$$p^*(\{\mathscr{A}^*_{\varepsilon,d}\}) \geqslant p^* = 2\max\{p_\gamma, p_\lambda\}$$

and this agrees with Theorem 1 which states that p^* is the ε-exponent of strong tractability of $\{\mathscr{S}_d\}$ for all $\mathscr{S}$ of rank at least two in the class Λ^{all}.

Observe that for small p_γ, the ε-exponent of the WTP algorithm may be arbitrarily close to p. If, in turn, $p = 2p_\lambda$, then the ε-exponent of the WTP algorithm may be arbitrarily close to p^*. We summarize this property in the corollary.

COROLLARY 2. *If* $p = 2p_\lambda$ *and*

$$p_\gamma < \min\left\{\frac{s}{2}, \frac{p_\lambda}{1+2p_\lambda/s}\right\}$$

then the ε*-exponent of the WTP algorithm defined by* (45) *and* (53) *with small* δ *is*

$$p^*(\{\mathscr{A}^*_{\varepsilon,d}\}) \approx 2p_\lambda$$

which is minimal even in the class Λ^{all}.

6. APPLICATIONS

In this section, we illustrate the general results by applying them to the integration and function approximation problems discussed in the Example from previous sections. We will also use two different families of sequences γ. Recall that in the Example, we consider functions of regularity $r = 1$. At the end of the section, we will briefly show how these results extend for the class of functions with higher regularity r. Although $r = 1$ is a special case of $r \geqslant 1$, we present the case $r = 1$ separately since it is simpler and the algorithms and their error formulas are more explicit.

Recall that the Example deals with functions from the following class $\mathscr{F}$

$$\mathscr{F} = \{f\colon [0,1] \to \mathbb{R} : f(0) = 0,\ f \text{ is abs. cont.},\ \|f'\|_2 \leqslant 1\}$$

with $\|g\|_2^2 = \int_0^1 g^2(x)\,dx$. We also have $S(f) = \int_0^1 f(x)\,dx$ for the integration problem, and $S(f) = f \in \mathscr{G} = L_2([0,1])$ for the approximation problem. Since $h^* \equiv 1$, we have $\|Sh^*\|_2 = 1$.

For the approximation problem, it is well known that the eigenvalues of $\mathscr{W}$ given by (1) satisfy $\lambda_i = \Theta(i^{-2})$. Hence, the sum-exponent, see (27), is

$$p_\lambda = 0.5. \tag{56}$$

We begin with the approximation problem. Consider $m_1 = 1$ and $m_i = 1 + 2^{i-2}$ for $i \geqslant 2$, and the following algorithms $B_{m_i - 1}$. We have $B_0 = 0$ and with

$$B_{m_i-1}(f)(x) = f(x_{j,i}) \qquad \text{if } x \in I_{j,i}$$

$$x_{j,i} = \frac{j}{2^{i-2}} \quad \text{and} \quad I_{j,i} = [x_{j,i} - 2^{-i+1}, x_{j,i} + 2^{-i+1}),\ \forall 0 \leqslant j \leqslant 2^{i-2},\ \forall i \geqslant 2.$$

Observe that B_{m_i-1} uses function values at $x_{j,i}$ for $j=1, 2, ..., 2^{i-2}=m_i-1$ since $f(x_{0,i})=f(0)$ is always zero.

We now estimate $\|B_{m_i-1}-B_{m_{i-1}-1}\|_{\mathscr{F}\to\mathscr{G}}$ for $i\geqslant 2$. For $i=2$, we have

$$\|B_{m_2-1}(f)-B_{m_1-1}(f)\|_{\mathscr{G}}^2=\|B_1(f)\|_{\mathscr{G}}^2=\int_{1/2}^1 f^2(1)\,dx$$

$$=\frac{1}{2}f^2(1)=\frac{1}{2}\left(\int_0^1 f'(x)\,dx\right)^2\leqslant\frac{1}{2}\|f'\|_2^2\leqslant\frac{1}{2}.$$

The last bound is sharp since we have equality for $f(x)=x$. Therefore,

$$\|B_{m_2-1}(f)-B_{m_1-1}(f)\|_{\mathscr{F}\to\mathscr{G}}^2=1/2.$$

Let now $i\geqslant 3$. Since $B_{m_i-1}(f)(x)=B_{m_{i-1}-1}(f)(x)$ for $x\in I_{j,i}$ with even values of j, we have

$$c_i(f):=\|B_{m_i-1}(f)-B_{m_{i-1}-1}(f)\|_{\mathscr{G}}^2$$

$$=\sum_{k=1}^{2^{i-3}}\int_{I_{2k-1,i}}|f(x_{2k-1,i})-B_{m_{i-1}-1}(f)(x)|^2\,dx.$$

Note that $B_{m_{i-1}-1}(f)(x)$ equals $f(x_{k-1,i-1})$ or $f(x_{k,i-1})$ depending on whether x is in the 1st or 2nd half of $I_{2k-1,i}$. Hence,

$$c_i(f)=\sum_{k=1}^{2^{i-3}}\left(\int_{x_{2k-1,i}-2^{-i+1}}^{x_{2k-1,i}}|f(x_{2k-1,i})-f(x_{k-1,i-1})|^2\,dx\right.$$

$$\left.+\int_{x_{2k-1,i}}^{x_{2k-1,i}+2^{-i+1}}|f(x_{2k-1,i})-f(x_{k,i-1})|^2\,dx\right)$$

$$=2^{-i+1}\sum_{k=1}^{2^{i-3}}(|f(x_{2k-1,i})-f(x_{k-1,i-1})|^2+|f(x_{k,i-1})-f(x_{2k-1,i})|^2)$$

$$=2^{-i+1}\sum_{k=1}^{2^{i-3}}\left(\left|\int_{x_{k-1,i-1}}^{x_{2k-1,i}}f'(t)\,dt\right|^2+\left|\int_{x_{2k-1,i}}^{x_{k,i-1}}f'(t)\,dt\right|^2\right)$$

$$\leqslant 2^{-i+1}\sum_{k=1}^{2^{i-3}}\left(2^{-i+2}\int_{x_{k-1,i-1}}^{x_{2k-1,i}}|f'(t)|^2\,dt+2^{-i+2}\int_{x_{2k-1,i}}^{x_{k,i-1}}|f'(t)|^2\,dt\right)$$

$$=2^{-2(i-1)}2\,\|f'\|_2^2\leqslant 2^{-2(i-1)}2.$$

Actually, for $f(x)=x$ we have equality and therefore

$$\|B_{m_i-1}-B_{m_{i-1}-1}\|_{\mathcal{F}\to\mathcal{G}}=2^{-i+1}\sqrt{2}, \qquad \forall i\geqslant 2.$$

Hence, for the approximation problem, (35) is satisfied with $C=\sqrt{2}$ and $D=1/2$, whereas (36) holds with $p=1$.

We now consider the integration problem. Let m_i, $x_{j,i}$, and $I_{j,i}$ be as before. The corresponding algorithms $B_{m_i-1}(f)$ are obtained as the integrals of the previous approximation algorithms. That is, $B_{m_1-1}(f)=0$ and

$$B_{m_i-1}(f)=\frac{1}{2^{i-2}}\left(\frac{f(1)}{2}+\sum_{j=1}^{2^{i-2}-1} f(x_{j,i})\right) \qquad \forall i\geqslant 2.$$

Hence, they are equal to the trapezoid rules (recall that $f(0)=0$). This also means that they are equal to integrals of piecewise linear functions interpolating f at 0, $1/2^{i-2}$, $2/2^{i-2}$, ..., 1, and thus are central.

Note that $|B_{m_2-1}(f)|=|f(1)/2|\leqslant 1/2$. As before, we have equality for $f(x)=x$, i.e., $\|B_{m_2}-1\|_{\mathcal{F}\to\mathcal{G}}=1/2$. For $i\geqslant 3$ we have

$$\begin{aligned}
&B_{m_i-1}(f)-B_{m_{i-1}-1}(f)\\
&\quad=2^{-i+2}\sum_{k=1}^{2^{i-3}} f(x_{2k-1,i})-2^{-i+2}\sum_{k=1}^{2^{i-3}-1} f(x_{2k,i})-2^{-i+1}f(1)\\
&\quad=2^{-i+1}\sum_{j=1}^{2^{i-2}} (-1)^{j+1}\,(f(x_{j,i})-f(x_{j-1,i})).
\end{aligned}$$

Since

$$|f(x_{j,i})-f(x_{j-1,i})|=\left|\int_{x_{j-1,i}}^{x_{j.i}} f'(t)\,dt\right|\leqslant 2^{-(i-2)/2}\left(\int_{x_{j-1,i}}^{x_{j.i}} (f'(t))^2\,dt\right)^{1/2}$$

and since $\sum_{k=1}^{n} a_k\leqslant\sqrt{n}\,\sqrt{\sum_{k=1}^{n} a_k^2}$, we conclude that $|B_{m_i-1}(f)-B_{m_{i-1}-1}|\leqslant 2^{-i+1}\,\|f'\|_2\leqslant 2^{-i+1}$. Thus,

$$\|B_{m_i-1}-B_{m_{i-1}-1}\|_{\mathcal{F}\to\mathcal{G}}\leqslant 2^{-i+1}, \qquad \forall i\geqslant 2.$$

As with the approximation problem, one can prove that we have equality above.

Hence, (35) and (36) now hold with $C=1$, $D=1/2$, and $p=1$.

We now discuss two families of sequences γ.

Family 1. Let $\gamma_{d,k}=\gamma_k$ be independent of d and

$$\gamma_k = \Theta(k^{-z}), \qquad \forall k \geqslant 1 \tag{57}$$

for some parameter $z>0$. Of course, the sum-exponent equals $p_\gamma = 1/z$. Therefore, for $\Lambda = \Lambda^{\mathrm{all}}$, the ε-exponent of strong tractability for the function approximation problem equals

$$p^*(\Lambda^{\mathrm{all}}) = \max\left\{1, \frac{2}{z}\right\}.$$

Note that the approximation problem is strongly tractable for every $z>0$; however, the exponent $p^*(\Lambda^{\mathrm{all}})$ converges to infinity with $z \to 0$.

Consider now $\Lambda = \Lambda^{\mathrm{std}}$. For the function approximation problem we need to assume additionally that $z>2$ since otherwise the results of Section 5 are not applicable. For the same reason, we need to assume $z>1$ for the integration problem. Then, from Theorem 5, we conclude that

$$p^*(\Lambda^{\mathrm{std}}) \leqslant p^*(\mathscr{A}_{\varepsilon,d}) \leqslant \begin{cases} \max\left\{\dfrac{1}{1-\delta}, \dfrac{2}{z-2}\right\} & \text{for approximation,} \\[2ex] \max\left\{\dfrac{1}{1-\delta}, \dfrac{2}{z-1}\right\} & \text{for integration.} \end{cases} \tag{58}$$

We do not know the ε-exponent $p^*(\Lambda^{\mathrm{std}})$ for all values of z. In particular, we do not know if the integration and approximation problems remain tractable when $z \leqslant 2$ or $z \leqslant 1$, respectively.

For $z \geqslant 4$ for the approximation problem, and $z \geqslant 3$ for the integration problem the bound (58) is sharp and $p^*(\Lambda^{\mathrm{std}}) \approx p^*(\mathscr{A}_{\varepsilon,d}) \approx 1$ since δ can be arbitrarily small. For other values of z, we do not know if the upper bounds (58) are sharp.

Family 2. We now take a special class of sequences γ suggested by a problem from economics. There is a well known family of functions studied by economists, called the Cobb Douglas family, [9], which consists of the functions of the form

$$f(t_1, t_2, \ldots, t_d) = \prod_{k=1}^{d} (t_k + a_k)^{\alpha_k},$$

where the a_k's are nonnegative and $\alpha_k \geqslant 0$ with $\sum_{k=1}^{d} \alpha_k = 1$. We now estimate the norm of f in the space $\mathscr{F}_d$. We have

$$\int_{[0,1]^{|u|}} \left(\prod_{k \in u} \frac{\partial}{\partial x_k} f_u(x) \right)^2 \prod_{k \in u} \frac{dx_k}{\gamma_{d,k}} = \prod_{k \in u} \frac{\alpha_k^2}{\gamma_{d,k}} \int_0^1 (t + a_k)^{2(\alpha_k - 1)}\, dt \prod_{k \notin u} a_k^{2\alpha_k}.$$

Since $2(\alpha_k - 1) \leqslant 0$ and $(t + a_k)^{2(a_k - 1)} \leqslant a_k^{2(a_k - 1)}$ we have

$$\int_{[0,1]^{|u|}} \left(\prod_{k \in u} \frac{\partial}{\partial x_k} f_u(x) \right)^2 \prod_{k \in u} \frac{dx_k}{\gamma_{d,k}} \leqslant \prod_{k \in u} \frac{\alpha_k^2}{a_k^2 \gamma_{d,k}} \prod_{k=1}^{d} a_k^{2\alpha_k}.$$

Hence,

$$\|f\|_{\mathscr{F}_d}^2 \leqslant \prod_{k=1}^{d} a_k^{2\alpha_k} \left(1 + \sum_{u \neq \varnothing} \prod_{k \in u} \frac{\alpha_k^2}{a_k^2 \gamma_{d,k}} \right).$$

It is known, see formula (40) in [10], that

$$\sum_{u \neq \varnothing} \prod_{k \in u} \gamma_k = \sum_{j=1}^{d} \gamma_j \prod_{k=j+1}^{d} (1 + \gamma_k) \leqslant \left(\sum_{j=k}^{d} \gamma_j \right) \exp\left(\sum_{j=1}^{d} \gamma_j \right).$$

Therefore,

$$\|f\|_{\mathscr{F}_d}^2 \leqslant \prod_{k=1}^{d} a_k^{2\alpha_k} \left(1 + \sum_{j=1}^{d} \frac{\alpha_j^2}{a_j^2 \gamma_{d,j}} \prod_{k=j+1}^{d} \left(1 + \frac{\alpha_k^2}{a_k^2 \gamma_{d,k}} \right) \right). \tag{59}$$

We now consider, as an example, two choices of $\gamma_{d,k}$ for which $\|f\|_{\mathscr{F}_d}$ is not too large.

(i) Let $a = \min_j a_j$ and $b = \max_j a_j$. Since $\sum_j \alpha_j = 1$ we have

$$\|f\|_{\mathscr{F}_d}^2 \leqslant b^2 + \frac{b^2}{a^2} \left(\sum_{j=1}^{d} \frac{\alpha_j^2}{\gamma_{d,j}} \right) \exp\left(\frac{1}{a^2} \sum_{j=1}^{d} \frac{\alpha_j^2}{\gamma_{d,j}} \right).$$

Setting $\gamma_{d,j} = \alpha_j$ we conclude that

$$\|f\|_{\mathscr{F}_d}^2 \leqslant b^2 + \frac{b^2}{a^2} \exp\left(\frac{1}{a_2} \right).$$

Hence, $\|f\|_{\mathscr{F}_d}$ has a bound which is not too large if a is not too small, and b not too large.

(ii) Assume for simplicity that $a_j = a$ for all j. Setting now $\gamma_{d,j} = \alpha_j^2$ we conclude from (59) that

$$\|f\|_{\mathscr{F}_d}^2 \leqslant a^2 \left(1 + \frac{1}{a^2} \sum_{j=1}^{d} (1 + a^{-2})^{d-j}\right) = a^2 (1 + a^{-2})^d.$$

For $a = \sqrt{d}$, say, we have

$$\|f\|_{\mathscr{F}_d}^2 \leqslant d(1 + d^{-1})^d \leqslant de.$$

Hence, $\|f\|_{\mathscr{F}_d} \leqslant \sqrt{de}$ which is not too large for reasonable d.

This example from economics suggests to consider the sequence of weights such that

$$\sum_{k=1}^{d} \gamma_{d,k}^q = 1, \qquad \forall d \geqslant 1,$$

for some positive q. Note that $q = 1$ and $q = 1/2$ were used in the example above. For instance, $\gamma_{d,k} = 1/d^{1/q}$ (for all k) or $\gamma_{d,k} = \delta_{d,k}$ are two extreme examples of such sequences. Depending on the sequence, the sum-exponent can be any number between 0 and q,

$$0 \leqslant p_\gamma \leqslant q.$$

It is q when, e.g., $\gamma_{d,k} = 1/d^{1/q}$, and is zero when, e.g., $\gamma_{d,k} = \delta_{d,k}$.

From Theorem 1 we conclude that for $\Lambda = \Lambda^{\text{all}}$, the function approximation problem is strongly tractable regardless of the sequence γ. That is,

$$p^*(\Lambda^{\text{all}}) = \max\{1, 2p_\gamma\}.$$

In particular, the exponent never exceeds the value of $2q$.

Consider now $\Lambda = \Lambda^{\text{std}}$. We do not know if the function approximation problem is tractable when $p_\gamma \geqslant 1/2$. Otherwise, when $p_\gamma < 1/2$, we have strong tractability with

$$p^*(\Lambda^{\text{std}}) \leqslant p^*(\mathscr{A}_{\varepsilon,d}) \leqslant \max\left\{\frac{1}{1-\delta}, \frac{2p_\gamma}{1-2p_\gamma}\right\}.$$

The integration problem is strongly tractable for every sequence γ with $p_\gamma < 1$ and

$$p^*(\Lambda^{\text{std}}) \leqslant p^*(\mathscr{A}_{\varepsilon,d}) \leqslant \max\left\{\frac{1}{1-\delta}, \frac{2p_\gamma}{1-p_\gamma}\right\}.$$

As before, the exponent of $\mathcal{A}_{\varepsilon,d}$ is optimal when $p_\gamma \leqslant 1/4$ for the approximation problem and $p_\gamma \leqslant 1/3$ for integration problem since then $p^*(\Lambda^{\rm std}) \approx p^*(\mathcal{A}_{\varepsilon,d}) \approx 1$.

We now turn to more smooth classes of functions. Consider the following class $\mathcal{F}$:

$$\mathcal{F} = \{ f\colon [0,1] \to \mathbb{R} : f^{(j)}(0) = 0,\ \forall j < r,\ f^{(r-1)} \text{ is abs. cont., } \|f^{(r)}\|_2 \leqslant 1 \}.$$

As before, $S(f) = \int_0^1 f(x)\,dx$ for the integration problem, $S(f) = f \in \mathcal{G} = L_2([0,1])$ for the approximation problem, and $h^* \equiv 1$. Here r is a positive integer that measures regularity of functions f, and $r = 1$ corresponds to the Example problem analyzed above.

For the approximation problem, it is well known that the eigenvalues of $\mathcal{W}$ given by (1) satisfy $\lambda_i = \Theta(i^{-2r})$. Hence, the sum-exponent, see (27), is now given by

$$p_\lambda = 1/(2r). \tag{60}$$

For both problems, consider $m_1 = 1$ and $m_i = 1 + 2^{i-2}$ for $i \geqslant 2$, and the following algorithms B_{m_i-1}. We have $B_0 = 0$ and for the approximation problem, B_{m_i-1} is based on interpolation by piecewise polynomials of degree $r-1$ at points $x_{j,i} = j/2^{i-2} (1 \leqslant j \leqslant m_i - 1)$. For the integration, B_{m_i-1} is given as $B_{m_i-1}(f) = \int_0^1 \sigma_i(x)\,dx$, where $\sigma_i = \sigma_i(f)$ is a spline that minimizes $\|\sigma_i^{(r)}\|_2$ among all functions from $\mathcal{F}$ that interpolate f at the points $x_{j,i}$ $(1 \leqslant j \leqslant m_i - 1)$. The choice of *splines* σ_i guarantees the centrality of the corresponding integration algorithms.

It is well known that for both the integration and approximation problems there exists a number Q depending only on r such that $\|\mathcal{S} - B_{m_i-1}\|_{\mathcal{F}\to\mathcal{G}} \leqslant Q(m_i - 1)^{-r}$ for any $i \geqslant 2$. Since $\|B_{m_i-1} - B_{m_{i-1}-1}\|_{\mathcal{F}\to\mathcal{G}} \leqslant \|\mathcal{S} - B_{m_i-1}\|_{\mathcal{F}\to\mathcal{G}} + \|\mathcal{S} - B_{m_{i-1}-1}\|_{\mathcal{F}\to\mathcal{G}}$, the conditions (35) and (36) are satisfied with

$$C = \max\{ \|\mathcal{S}\|_{\mathcal{F}\to\mathcal{G}} + Q,\ Q(4^r + 2^r) \}, \qquad D = 2^{-r}, \qquad \text{and} \qquad p = \frac{1}{r}.$$

Thus, we have the following estimates for the exponents p^*:

Family 1. $p^*(\Lambda^{\rm all}) = \max\{1/r, 2/z\}$ and

$$p^*(\Lambda^{\rm std}) \leqslant p^*(\mathcal{A}_{\varepsilon,d}) \leqslant \begin{cases} \max\left\{ \dfrac{1}{r(1-\delta)}, \dfrac{2}{(z-2)_+} \right\} & \text{for approximation,} \\[2ex] \max\left\{ \dfrac{1}{r(1-\delta)}, \dfrac{2}{(z-1)_+} \right\} & \text{for integration.} \end{cases}$$

Family 2. $p^*(\Lambda^{\mathrm{all}}) = \max\{1/r, 2p_\gamma\}$ and

$$p^*(\Lambda^{\mathrm{std}}) \leqslant p^*(\mathscr{A}_{\varepsilon,d}) \leqslant \begin{cases} \max\left\{\dfrac{1}{r(1-\delta)}, \dfrac{2p_\gamma}{(1-2p_\gamma)_+}\right\} & \text{for approximation,} \\ \max\left\{\dfrac{1}{r(1-\delta)}, \dfrac{2p_\gamma}{(1-p_\gamma)_+}\right\} & \text{for integration.} \end{cases}$$

It is clear that similar results can be obtained for more general domains $\mathscr{D}$, more general spaces $\mathscr{F}$ and solution operators $\mathscr{S}$. What is really needed is the rate of convergence of the minimal errors $e(n)$ after n evaluations. As long as we know that $e(n) = O(n^{-r})$ then $p \leqslant 1/r$.

For linear functionals (such as integration), we also need nested information. The nested information can be obtained by adding the information already used at previous steps. More precisely, suppose that the optimal information N_i of cardinality, say, 2^{i-2} is not nested and the error of the central algorithm using N_i is of order $2^{-r(i-2)}$ for some positive r. To obtain the nested information we add to N_i all the previous information N_k with $k \leqslant i$ so that with the evaluation at zero, $m_i \leqslant 1 + \sum_{k=1}^{i} 2^{k-2} = 2^{i-1}$. Of course, any central algorithm B_{m_i-1} that uses this nested information has error of order m_i^{-r}. By taking $D = 2^{-r}$ in (35) and (36), we have $p = 1/r$. Hence, we lose only a multiplicative factor and we have the same p for optimal and nested information.

We plan to implement the WTP algorithm and we hope that in the process of implementation we will learn how to select the parameters of the WTP algorithm to make it more efficient. We will report on the results of implementation of the WTP algorithm in the future.

ACKNOWLEDGMENTS

We thank E. Novak and J. F. Traub for their valuable suggestions and comments.

REFERENCES

1. N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68** (1950), 337–404.
2. E. Novak and K. Ritter, High dimensional integration of smooth functions over cubes, *Numer. Math.* **75** (1996), 79–97.
3. E. Novak, K. Ritter, R. Schmitt, and A. Steinbauer, On an interpolatory method for high dimensional integration, *J. Comput. Appl. Math.*, in press.
4. E. Novak, I. H. Sloan, and H. Woźniakowski, Tractability of tensor product linear operators, *J. Complexity* **13** (1997), 387–418.

5. A. Papageorgiou and J. F. Traub, Beating Monte Carlo, *Risk* **9**, No. 6 (1996), 63–65.
6. B. P. Parlett, "The Symmetric Eigenvalue Problem," Prentice–Hall, London, 1980.
7. S. H. Paskov, New methodologies for valuing derivatives, *in* "Mathematics of Derivative Securities" (S. Pliska and M. Dempster, Eds.), Isaac Newton Institute, Cambridge Univ. Press, Cambridge, UK, 1997.
8. S. H. Paskov and J. F. Traub, Faster valuation of financial derivatives, *J. Portfolio Manage.* **22** (1995), 113–120.
9. J. Rust, private communication.
10. I. H. Sloan and H. Woźniakowski, When are Quasi-Monte Carlo algorithms efficient for high dimensional integrals, *J. Complexity* **14** (1998), 1–33.
11. S. A. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, *Dokl. Akad. Nauk SSSR* (1963), 240–243.
12. S. Tezuka, Financial applications of Monte Carlo and quasi-Monte Carlo methods, to appear.
13. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, "Information-Based Complexity," Academic Press, New York, 1988.
14. G. W. Wasilkowski and H. Woźniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems, *J. Complexity* **11** (1995), 1–56.
15. H. Woźniakowski, Tractability and strong tractability of linear multivariate problems, *J. Complexity* **10** (1994), 96–128.
16. H. Woźniakowski, Tractability and strong tractability of multivariate tensor product problems, *J. Comput. Inform.* **4** (1994), 1–19.
17. H. Woźniakowski, Strong tractability of weighted tensor products, *Ann. Numer. Math.* **4** (1997), 607–662.

Printed in Belgium