Polynomial-Time Algorithms
for Multivariate Linear Problems
with Finite-Order Weights;
Worst Case Setting[*]

G. W. Wasilkowski

Department of Computer Science, University of Kentucky

Lexington, KY 40506, USA, email: greg@cs.uky.edu

and

H. Woźniakowski

Department of Computer Science, Columbia University

New York, NY 10027, USA,

and Institute of Applied Mathematics, University of Warsaw

ul. Banacha 2, 02-097 Warszawa, Poland, email: henryk@cs.columbia.edu

October 25, 2004

Abstract

We consider approximation of linear multivariate problems defined over weighted tensor product Hilbert spaces with *finite-order* weights. This means we consider functions of d variables that can be represented as sums of functions of at most q^* variables. Here, q^* is fixed (and presumably small) and d may be arbitrarily large.

For the univariate problem, $d = 1$, we assume we know algorithms $A_{1,\varepsilon}$ that use $O(\varepsilon^{-p})$ function or linear functional evaluations to achieve an error ε in the worst

---

case setting. Based on these algorithms $A_{1,\varepsilon}$, we provide a construction of polynomial-time algorithms $A_{d,\varepsilon}$ for the general d-variate problem with the number of evaluations bounded roughly by $\varepsilon^{-p}d^{q^*}$ to achieve an error ε in the worst case setting.

1   Introduction

There is a host of practical problems that deal with functions of very many variables. In many cases, the required error tolerance, ε, for such problems is not too small. The classical estimates on approximation errors are asymptotic in the number, n, of evaluations and for the number, d, of variables fixed. They are usually of no practical value if n is fixed and d is very large. For instance, the classical discrepancy bounds are of the form $n^{-1}(\log n)^{d-1}$ and become meaningful only when the number n of evaluations significantly exceeds $\exp(d)$. This is why, since its introduction in 1994, see [43], there has been an increasing interest in the study of tractability of multivariate problems. The definition of tractability is recalled in the next section. Here we only mention that a problem is *tractable* if approximation errors go to zero with n, and are bounded by a polynomial in d and n^{-1} for every d and n. Equivalently, a problem is tractable if it is possible to reduce the initial error ε times by using a polynomial number of evaluations in ε^{-1} and d; and it is *strongly tractable*, if this number is independent of d. We stress that the upper bound on the number of evaluations should hold for all $\varepsilon \in (0,1)$ and all $d = 1, 2, \ldots$, including the case of huge d and relatively large ε, say $\varepsilon = 10^{-1}$. Algorithms that compute an ε-approximation and use a polynomial number of evaluations in ε^{-1} and d are called *polynomial-time* algorithms, and if this number does not depend on d they are called *strongly polynomial-time* algorithms.

There are many results on tractability of multivariate problems; however, quite a few of them are not constructive, see the survey paper [22] and many papers cited there. The results are obtained for problems defined over general tensor product spaces, mostly for reproducing kernel Hilbert spaces but some of the results are also for Banach spaces, see e.g., [15].

As observed in a number of papers, see, e.g., [4, 27, 34, 35], there are important problems, including problems in mathematical finance and physics, that deal with functions which only depend on groups of few variables. That is, the function depends on all d variables; however, it is given as the sum of terms each of which depends only on few, say q^*, variables. For some finance applications, the number q^* is fairly small, e.g., $q^* = 1$ or 2. An example of such functions in physics with $q^* = 6$ is provided by a sum of Coulomb pair potentials where $f(\mathbf{x}) = \sum_{1 \leq i < j \leq d} \|\vec{x}_i - \vec{x}_j\|^{-1}$ for vector $\mathbf{x} = [\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_d]$ with $\vec{x}_j \in \mathbb{R}^3$, and the Euclidean norm $\|\cdot\|$, see e.g., [10]. Since this function is not defined for $\vec{x}_i = \vec{x}_j$, we can modify it by taking a small positive α and consider $f_\alpha(\mathbf{x}) = \sum_{1 \leq i < j \leq d}(\|\vec{x}_i - \vec{x}_j\|^2 + \alpha)^{-1/2}$. That is, f and

f_α only depend on groups of two variables each being a 3-dimensional vector, see Section 7.4 where we discuss how such functions f_α can be efficiently approximated.

Functions of d variables can be represented as the sum of functions of groups $\mathbf{x}_u$ of variables with u varying through all subsets of the index set $\{1, 2, \ldots, d\}$. Probably, the first such representation was the ANOVA (for 'analysis of variance') decomposition of functions from a specific space as the sum of lower-dimensional terms, see [9, 29]. For functions belonging to a general tensor product reproducing kernel Hilbert space, and for $\mathbf{x} = [x_1, x_2, \ldots, x_d]$ we may decompose the function f as

$$f(\mathbf{x}) = \sum_{u \subset \{1,2,\ldots,d\}} \gamma_{d,u} f_u(\mathbf{x}) \tag{1}$$

for some functions f_u depending only on x_j for $j \in u$, and non-negative weights $\gamma_{d,u}$. The essence of the example with the Coulomb potential function is that we can set $\gamma_{d,u} = 0$ for all u with $|u| \neq 6$ and $\gamma_{d,u} = 1$ for all u with $|u| = 6$. If functions f from a given space satisfy (1) with $\gamma_{d,u} = 0$ for all u with the cardinality greater than, say, q^*, then we say that the space is equipped with *finite-order weights*; see the next section for the formal definition as well as the paper [8] where finite-order weights were introduced, and the paper [27] where finite-order weights were studied for multivariate integration. The concept of finite-order weights is related to the concept of *effective dimension* which has been studied in a number of papers, see e.g., [4, 34, 35] as well as a discussion at the end of the introduction and Remark 2 in Section 3.

For some spaces with finite-order weights we are able to obtain efficient algorithms that are polynomial-time or even strongly polynomial-time algorithms in the worst case. Indeed, it has been recently shown in [8, 27] that this is the case for approximating integrals $\int_{[0,1]^d} f(\mathbf{x}) \, d\mathbf{x}$ for Sobolev and Korobov spaces of functions equipped with finite-order weights. In this case, the quasi-Monte Carlo algorithms based on such classical low discrepancy points as Halton, Niederreiter, Sobol, as well as lattice rules and shifted lattice rules are polynomial or even strongly polynomial-time algorithms.

More general problems, including weighted L_2-approximation, were studied in a recent paper [42]. It was shown there that, under assumption (5) stated in Section 2, these problems are tractable or even strongly tractable for reproducing kernel Hilbert spaces equipped with finite-order weights. More specifically, an upper bound on the number of evaluations needed to compute an ε-approximation was shown to be independent on d and of order ε^{-2} or ε^{-4}; the former dependence for algorithms that use properly chosen linear functional evaluations, and the latter for algorithms that use only function evaluations at properly chosen points. For some problems these bounds are not sharp; however, in full generality, the bound of

order ε^{-2} cannot be improved. The bound ε^{-4} is probably not sharp. We stress that the proof of the latter result is *non-constructive*.

The present paper may be viewed as a continuation of [42]. Indeed, under slightly different assumptions and using different proof techniques, we provide constructions of polynomial-time algorithms that use either only function evaluations or arbitrary linear functionals for approximating linear problems over reproducing kernel Hilbert spaces equipped with finite-order weights. These algorithms are derived for arbitrary $d \geq 2$ in terms of tensor products of algorithms for $d = 1$ in a way similar to weighted tensor product algorithms studied in [38], see also [28]. Upper bounds on the number of evaluations needed to compute an ε-approximation for general d are practically the same as for $d = 1$ as far as the dependence on ε^{-1} is concerned. Hence, these upper bounds are sharp in ε^{-1} if we use optimal algorithms for $d = 1$. The dependence on d is polynomial and the degree of this polynomial depends on the order of the weights, i.e., on the largest cardinality of u for which $\gamma_{d,u}$ is still non-zero.

We stress that our algorithms are based on a modified Smolyak's construction, see [28]. This construction is also known in the literature as *sparse grid*, *hyperbolic cross*, and *Boolean blending*. Smolyak's construction and its variants proved very efficient for many problems including, in particular, differential and integral equations, integration and approximation of multivariate functions, and wavelets construction, see e.g., [2, 3, 6, 7, 12, 13, 14, 19, 21, 24, 25, 30, 31, 37, 38] and papers cited therein. This is especially the case when d is not very large. The case of large d and polynomial-time properties of algorithms based on Smolyak's construction have also been studied in some of the papers cited above.

We now explain our results in a more technical terms for the following simplified version of weighted approximation problem, where one wants to recover f with the error measured in a weighted L_2-norm,

$$\sqrt{\int_{D_d} |f(\mathbf{x}) - (Af)(\mathbf{x})|^2 \, \rho_d(\mathbf{x}) \, \mathrm{d}\mathbf{x}}.$$

Here $D_d = D \times \cdots \times D$ with $D \subset \mathbb{R}$, $\rho_d = \prod_{k=1}^{d} \rho(x_k)$ is a probability density function on D_d, and Af is an approximation given by an algorithm A. We assume that functions f belong to a reproducing kernel Hilbert space F_d whose formal definition is presented in the next section. For $d = 1$, we assume that functions from F_1 can be approximated with an error ε in the worst case setting using $O(\varepsilon^{-p})$ function or linear functional evaluations for some positive p. For many spaces F_1, the smallest exponent p is known. Typically, p depends on the smoothness of functions from F_1. For example, if F_1 is a space of r times differentiable functions then $p = 1/r$. This is the case for problems studied in Section 7. In general, under assumption (5), the exponent $p \leq 4$. If (5) does not hold then p can be arbitrary large, see Remark 1 of Section 2.

4

We construct a polynomial-time algorithm $A_{d,\varepsilon}$ for arbitrary d and for finite-order weights whose order is q^*, i.e., $\gamma_{d,u} = 0$ for all $u \subset \{1, 2, \ldots, d\}$ with $|u| > q^*$. This algorithm computes an ε-approximation in the worst case setting and uses roughly

$$\varepsilon^{-p} \, d^{q^*} \, \left(\ln(d/\varepsilon)\right)^{q^*(p+1)} \quad \forall \, d, \varepsilon \tag{2}$$

function or linear functional evaluations. The essence of (2) is that we have essentially the same dependence on ε^{-1} for all d, and a polynomial dependence on d with the exponent given by the order q^* of finite-order weights. We refer the reader to Theorem 1 in Section 3 for more details and all assumptions.

The algorithms $A_{d,\varepsilon}$ have the following additional property. Suppose that, as in the case of *effective dimension*, the function f is only approximately equal to a sum of functions of at most q^* variables. That is, $f = f_1 + f_2$, where $f_1 = \sum_{u:\, |u| \leq q^*} \gamma_{d,u} f_u$ is of the form required by finite-order weights, and $f_2 = \sum_{u:\, |u| > q^*} \gamma_{d,u} f_u$ is the remaining part of the decomposition of f, see (1), and has small $\|f_2\|_{F_d}$. We show in Remark 2 of Section 3 that, due to assumption (12),

$$A_{d,\varepsilon} f_2 \;=\; 0.$$

Hence $f - A_{d,\varepsilon} f = (f_1 - A_{d,\varepsilon} f_1) + f_2$, and the error of the algorithm $A_{d,\varepsilon}$ may increase only by the norm of f_2.

We now comment on the results concerning algorithms that may use arbitrary functional evaluations. As already mentioned, general results with constructive proofs have been obtained in [42] with the exponent $p = 2$. In Section 4, under an additional assumption (12) and using different proof techniques, we construct optimal algorithms $A_{d,\varepsilon}$ that compute an ε-approximation and use the number of evaluations as in (2). Hence, we may have the exponent p much smaller than 2. We also show that this bound is sharp in both ε^{-1} and d.

In Section 5, we study strong tractability of multivariate approximation. We present a necessary and sufficient condition for strong tractability. Due to optimality of $A_{d,\varepsilon}$, this condition is also necessary and sufficient for the strong polynomial-time property of $A_{d,\varepsilon}$. We also show that there is sometimes a tradeoff between the minimal exponents of ε^{-1} and d. Indeed, strong tractability may require a larger exponent in ε^{-1} than the exponent p for the univariate case $d = 1$. On the other hand, the exponent p can be always obtained at the expense of a polynomial dependence on d.

Some of the assumptions in Sections 2–5 are made to simplify the presentation and proofs. In Section 6, we discuss how some of these assumptions can be removed and explain how the results can be extended. In particular, we discuss more general problems than the weighted L_2-approximation problem. Finally, in Section 7 we illustrate the results by applying them to a number of specific problems including approximation of perturbed Coulomb potentials.

2   Problem Formulation

In this paper we study efficient algorithms for linear problems defined over spaces F_d of multivariate functions. These spaces are defined by tensor products of *basic* spaces, and therefore we begin with the definition of these basic spaces of functions.

For an arbitrary (and fixed) positive integer m, let $D \subset \mathbb{R}^m$ be a Lebesgue measurable set and let $\rho : D \to \mathbb{R}_+$ be a Lebesgue integrable function such that

$$\rho \geq 0 \qquad \text{and} \qquad \int_D \rho(t)\,\mathrm{d}t = 1.^1$$

Let $H(K)$ be a separable reproducing kernel Hilbert space of Lebesgue measurable real functions defined on D with the kernel $K : D \times D \to \mathbb{R}$. For a definition and properties of reproducing kernel Hilbert spaces we refer the reader to [1]. Here we only recall that

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

To simplify the presentation and derivation of the results, we assume up to Section 6 that the constant function $f(x) \equiv 1$ does *not* belong to $H(K)$,

$$1 \notin H(K). \tag{3}$$

This assumption is without any loss of generality; in Section 6 we explain how the results of this paper can be used for problems that do not satisfy (3).

Let $L_{2,\rho(D)}$ be the space of functions with norm

$$\|f\|_{L_{2,\rho}(D)} := \left(\int_D \rho(t)\,f^2(t)\,\mathrm{d}t\right)^{1/2} < \infty.$$

We also assume that

$$\sup_{f \in H(K)} \frac{\|f\|_{L_{2,\rho}(D)}}{\|f\|_{H(K)}} < \infty, \tag{4}$$

which means that $H(K)$ is continuously embedded into $L_{2,\rho}(D)$. Note that the condition

$$\int_D \rho(t)\, K(t, t)\,\mathrm{d}t < \infty, \tag{5}$$

---

[1] This assumption is made only for simplicity of presentation; all results hold as long as $\int_D \rho(t)\,\mathrm{d}t < \infty$.

that has been assumed in [42], is sufficient for (4). Indeed, for any $f \in H(K)$ we have $f^2(t) \leq \|f\|_{H(K)}^2 K(t,t)$ and, hence,

$$\|f\|_{L_{2,\rho}(D)} = \left(\int_D \rho(t) f^2(t) \, \mathrm{d}t \right)^{1/2} \leq \|f\|_{H(K)} \left(\int_D \rho(t) K(t,t) \right)^{1/2} \mathrm{d}t < \infty,$$

as claimed. However, (5) is not necessary for (4) as illustrated in Remark 1 that concludes this section.

We are ready to define the multivariate problems studied in this paper. For a given integer $d \geq 1$, define

$$D_d := D \times D \times \cdots \times D \subset \mathbf{R}^{dm} \qquad \text{and} \qquad \rho_d(\mathbf{t}) := \prod_{j=1}^{d} \rho(t_j),$$

where $\mathbf{t} = [t_1, t_2, \ldots, t_d]$ with $t_j \in D$. Clearly, $\int_{D_d} \rho(\mathbf{t}) \, \mathrm{d}\mathbf{t} = 1$.

In what follows, we assume that u is a subset of indices from the set $\{1, 2, \ldots, d\}$. By $|u|$ we denote the cardinality of u. Let $\gamma = \{\gamma_{d,u}\}$ be a non-zero sequence of non-negative numbers, called *weights*, indexed by d and u. That is, for each d we have 2^d non-negative weights $\gamma_{d,u}$. As in [8, 27], we say that $\gamma = \{\gamma_{d,u}\}$ are *finite-order* weights if there exists an integer q such that

$$\gamma_{d,u} = 0 \qquad \text{for all } (d,u) \text{ with } |u| > q. \tag{6}$$

Finite-order weights γ are of *order* q^* if q^* is the smallest integer q satisfying (6). Let $\mathcal{U}_d$ denote the set of nonempty subsets u with positive $\gamma_{d,u}$, i.e.,

$$\mathcal{U}_d := \{u \subset \{1, 2, \ldots, d\} : u \neq \emptyset, \gamma_{d,u} > 0\}.$$

It should be clear that for weights of order q^* the cardinality of $\mathcal{U}_d$ is proportional to d^{q^*}. Actually, by a simple induction on q^* it can be shown that

$$|\mathcal{U}_d| \leq 2 \cdot d^{q^*} \quad \forall d, \tag{7}$$

and $|\mathcal{U}_d| \leq d^{q^*}/(q^* - 1)!$ when $d \geq 2 q^*$.

Consider now the following weighted reproducing kernel Hilbert space

$$F_d := H(K_d)$$

of real functions with $d\,m$ variables defined on D_d with the kernel

$$K_d(\mathbf{x}, \mathbf{y}) = \gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u} \prod_{j \in u} K(x_j, y_j) \qquad \forall \mathbf{x}, \mathbf{y} \in D_d. \tag{8}$$

7

We now characterize functions from $H(K_d)$. Let $K_{d,u}(\mathbf{x}, \mathbf{y}) = \prod_{j \in u} K(x_j, y_j)$ for $\mathbf{x}, \mathbf{y} \in D_d$ denote a term in (8). Clearly, $K_{d,u}$ is the reproducing kernel of the Hilbert space $H(K_{d,u})$ of functions $f(t_1, t_2, \ldots, t_d)$ defined on D_d which do not depend on t_j for all $j \notin u$, and is the tensor product space with respect to the spaces $H(K)$ of functions depending on variables t_i with $i \in u$. Since the constant function $f(\mathbf{x}) = 1$ does not belong to $H(K_{d,u})$, we have $H(K_{d,u}) \cap H(K_{d,v}) = \{0\}$ for all distinct u and v form $\mathcal{U}_d$. This also means that the Hilbert space F_d is the direct sum of Hilbert spaces $H(K_{d,u})$ for all subsets of u. Here $K_{d,\emptyset} = 1$ and $H(1) = \text{span}(1)$. Moreover the subspaces $\text{span}(1)$ and $H(K_{d,u})$ for $u \in \mathcal{U}_d$ are pairwise orthogonal. Hence, any $f \in F_d$ can be uniquely decomposed as

$$f = f_\emptyset + \sum_{u \in \mathcal{U}_d} f_u = \gamma_{d,\emptyset} f_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u} f_{d,u} \qquad \text{with} \quad f_u = \gamma_{d,u} f_{d,u} \in H(K_{d,u})$$

and with $f_\emptyset \equiv \text{constant}$, where $f_\emptyset = 0$ if $\gamma_{d,\emptyset} = 0$. Moreover,

$$
\begin{aligned}
\langle f, g \rangle_{H(K_d)} &= \gamma_{d,\emptyset}^{-1} f_\emptyset\, g_\emptyset + \sum_{u \in \mathcal{U}_d} \gamma_{d,u}^{-1} \langle f_u, g_u \rangle_{H(K_{d,u})} \\
&= \gamma_{d,\emptyset} f_{d,\emptyset}\, g_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u} \langle f_{d,u}, g_{d,u} \rangle_{H(K_{d,u})},
\end{aligned}
$$

for all $f, g \in H(K_d)$. (Here, for $\gamma_{d,\emptyset} = 0$ we take $0/0 = 0$.)

We stress that the term $f_{d,u}$ is a function of $m \cdot |u|$ variables, i.e., it depends on $|u|$ points $t_k \in D \subseteq \mathbb{R}^m$ with $k \in u$. For finite-order weights of order q^*, there are only $O(d^{q^*})$ such terms.

Observe also that the assumption (4) implies a continuous embedding $F_d \subset L_{2,\rho_d}(D_d)$.

Consider now linear multivariate operators defined over the spaces $F_d = H(K_d)$. More precisely, for $d = 1, 2, \ldots$, let

$$S_d \,:\, F_d \;\to\; G_d$$

be a continuous linear operator and G_d a separable Hilbert space. Similarly as in [42, 43], we assume that the operator S_d is also continuous with respect to the norm of the space $L_{2,\rho_d}(D_d)$. That is, there exists a non-negative number C_d such that

$$\|S_d f\|_{G_d} \leq C_d \|f\|_{L_{2,\rho_d}(D_d)} \qquad \forall\, f \in F_d. \tag{9}$$

For instance, the *multivariate weighted integration problem* with $G_d = \mathbb{R}$ and $S_d = \text{INT}_d$,

$$S_d f = \text{INT}_d f = \int_{D_d} f(\mathbf{t})\, \rho_d(\mathbf{t})\, d\mathbf{t},$$

satisfies (9) with $C_d = 1$. The *multivariate weighted approximation problem* is yet another example of an important problem satisfying (9) with $C_d = 1$. It is defined by $G_d = L_{2,\rho_d}(D_d)$ and $S_d = \mathrm{APP}_d$,

$$S_d f = \mathrm{APP}_d f = f \quad \forall f \in F_d.$$

Note that APP_d is a continuous linear operator due to (4).

Our goal is to approximate elements $S_d f$ for $f \in F_d$. We approximate $S_d f$ by computing a number of values $L(f)$ of continuous linear functionals belonging to a class Λ of permissible functionals. We study two classes of Λ. The first class, $\Lambda = \Lambda^{\mathrm{all}} = F_d^*$, consists of all continuous linear functionals. The second class, $\Lambda = \Lambda^{\mathrm{std}}$, consists of function evaluations only.[2] That is, $L \in \Lambda^{\mathrm{std}}$ iff there exists $\mathbf{t} \in D_d$ such that $L(f) = f(\mathbf{t})$ for all $f \in F_d$. Obviously, L is also continuous since $L(f) = \langle f, K_d(\cdot, \mathbf{t}) \rangle_{F_d}$ and $\|L\| = \sqrt{K_d(\mathbf{t}, \mathbf{t})}$.

For problems considered in this paper, it is known that adaptive choice of linear functionals as well as nonlinear algorithms do not help, see [11, 33]. Hence, we can restrict our attention to linear algorithms,

$$Af = \sum_{j=1}^{n} L_j(f)\, a_j,$$

where $L_j \in \Lambda$ and $a_j \in G_d$ for $j = 1, 2, \dots, n$. The number n of functional evaluations is called *the cardinality of A* and is denoted by $\mathrm{card}(A)$,

$$\mathrm{card}(A) = n.$$

The worst case error of the algorithm A (with respect to the space F_d) is defined as

$$e^{\mathrm{wor}}(A; F_d) := \sup_{f \in F_d} \frac{\|S_d f - Af\|_{G_d}}{\|f\|_{F_d}}.$$

Due to linearity of S_d and A we obviously have $e^{\mathrm{wor}}(A; F_d) = \|S_d - A\|$. Here the operator norm is from F_d to G_d. This implies that

$$\|S_d f - Af\|_{G_d} \le \|f\|_{F_d} \cdot e^{\mathrm{wor}}(A; F_d) \qquad \forall f \in F_d.$$

For $n = 0$, we formally set $A \equiv 0$ and then $e^{\mathrm{wor}}(A; F_d) = \|S_d\|$ is the initial error which can be obtained without sampling the functions f from F_d. We want to reduce this initial error by a factor $\varepsilon \in (0, 1)$ and we are interested in finding the smallest number n of evaluations for which it is possible. Let

$$n(\varepsilon, S_d, \Lambda) := \min\{\mathrm{card}(A) : A \text{ uses } L_j \in \Lambda \text{ and } e^{\mathrm{wor}}(A; F_d) \le \varepsilon\,\|S_d\|\}.$$

---

[2] $\Lambda = \Lambda^{\mathrm{std}}$ is considered only for simplicity of presentation; see Section 6 for a more general case.

Since we are using different spaces and different operator norms, we will sometimes write $\|S_d\| = \|S_d\|_{F_d \to G_d}$ to make it clear what spaces are involved in the operator norm.

As in many papers dealing with tractability, we say that the multivariate problem $\{S_d\}$ is *tractable* in the class Λ if there exist non-negative numbers C, p and q such that

$$n(\varepsilon, S_d, \Lambda) \leq C\,\varepsilon^{-p}\,d^q \qquad \forall\,\varepsilon \in (0,1),\ \forall\,d = 1, 2, \ldots . \tag{10}$$

Algorithms $A_{d,\varepsilon}$ are called *polynomial-time* algorithms if for every d and ε, the algorithm $A_{d,\varepsilon}$ uses at most $C\,\varepsilon^{-p}\,d^q$ functional evaluations and has the error bounded by $\varepsilon\|S_d\|$, i.e.,

$$\mathrm{card}(A_{d,\varepsilon}) \leq C\,\varepsilon^{-p}\,d^q \qquad \text{and} \qquad e^{\mathrm{wor}}(A_{d,\varepsilon}; F_d) \leq \varepsilon\,\|S_d\|. \tag{11}$$

If $q = 0$ in (10) then we say that the multivariate problem $\{S_d\}$ is *strongly tractable* in the class Λ. Moreover, algorithms $A_{d,\varepsilon}$ are called *strongly polynomial-time* algorithms if (11) holds with $q = 0$. The infimum of p satisfying (11) with $q = 0$ over all strongly polynomial-time algorithms is called the *exponent of strong tractability.*

The algorithms proposed in this paper will have cardinality bounds that are slightly different than (11). They will be of the form

$$\mathrm{card}(A_{d,\varepsilon}) \leq C\varepsilon^{-p}\,d^q\,\ln^\alpha(d/\varepsilon)$$

for some positive number α independent of d and ε. Such bounds imply polynomial-time property since for every $\delta > 0$ there exists a number C_δ such that

$$C\varepsilon^{-p}\,d^q\,\ln^\alpha(d/\varepsilon) \leq C_\delta\,\varepsilon^{-p(1+\delta)}\,d^{q(1+\delta)} \qquad \forall\,\varepsilon,\,d.$$

We stress that for many applications, $m = 1$, i.e., the basic space $H(K)$ consists of univariate functions. However, there are important problems with $m \geq 2$, see e.g., [16, 17, 18, 19] as well as the second and fourth problems in Section 7. Of course, if $H(K)$ is by itself a tensor product of spaces of univariate functions, the problem with $m \geq 2$ can be easily treated as a tensor product problem with $m = 1$ and d replaced by $d\,m$. However, this is not the case when, for instance, $H(K)$ is an isotropic reproducing kernel Hilbert space and/or D is not a Cartesian product of subsets of $\mathbb{R}$, e.g., D is a ball.

We end this section with the following remark.

Remark 1 As mentioned in the introduction, the assumption (5) implies that the exponent p in (10) is at most 2 for the class Λ^{all}. This assumption is not necessary for the weighted approximation problem to be well defined and to satisfy the assumptions of this

paper. However, without (5), the exponent p in (10) can be arbitrarily large (including infinity) even for the scalar case $d = 1$. To see this, consider the following problem.

Let $D = [0, 1)$ and $\rho \equiv 1$. Consider an arbitrary ordered sequence $\{\lambda_i\}$ of positive numbers, $\lambda_i \geq \lambda_{i+1} > 0$, and the following sequence $a_i = i/(i+1)$ for $i = 0, 1, \ldots$. Define

$$e_i(x) := \sqrt{i(i+1)}\, \mathbf{1}_{[a_{i-1}, a_i)}(x) \quad \text{and} \quad \eta_i(x) := \sqrt{\lambda_i}\, e_i(x) \quad \text{for} \quad i = 1, 2, \ldots$$

and the following kernel

$$K(x, t) := \sum_{i=1}^{\infty} \eta_i(x) \cdot \eta_i(t).$$

Here $\mathbf{1}_{[a_{i-1}, a_i)}$ is the characteristic function of the interval $[a_{i-1}, a_i)$. Hence e_i are piece-wise constant and orthonormal functions in $L_2([0, 1))$ (similar construction can be also done for smooth functions).

Of course, K is a well-defined reproducing kernel. It follows from the general results concerning reproducing kernel Hilbert spaces that the corresponding space $H(K)$ has $\{\eta_i\}$ as its orthonormal system. Moreover, as can be easily verified, (λ_i, η_i) are the eigenpairs of the operator $W := \mathrm{APP}_1^* \, \mathrm{APP}_1 : H(K) \to H(K)$,

$$\lambda_i\, \eta_i(\cdot) \;=\; W\eta_i(\cdot) \;=\; \int_0^1 \eta_i(t)\, K(\cdot, t)\, \mathrm{d}t.$$

This means that (4) holds since

$$\sup_{f \in H(K)} \frac{\|f\|_{L_2}}{\|f\|_{H(K)}} \;=\; \|\mathrm{APP}_1\| \;=\; \sqrt{\lambda_1}.$$

However, if $\sum_{i=1} \lambda_i = \infty$ then (5) does not hold since

$$\int_0^1 K(x, x)\, \mathrm{d}x \;=\; \sum_{i=1}^{\infty} \lambda_i.$$

It follows from general facts, see, e.g., [33] and/or Section 4, that

$$n(\varepsilon, \mathrm{APP}_1, \Lambda^{\mathrm{all}}) \;=\; \inf\{k \;:\; \lambda_{k+1} \leq \varepsilon^2 \lambda_1\}.$$

Hence, depending on the choice of the eigenvalues λ_i, the function $n(\cdot, \mathrm{APP}_1, \Lambda^{\mathrm{all}})$ may increase to infinity very rapidly when $\varepsilon \to 0$. For instance, for an arbitrary positive p and for $\lambda_i = i^{-2/p}$ we have $n(\varepsilon, \mathrm{APP}_1, F_1) = \Theta(\varepsilon^{-p})$. Moreover, if λ_i do not converge to zero then $n(\varepsilon, \mathrm{APP}_1, F_1) = \infty$ whenever $\varepsilon < \sqrt{\lim_i \lambda_i / \lambda_1}$.

11

3 Polynomial-Time Algorithms for Λ^{std} or Λ^{all}

We provide a construction of polynomial-time algorithms $A_{d,\varepsilon}$ under the following additional assumption

$$K(a,a) = 0 \quad \text{for some } a \in D. \tag{12}$$

This assumption implies that $K(t,a) = 0$ for any $t \in D$, and hence it is equivalent to assuming that every function from $H(K)$ vanishes at a. It can be replaced by another two assumptions as explained in Section 6.3.

The construction is based on a sequence of algorithms B_i for the basic case $d = 1$. More specifically, let $\{B_i\}_{i=0}^{\infty}$ be a sequence of algorithms using functionals from the class $\Lambda \in \{\Lambda^{\mathrm{std}}, \Lambda^{\mathrm{all}}\}$ for approximation of functions from the space $H(K)$. We assume that the algorithms B_i have the following properties:

$$B_0 \equiv 0, \qquad \lim_{i \to \infty} \|\mathrm{APP}_1 - B_i\|_{H(K) \to L_{2,\rho}(D)} = 0, \tag{13}$$

$$\|B_i - B_{i-1}\|_{H(K) \to L_{2,\rho}(D)} \leq E_0 \, 2^{-i} \quad \forall \, i \geq 2, \tag{14}$$

and

$$\mathrm{card}(B_i) \leq D_0 \, 2^{ip} \quad \forall \, i \geq 1, \tag{15}$$

for positive numbers E_0, D_0 and p. We also assume that the algorithms B_i use "nested" information, i.e., functionals used by B_{i-1} are also used by B_i. Of course, we can assume that such algorithms exist since, otherwise, the problem would require more than a polynomial number of evaluations in ε^{-1} even for $d = 1$.

For $f \in H(1 + K)$, consider now

$$A_1 f := f(a) \qquad \text{and} \qquad A_i f := B_{i-1}(f - f(a)) \quad \text{for } i \geq 2.$$

Here and in the rest of this paper, we use multi-indices $\mathbf{i}$. For instance, $\mathbf{i} \in \mathbb{N}_+^k$ means that $\mathbf{i}$ is a k-dimensional vector $[i_1, \ldots, i_k]$ whose components i_ℓ are positive integers. By $|\mathbf{i}|$ we denote $\sum_{\ell=1}^{k} i_\ell$, and we write $\mathbf{i} \geq \mathbf{j}$ to denote that $i_\ell \geq j_\ell$ for all $\ell = 1, 2, \ldots, k$.

For a subset $u \in \mathcal{U}_d$ and a multi-index $\mathbf{i} \in \mathbb{N}_+^{|u|}$ whose all components are greater than 1 (i.e., $\mathbf{i} \geq \mathbf{2}$), define

$$\Delta_{u,\mathbf{i}} := \bigotimes_{k=1}^{d} G_k(u,\mathbf{i}) \qquad \text{with} \qquad G_k(u,\mathbf{i}) := \begin{cases} A_1 & \text{if } k \notin u, \\ A_{i_k} - A_{i_k-1} & \text{if } k \in u. \end{cases}$$

For the empty set $u = \emptyset$,

$$\Delta_\emptyset := \bigotimes_{k=1}^{d} A_1, \quad \text{i.e.,} \quad \Delta_\emptyset(f) = f(\mathbf{a}) \quad \text{with} \quad \mathbf{a} = [a, \dots, a].$$

Since $A_i(1) = 0$ for $i \geq 2$, and $A_1(K(\cdot, t)) = 0$ for any $t \in D$, the operators Δ have the following important properties:

$$\Delta_\emptyset(1) = 1 \quad \text{and} \quad \Delta_\emptyset(H(K_{d,v})) = \{0\} \quad \forall\, v \neq \emptyset,$$

as well as

$$\Delta_{u,\mathbf{i}}(H(K_{d,v})) = \{0\} \quad \forall\, u \neq \emptyset,\ \forall\, \mathbf{i} \geq \mathbf{2},\ \forall\, v \neq u. \tag{16}$$

We are ready to define algorithms for approximation of functions from the space F_d. These algorithms will depend on the parameters d, ε, and γ. We suppress the dependence on γ in the definition of the algorithms and we list only the dependence on d and ε. Hence, we have the algorithms $A_{d,\varepsilon}$ defined by

$$A_{d,\varepsilon} f := f(\mathbf{a}) + \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \in Q(u)} \Delta_{u,\mathbf{i}}(f). \tag{17}$$

The sets $Q(u) = Q(u, \varepsilon)$ defining the algorithm are of the form

$$Q(u) := \left\{ \mathbf{i} \in \mathbb{N}^{|u|} : \mathbf{i} \geq \mathbf{2} \text{ and } |\mathbf{i}| \leq m(u) \right\} \tag{18}$$

with the integer numbers $m(u) = m(u, \varepsilon)$ that will be specified later. Of course, to guarantee that the set $Q(u) \neq \emptyset$ we will choose $m(u) \geq 2|u|$.

Let $Q^c(u)$ denote the complement of $Q(u)$

$$Q^c(u) := \left\{ \mathbf{i} \in \mathbb{N}^{|u|} : \mathbf{i} \geq \mathbf{2} \text{ and } |\mathbf{i}| \geq m(u) + 1 \right\}.$$

We now estimate the error of $f - A_{d,\varepsilon} f$. From (13), we easily conclude that $f = f(\mathbf{a}) + \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \in \mathbb{N}_+^{|u|}} \Delta_{u,\mathbf{i}}(f)$. Hence

$$f - A_{d,\varepsilon} f = \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \in Q^c(u)} \bigotimes_{k=1}^{d} G_k(u, \mathbf{i})(f).$$

Using the representation $f = f_\emptyset(c) + \sum_{u \in \mathcal{U}_d} f_u$ with $f_u \in H(K_{d,u})$ and (16), we conclude

$$f - A_{d,\varepsilon} f = \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \in Q^c(u)} \bigotimes_{k \in u} (B_{i_k - 1} - B_{i_k - 2})(f_u),$$

13

and

$$\|f - A_{d,\varepsilon} f\|_{L_{2,\rho_d}(D_d)}$$

$$\leq \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \in Q^c(u)} \|f_u\|_{H(K_{d,u})} \left\| \bigotimes_{k \in u} (B_{i_k-1} - B_{i_k-2}) \right\|_{H(K_{d,u}) \to L_{2,\rho_d}(D_d)}$$

$$= \sum_{u \in \mathcal{U}_d} \sqrt{\gamma_{d,u}} \|f_u\|_{F_d} \sum_{\mathbf{i} \in Q^c(u)} \prod_{k \in u} \|B_{i_k-1} - B_{i_k-2}\|_{H(K) \to L_{2,\rho}(D)}.$$

Due to (14),

$$\sum_{\mathbf{i} \in Q^c(u)} \prod_{k \in u} \|B_{i_k-1} - B_{i_k-2}\|_{H(K) \to L_{2,\rho}(D)} \leq E_0^{|u|} \sum_{\mathbf{i} \in Q^c(u)} 2^{-|\mathbf{i}|+|u|}.$$

The last sum is given in [26, Lemma 2] and therefore we have

$$\|f - A_{d,\varepsilon} f\|_{L_{2,\rho_d}(D_d)} \leq \sum_{u \in \mathcal{U}_d} \sqrt{\gamma_{d,u}} \|f_u\|_{F_d} E_0^{|u|} 2^{-(m(u)-|u|)} \sum_{j=0}^{|u|-1} \binom{m(u) - |u|}{j}. \tag{19}$$

We estimate the sum of these binomial coefficients in the following lemma.

Lemma 1 *The following inequality holds*

$$\sum_{j=0}^{|u|-1} \binom{m - |u|}{j} \leq c_{q^*} \left(\frac{m - |u|}{|u|} \right)^{|u|} \qquad \forall \, |u| \leq q^*, \; \forall \, m \geq 2 \, |u| \tag{20}$$

with

$$c_{q^*} := e^{q^*} \cdot \sqrt{q^* e^{1/12} \pi^{-1}} \cdot \max \left(1, \, (\ln 2)^{q^*} \sqrt{3 \, e} \right)$$

Proof: In what follows, we will represent m as $m = x \, |u|$ with $x \geq 2$. Consider first the case $x \in [2, 3]$. The largest value of $\binom{m - |u|}{j}$ is for $j = \lceil (m - |u|)/2 \rceil$ and the sum in (20) can be estimated by

$$|u| \binom{(x - 1)|u|}{\lceil (x - 1)|u|/2 \rceil}.$$

Using Stirling's formula, $n! = (n/e)^n \sqrt{2\pi n} \, e^{\theta_n/(12n)}$ for some $\theta_n \in (0, 1)$, it is easy to show that

$$\binom{n}{k} \leq \left(\frac{n}{k} \right)^k \left(\frac{n}{n - k} \right)^{n-k} \sqrt{\frac{n \, e^{1/(6n)}}{2\pi k(n - k)}}. \tag{21}$$

We use the last estimate for $n = (x-1)|u|$ and $k = \lceil (x-1)|u|/2 \rceil$. If $n = 1$ then $k = 1$ and $\binom{n}{k} = 1$. If n is odd and at least equal to 3, i.e., $n = 2p+1$ with $p \geq 1$, then $k = p+1$ and we estimate

$$\left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} \sqrt{\frac{n\, e^{1/(6n)}}{2\pi k(n-k)}} \leq \left(\frac{2p+1}{p+1}\right)^{p+1} \left(\frac{2p+1}{p}\right)^p \sqrt{\frac{(2p+1)e^{1/(6n)}}{2\pi(p+1)p}}$$

$$\leq 2^n \left(1 + \frac{1}{2p}\right)^p \sqrt{\frac{(2+1/p)e^{1/(6n)}}{2\pi(p+1)}} \leq 2^n \sqrt{\frac{3e^{1+1/(6n)}}{\pi n}}.$$

The last estimate is also true if n is even. Therefore for $(x-1)|u| \geq 2$ we have

$$|u| \left(\begin{array}{c} (x-1)|u| \\ \lceil (x-1)|u|/2 \rceil \end{array} \right) \leq |u|\, 2^{(x-1)|u|} \sqrt{\frac{3e^{13/12}}{\pi(x-1)|u|}} \leq 2^{(x-1)|u|} \sqrt{3q^* e^{13/12}/\pi}.$$

Since $2^{x-1}/(x-1) \leq e\ln 2$, we obtain (20).

Consider now the remaining case $x \geq 3$. The terms of the sum in (20) are now increasing since $j \leq (m-|u|)/2 - 1$. Therefore the sum in (20) is smaller than $|u| \left(\begin{array}{c} (x-1)|u| \\ |u| \end{array} \right)$. Applying (21) for $n = (x-1)|u| \geq 2$ and $k = |u|$ we easily conclude that

$$|u| \left(\begin{array}{c} (x-1)|u| \\ |u| \end{array} \right) \leq \sqrt{|u|e^{1/12}/\pi}\, (x-1)^{|u|}\, e^{|u|} \leq \sqrt{q^* e^{1/12}\pi^{-1}}\, (x-1)^{|u|}\, e^{q^*}$$

which implies (20) and completes the proof of the lemma. □

We continue to estimate the error $\|f - A_{d,\varepsilon}f\|_{L_{2,\rho_d}(D_d)}$. Since $m(u) \geq 2|u|$ we can apply Lemma 1 to (19), and obtain

$$\|f - A_{d,\varepsilon}f\|_{L_{2,\rho_d}(D_d)} \leq \sum_{u \in \mathcal{U}_d} \sqrt{\gamma_{d,u}}\, \|f_u\|_{F_d} c_{q^*}\, E_0^{|u|}\, 2^{-(m(u)-|u|)} \left(\frac{m(u)-|u|}{|u|}\right)^{|u|}.$$

Since $\|f\|_{F_d}^2 = \sum_{u \in \mathcal{U}_d} \|f_u\|_{F_d}^2 \leq 1$, we conclude that

$$\|f - A_{d,\varepsilon}f\|_{L_{2,\rho_d}(D_d)} \leq c_{q^*} \left(\sum_{u \in \mathcal{U}_d} \gamma_{d,u}\, E_0^{2|u|}\, 2^{-2(m(u)-|u|)} \left(\frac{m(u)-|u|}{|u|}\right)^{2|u|} \right)^{1/2}. \tag{22}$$

Consider the function $f(x) = (x-1)2^{-(x-1)}$ for $x \geq 2$. It is easy to check that $\max_{x \geq 2} f(x) = f(1 + 1/\ln 2) = 1/(e \ln 2)$, and that f is decreasing to zero for $x \geq 1 + 1/\ln 2 = 2.44\ldots$.

We are ready to define $m(u)$ as

$$m(u) = \lceil x_u |u| \rceil, \tag{23}$$

where x_u is the smallest number $x \geq 1 + 1/\ln 2$ for which

$$\left((x-1)2^{-(x-1)} \right)^{|u|} \leq \sqrt{y_u} \tag{24}$$

with

$$y_u := \frac{\varepsilon^2 \|\text{APP}_d\|^2}{c_{q^*}^2} \left(\frac{D_0^{|u|}}{\gamma_{d,u} E_0^{2|u|}} \right)^{1/(1+p/2)} \frac{1}{\sum_{v \in \mathcal{U}_d} \left(\gamma_{d,v} E_0^{2|v|} \right)^{p/(2+p)} D_0^{|v|/(1+p/2)}}.$$

Here p and D_0 are the numbers from (15). That is, if $y_u \geq (e \ln 2)^{-2|u|}$ then $x_u = 1 + 1/\ln 2$, and if $y_u < (e \ln 2)^{-2|u|}$ then x_u is the unique solution of the non-linear equation

$$(x-1)2^{-(x-1)} = y_u^{1/(2|u|)}.$$

Clearly, $\lceil x_u |u| \rceil$ can be computed exactly by using just a few steps of Newton's iterations.

This concludes the definition of the algorithms $A_{d,\varepsilon}$ for the approximation problem APP_d. We stress that $m(u)$ depends on all parameters d, ε, and γ. From the definition of $m(u)$ and (22) we obtain that

$$\|f - A_{d,\varepsilon}\|_{L_{2,\rho_d}(D_d)} \leq c_{q^*} \left(\sum_{u \in \mathcal{U}_d} \gamma_{d,u} E_0^{2|u|} y_u \right)^{1/2} = \varepsilon \|\text{APP}_d\|.$$

Hence, the error of $A_{d,\varepsilon}$ is at most $\varepsilon \|\text{APP}_d\|$.

For the general linear problems $\{S_d\}$, we define the algorithms $\widehat{A}_{d,\varepsilon}$ by

$$\widehat{A}_{d,\varepsilon} := S_d \circ A_{d,\varepsilon^*} \quad \text{with} \quad \varepsilon^* := \frac{\varepsilon \|S_d\|}{C_d \|\text{APP}_d\|} \tag{25}$$

with the numbers C_d from (9). It is clear that the error of $\widehat{A}_{d,\varepsilon}$ is bounded by $\varepsilon \|S_d\|$. We summarize this analysis in the following lemma.

16

Lemma 2 *Let (12-15) hold. The algorithms $A_{d,\varepsilon}$ given by (17), (18), and (23), and $\widehat{A}_{d,\varepsilon}$ given by (25) have errors satisfying*

$$e^{\mathrm{wor}}(A_{d,\varepsilon};F_d) \leq \varepsilon \,\|\mathrm{APP}_d\| \quad and \quad e^{\mathrm{wor}}(\widehat{A}_{d,\varepsilon};F_d) \leq \varepsilon \,\|S_d\| \quad \forall\, d \in \mathbb{N}_+, \,\forall\, \varepsilon \in (0,1).$$

The definitions of the algorithms $A_{d,\varepsilon}$ and $\widehat{A}_{d,\varepsilon}$ formally require the knowledge of all weights $\gamma_{d,u}$ as well as the norm $\|\mathrm{APP}_d\|$ which are used in the definition of $m(u)$. Note that $m(u)$ is a decreasing function of $\|\mathrm{APP}_d\|$, and therefore we can use a lower bound on $\|\mathrm{APP}_d\|$ in the definition of $m(u)$. Hence, if this norm is unknown, it can be replaced by the following lower bound obtained through the integration problem, see [42, Theorem 1],

$$\|\mathrm{APP}_d\|^2 \geq \|\mathrm{INT}_d\|^2 = \gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u}\, c^{|u|},$$

with

$$c := \int_{D^2} \rho(t)\, \rho(x)\, K(t,x)\, \mathrm{d}(t,x).$$

Lemma 2 as well as the next theorem hold for the algorithms with $\|\mathrm{APP}_d\|$ replaced by $\|\mathrm{INT}_d\|$ as above.

We now estimate the cardinalities of the algorithms $\widehat{A}_{d,\varepsilon}$ and $A_{d,\varepsilon}$. The cardinality of $\widehat{A}_{d,\varepsilon}$ does not exceed the cardinality of $A_{d,\varepsilon}$, and the latter cardinality can be estimated by the sum of cardinalities of the corresponding algorithms $\Delta_{u,\mathbf{i}}$. That is,

$$\mathrm{card}(A_{d,\varepsilon}) \leq 1 + \sum_{u \in \mathcal{U}_d} \mathrm{card}(u) \qquad \text{with} \qquad \mathrm{card}(u) := \sum_{\mathbf{i} \in Q(u)} \mathrm{card}\left(\bigotimes_{k=1}^{d} G_k(u,\mathbf{i})\right).$$

As in [37], $\mathrm{card}(u) \leq \sum_{\mathbf{i} \in Q(u)} \mathrm{card}\left(\bigotimes_{k \in u} B_{i_k-1}\right) \leq \sum_{\mathbf{i} \in Q(u)} \prod_{k \in u} \mathrm{card}(B_{i_k-1})$. Therefore due to (15),

$$\mathrm{card}(u) \leq \sum_{\mathbf{i} \in Q(u)} D_0^{|u|}\, 2^{p(|\mathbf{i}|-|u|)} = D_0^{|u|} \sum_{\ell=|u|}^{m(u)-|u|} 2^{p\ell} \binom{\ell-1}{|u|-1}.$$

The binomials in the last sum are increasing, so we can estimate them by the largest one obtained for $\ell = m(u) - |u|$. We then get

$$\mathrm{card}(u) \leq \frac{D_0^{|u|}}{2^p - 1}\, 2^{p(m(u)-|u|+1)} \binom{m(u)-|u|-1}{|u|-1}.$$

Using (20) we finally obtain

$$\mathrm{card}(u) \leq \frac{c_{q^*} \, 2^p}{2^p - 1} \, D_0^{|u|} \, 2^{\, p(m(u) - |u|)} \left(\frac{m(u) - |u|}{|u|} \right)^{|u|}.$$

From (23) we have $m(u) - |u| = \lceil (x_u - 1)|u| \rceil \leq 1 + (x_u - 1)|u|$ and

$$2^{p(m(u)-|u|)} \left(\frac{m(u) - |u|}{|u|} \right)^{|u|} \leq 2^{p(x_u-1)|u|} (x_u - 1)^{|u|} 2^p e^{1/(x_u-1)} \leq 2^{p+1} 2^{p(x_u-1)|u|} (x_u - 1)^{|u|}$$

since $1/(x_u - 1) \leq \ln 2$. From this we obtain

$$\mathrm{card}(A_{d,\varepsilon}) \leq 1 + \frac{c_{q^*} \, 2^{2p+1}}{2^p - 1} T(A_{d,\varepsilon}) \qquad \text{with} \qquad T(A_{d,\varepsilon}) := \sum_{u \in \mathcal{U}_d} D_0^{|u|} \, 2^{\, p(x_u-1)|u|} \, (x_u - 1)^{|u|}.$$

(26)

To estimate $T(A_{d,\varepsilon})$ we need the following inequality which is easy to check. For any $p > 0$, $x \geq 1$ and $a \geq 1$

$$x \leq \widehat{c}_p \, x^{-p} \left(\ln \left(2^{xa} \, x^{-a} \right) \right)^{p+1} \qquad \text{with} \qquad \widehat{c}_p := \left(\ln(2) - e^{-1} \right)^{-(p+1)}.$$

(27)

Then letting $x = x_u - 1$ and $a = |u|$, we obtain

$$T(A_{d,\varepsilon}) \leq \sum_{u \in \mathcal{U}_d} (\widehat{c}_p \, D_0)^{|u|} \left(\frac{2^{\, (x_u-1)|u|}}{(x_u - 1)^{|u|}} \right)^p \left(\ln \left(2^{\, (x_u-1)|u|} (x_u - 1)^{-|u|} \right) \right)^{(p+1)|u|}.$$

From the definition of x_u we have

$$\left(\frac{2^{(x_u-1)}}{x_u - 1} \right)^{2|u|} = \frac{1}{z_u},$$

where $z_u = \min \left(y_u, (e \ln 2)^{-2|u|} \right)$. Hence,

$$
\begin{aligned}
T(A_{d,\varepsilon}) \quad &\leq \quad \sum_{u \in \mathcal{U}_d} (\widehat{c}_p D_0)^{|u|} \, z_u^{-p/2} \left(\frac{1}{2} \ln z_u^{-1} \right)^{(p+1)|u|} \\
&\leq \quad \max_{u \in \mathcal{U}_d} \left(\ln z_u^{-p/2} \right)^{(p+1)|u|} \sum_{u \in \mathcal{U}_d} \left(\widehat{c}_p D_0 p^{-(p+1)} \right)^{|u|} z_u^{-p/2} \\
&\leq \quad \max \left(\frac{\widehat{c}_p}{p^{p+1}}, \left(\frac{\widehat{c}_p}{p^{p+1}} \right)^{q^*} \right) \max_{u \in \mathcal{U}_d} \left(\ln z_u^{-p/2} \right)^{(p+1)|u|} \sum_{u \in \mathcal{U}_d} D_0^{|u|} z_u^{-p/2}.
\end{aligned}
$$

18

We estimate the logarithmic factor as follows:

$$
\max_{u \in \mathcal{U}_d} \left(\ln(z_u^{-p/2}) \right)^{(p+1)|u|} = \max_{u \in \mathcal{U}_d} \left(\ln(D_0^{|u|} z_u^{-p/2}) - \ln(D_0^{|u|}) \right)^{(p+1)|u|}
$$

$$
\leq \left(\ln \left(\sum_{u \in \mathcal{U}_d} D_0^{|u|} z_u^{-p/2} \right) + \max \left(\ln(D_0^{-1}),\, q^* \ln(D_0^{-1}) \right) \right)^{(p+1)q^*}.
$$

To estimate the sum of $D_0^{|u|} z_u^{-p/2}$ note that

$$
z_u^{-p/2} = \max \left(y_u^{-p/2}, (e \ln 2)^{p|u|} \right) \leq y_u^{-p/2} + (e \ln 2)^{p|u|}.
$$

We have

$$
\sum_{u \in \mathcal{U}_d} D_0^{|u|} y_u^{-p/2} = \frac{c_{q^*}^p}{\varepsilon^p \|\mathrm{APP}_d\|^p} \sum_{u \in \mathcal{U}_d} D_0^{|u|} \left(\frac{\gamma_{d,u} E_0^{2|u|}}{D_0^{|u|}} \right)^{p/(2+p)}
$$

$$
\times \left(\sum_{v \in \mathcal{U}_d} \left(\gamma_{d,v} E_0^{2|v|} \right)^{p/(2+p)} D_0^{|v|/(1+p/2)} \right)^{p/2}
$$

$$
= \frac{c_{q^*}^p}{\varepsilon^p \|\mathrm{APP}_d\|^p} \left(\sum_{u \in \mathcal{U}_d} \left(\gamma_{d,u} E_0^{2|u|} \right)^{p/(2+p)} D_0^{|u|/(1+p/2)} \right)^{1+p/2} \tag{28}
$$

$$
\leq \frac{c_{q^*}^p |\mathcal{U}_d|}{\varepsilon^p \|\mathrm{APP}_d\|^p} \left(\sum_{u \in \mathcal{U}_d} \gamma_{d,u} \right)^{p/2} \frac{\sum_{u \in \mathcal{U}_d} D_0^{|u|} E_0^{|u|p}}{|\mathcal{U}_d|}
$$

$$
\leq c_{q^*}^p \max \left(D_0 E_0^p, (D_0 E_0^p)^{q^*} \right) \frac{|\mathcal{U}_d|}{\varepsilon^p} \left(\frac{\sum_{u \in \mathcal{U}_d} \gamma_{d,u}}{\|\mathrm{APP}_d\|^2} \right)^{p/2}.
$$

Since $\sum_{u \in \mathcal{U}_d} \left(D_0 (e \ln 2)^p \right)^{|u|} \leq \max \left(D_0 (e \ln 2)^p, (D_0 (e \ln 2)^p)^{q^*} \right) |\mathcal{U}_d|$, we conclude that

$$
\mathrm{card}(A_{d,\varepsilon}) \leq 1 + b_1 \left(b_2 \frac{|\mathcal{U}_d|}{\varepsilon^p} \left(\frac{\sum_{u \in \mathcal{U}_d} \gamma_{d,u}}{\|\mathrm{APP}_d\|^2} \right)^{p/2} + b_3 |\mathcal{U}_d| \right) \tag{29}
$$

$$
\times \left(\left(\ln \left(\frac{b_2 |\mathcal{U}_d|}{\varepsilon^p} \left(\frac{\sum_{u \in \mathcal{U}_d} \gamma_{d,u}}{\|\mathrm{APP}_d\|^2} \right)^{p/2} + b_3 |\mathcal{U}_d| \right) + b_4 \right) \right)^{(p+1)q^*}
$$

with the numbers b_i given by

$$b_1 \quad := \quad \frac{c_{q^*} \, 2^{2p+1}}{2^p - 1} \, \max \left(\frac{\widehat{c}_p}{p^{p+1}} \, , \, \left(\frac{\widehat{c}_p}{p^{p+1}} \right)^{q^*} \right) \tag{30}$$

$$b_2 \quad := \quad c_{q^*}^p \, \max \left(D_0 E_0^p \, , \, (D_0 E_0^p)^{q^*} \right) , \tag{31}$$

$$b_3 \quad := \quad \max \left(D_0 (e \ln 2)^p \, , \, (D_0 (e \ln 2)^p)^{q^*} \right) , \tag{32}$$

$$b_4 \quad := \quad \max(\ln(D_0^{-1}) \, , \, q^* \, \ln(D_0^{-1})). \tag{33}$$

We are ready to prove the following theorem.

Theorem 1 *Let (12-15) hold and let $\gamma = \{\gamma_{d,u}\}$ be arbitrary finite-order weights with the order q^*.*

Then the weighted approximation problem is tractable and the algorithms $\{A_{d,\varepsilon}\}$ defined by (17), (18), and (23) are polynomial-time algorithms.

The general linear problem $\{S_d\}$ is tractable if, in addition to (9), there exists a number k such that

$$\sup_{d=1,2\ldots} \frac{C_d \, \|\mathrm{APP}_d\|_{F_d \to L_{2,\rho_d}(D_d)}}{d^k \, \|S_d\|_{F_d \to G_d}} < \infty. \tag{34}$$

Then the algorithms $\{\widehat{A}_{d,\varepsilon}\}$ defined by (25) are polynomial-time algorithms.

More precisely, the following bounds hold.

(i) *If $\int_{D^2} \rho(t) \, \rho(x) \, K(t,x) \, \mathrm{d}(t,x) > 0$ then there exists constants a_i such that for every d and every $\varepsilon \in (0,1)$,*

$$\begin{aligned}
\mathrm{card}(A_{d,\varepsilon}) \quad &\leq \quad a_1 \, |\mathcal{U}_d| \, \varepsilon^{-p} \, \left(1 + \ln(|\mathcal{U}_d|/\varepsilon) \right)^{(p+1)q^*} \\
&\leq \quad 2 \, a_1 \, d^{q^*} \, \varepsilon^{-p} \, \left(1 + \ln(2 \, d^{q^*}/\varepsilon) \right)^{(p+1)q^*} , \tag{35} \\
\mathrm{card}(\widehat{A}_{d,\varepsilon}) \quad &\leq \quad a_2 \, |\mathcal{U}_d|^{1+kp} \, \varepsilon^{-p} \, \left(1 + \ln(|\mathcal{U}_d|/\varepsilon) \right)^{(p+1)q^*} \\
&\leq \quad 2 \, a_2 \, d^{q^*(1+kp)} \, \varepsilon^{-p} \, \left(1 + \ln(2d^{q^*}/\varepsilon) \right)^{(p+1)q^*} . \tag{36}
\end{aligned}$$

(ii) *$\int_{D^2} \rho(t) \, \rho(x) \, K(t,x) \, \mathrm{d}(t,x) = 0$ then there exists constants a_i such that for every d and every $\varepsilon \in (0,1)$,*

$$\begin{aligned}
\mathrm{card}(A_{d,\varepsilon}) \quad &\leq \quad a_3 \, d^{q^*(1+p/2)} \, \varepsilon^{-p} \, \left(1 + \ln(d/\varepsilon) \right)^{(p+1)q^*} , \tag{37} \\
\mathrm{card}(\widehat{A}_{d,\varepsilon}) \quad &\leq \quad a_4 \, d^{q^*(1+p/2)+kp} \, \varepsilon^{-p} \, \left(1 + \ln(d/\varepsilon) \right)^{(p+1)q^*} . \tag{38}
\end{aligned}$$

Proof: The theorem follows immediately from Lemma 2, (29), (7), (25), and the following estimates on $\|\text{APP}_d\|$. If $c = \int_{D^2} \rho(x)\rho(t)K(x,t)\,\mathrm{d}(x,t)$ is positive then, as already mentioned after Lemma 2, $\|\text{APP}_d\|^2 \geq \|\text{INT}_d\|^2 = \gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u} c^{|u|}$. Therefore,

$$\frac{\sum_{u \in \mathcal{U}_d} \gamma_{d,u}}{\|\text{APP}_d\|^2} \leq \frac{\gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u}}{\gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u} c^{|u|}} \leq \max(1, c^{-q^*});$$

so they are uniformly bounded in d.

Otherwise, if $\int_{D^2} \rho(x)\rho(t)K(x,t)\,\mathrm{d}(x,t) = 0$ then

$$\|\text{APP}_d\|^2 = \max\left(\gamma_{d,\emptyset}, \max_{u \in \mathcal{U}_d} \gamma_{d,u} \|W\|^{|u|}\right),$$

where $\|W\|$ is the norm of $W : H(K) \to H(K)$ given by $W(f) = \int_D f(t)K(t,\cdot)\rho(t)\,\mathrm{d}t$, see [42, Lemma 2], as well as Remark 1. In this case, $\sum_{u \in \mathcal{U}_d} \gamma_{d,u}/\|\text{APP}_d\|^2$ are at most proportional to d^{q^*}. $\qquad\square$

The essence of Theorem 1 is that we have polynomial-time algorithms for *arbitrary* finite-order weights. For some specific cases, these algorithms might be even strongly polynomial-time. For instance, (35) implies the strong polynomial-time property of $\{A_{d,\varepsilon}\}$ for the approximation problem if the cardinalities of the sets $\mathcal{U}_d$ are uniformly bounded for all d.

In general, when the cardinality of $\mathcal{U}_d$ increases like d^{q^*}, the cost of the algorithms $A_{d,\varepsilon}$ is essentially as large as $d^{q^*}\varepsilon^{-p}$ which could be very large when d and q^* are large and ε is small. However, under an additional assumption on the weights, we can propose algorithms $A_{d,\varepsilon,p'}$ whose cardinalities are essentially bounded by $|\mathcal{U}_d| + \varepsilon^{-p'}$ for some $p' \geq p$.

To define the algorithm $A_{d,\varepsilon,p'}$, we first observe that the algorithms $A_{d,\varepsilon}$ and $\hat{A}_{d,\varepsilon}$ depend on all parameters of the problem S_d. In particular, they depend on p which is used in the definition of y_u and effects $m(u)$ given by (23). The parameter p bounds the cardinality of the algorithms B_i, see (15). Obviously if we replace p by a larger number p' then (15) holds and the analysis of this section goes through. That is how new algorithms are defined. Namely, the algorithms $A_{d,\varepsilon,p'}$ and $\hat{A}_{d,\varepsilon,p'}$ are defined as $A_{d,\varepsilon}$ and $\hat{A}_{d,\varepsilon}$ with the parameter p replaced by p' with $p' \geq p$. For simplicity we now only analyze the algorithm $A_{d,\varepsilon,p'}$ for the case when $\int_{D^2} \rho(t)\rho(x)K(t,x)\,\mathrm{d}(t,x)$ is positive.

Theorem 2 *Let $\int_{D^2} \rho(t)\rho(x)K(t,x)\,\mathrm{d}(t,x) > 0$. Let*

$$r^* := \sup\left(r \geq 1 \ : \ \sup_d \frac{\sum_{u \in \mathcal{U}_d} \gamma_{d,u}^{1/r}}{\left(\sum_{u \in \mathcal{U}_d} \gamma_{d,u}\right)^{1/r}} < \infty\right)$$

If $r^* > 1$ then for any

$$p' > \max(p, \, 2/(r^* - 1))$$

(or any $p' \geq \max(p, 2/(r^* - 1))$ if the infimum for r^* is attained) the algorithms $\{A_{d,\varepsilon,p'}\}$ are polynomial-time algorithms and there exist positive numbers a_i such that

$$\text{card}(\mathcal{A}_{d,\varepsilon,p'}) \leq a_1 \, |\mathcal{U}_d| + a_2 \, \varepsilon^{-p'} (\ln \varepsilon^{-1})^{(p'+1)q^*} \qquad \forall \, d, \, \varepsilon. \qquad (39)$$

Proof: The proof is the same as the proof of Theorem 1 up to (28). That is, we differently estimate $\sum_{u \in \mathcal{U}_d} D_0^{|u|} y_u^{-p'/2}$. More specifically, by using p' instead of p, and keeping in mind that $\|\text{APP}_d\|^2 \geq \min(1, c^{q^*}) \sum_{u \in \mathcal{U}_d} \gamma_{d,u}$ with $c = \int_{D^2} \rho(t) \, \rho(x) \, K(t, x) \, \mathrm{d}(t, x)$, we have

$$\sum_{u \in \mathcal{U}_d} D_0^{|u|} y_u^{-p'/2} \leq \frac{c_{q^*}^{p'}}{\varepsilon^{p'}} \left(\frac{\sum_{u \in \mathcal{U}_d} \gamma_{d,u}^{p'/(2+p')}}{\left(\sum_{u \in \mathcal{U}_d} \gamma_{d,u} \right)^{p'/(2+p')}} \right)^{1+p'/2} \frac{\max(D_0 \, E_0^{p'}, \, (D_0 \, E_0^{p'})^{q^*})}{\min(1, \, c^{q^*})^{p'/2}}.$$

Then denoting $(2 + p')/p'$ by r we see that $r < r^*$ and hence

$$\left(\frac{\sum_{u \in \mathcal{U}_d} \gamma_{d,u}^{p'/(2+p')}}{\left(\sum_{u \in \mathcal{U}_d} \gamma_{d,u} \right)^{p'/(2+p')}} \right)^{1+p'/2} = \left(\frac{\sum_{u \in \mathcal{U}_d} \gamma_{d,u}^{1/r}}{\left(\sum_{u \in \mathcal{U}_d} \gamma_{d,u} \right)^{1/r}} \right)^{1+p'/2}$$

is bounded independently of d. The rest of the proof is the same as before. $\qquad \square$

Although the bound (39) does not imply strong polynomial-time property, it exhibits a weaker dependence on d than (35) since it it separates the dependence on d from the one on $1/\varepsilon$. We end this section with the following remark.

Remark 2 The error of the algorithm $A_{d,\varepsilon}$ is derived under the finite-order weights assumption, i.e., for functions f belonging to $\bigoplus_{|u| \leq q^*} H(K_u)$. It is interesting to check what happens with the error of $A_{d,\varepsilon}$ if the algorithm is applied to functions with 'small' components outside the space $\bigoplus_{|u| \leq q^*} H(K_u)$, as in the case of small effective dimension. That is, let f be of the form

$$f = f_1 + f_2 \quad \text{with} \quad f_1 \in \bigoplus_{|u| \leq q^*} H(K_{d,u}), \quad f_2 \in \bigoplus_{|u| \geq q^*+1} H(K_{d,u}) \quad \text{and} \quad \|f_2\|_{H_{K_{d,u}}} \text{ is 'small'.}$$

Since $f_2(\mathbf{a}) = 0$, the property (16) implies that the algorithm $A_{d,\varepsilon}$ vanishes on f_2. In fact

$$A_{d,\varepsilon} \left(\bigoplus_{|u| \geq q^*+1} H(K_{d,u}) \right) = \{0\}.$$

22

Hence its error,
$$f - A_{d,\varepsilon} f = (f_1 - A_{d,\varepsilon} f_1) + f_2,$$
increases at most by $\|f_2\|_{G_d} \le \|\mathrm{APP}_d\| \cdot \|f_2\|_{F_d}$.

4 Polynomial-Time Algorithms for Λ^{all} when

$$\int_{D^2} \rho(x)\rho(t)K(x,t)\,\mathrm{d}(t,x) = 0$$

In this section, we describe polynomial-time algorithms $A_{d,\varepsilon}$ that use information functionals from Λ^{all}, and lead to better tractability bounds in d. As mentioned in the introduction, general results for Λ^{all} have been obtained in [42] under the assumption (5). Here we do not assume (5) and strengthen the results of [42] under the following assumption:

$$\int_{D^2} \rho(x)\rho(t)K(x,t)\,\mathrm{d}(t,x) = 0. \tag{40}$$

As explained in [42], this is equivalent to assuming that $\int_D f(x)\rho(x)\,\mathrm{d}x = 0$ for any $f \in H(K)$.

As in the last part of the proof of Theorem 1, consider the operator $W : H(K) \to H(K)$,

$$W(f)(\cdot) := \int_D f(t)\,K(t,\cdot)\,\rho(t)\,\mathrm{d}t.$$

We assume that the operator W is compact. This assumption is without loss of generality since otherwise the complexity of the approximation problem would be infinite for small ε even for $d = 1$, see e.g., [33]. Let $\{(\lambda_i, \zeta_i)\}_{i \in \mathbb{N}_+}$ be the eigenpairs of W with orthonormalized ζ_i and ordered so that

$$\|W\| = \lambda_1 \ge \cdots \ge \lambda_i \ge \lambda_{i+1} \ge \cdots \ge 0.$$

Then $\{\zeta_i\}$ forms an ON (orthonormal) system in $H(K)$. Moreover

$$\langle \zeta_i, \zeta_j \rangle_{L_{2,\rho}} = \langle W(\zeta_i), \zeta_j \rangle_{H(K)} = \lambda_i\,\delta_{i,j} \qquad \forall\, i,j \in \mathbb{N}_+. \tag{41}$$

It is well-known, see e.g., [33], that $\sqrt{\lambda_{n+1}}$ is the minimal error for solving the basic, $d = 1$, problem APP_1 among all algorithms that use at most n evaluations from Λ^{all}. To guarantee tractability we must assume that λ_i goes polynomially fast to zero with n^{-1}. Hence, we assume that

$$\exists\, B > 0, \quad \exists\, p > 0, \quad \forall\, i = 1, 2, \ldots, \quad \text{we have} \quad \lambda_i \le B\,i^{-2/p}, \tag{42}$$

which corresponds to assumptions (14) and (15) from Section 3.

For a nonempty set u, define

$$\zeta_{\mathbf{i},u}(\mathbf{x}) := \prod_{k \in u} \zeta_{i_k}(x_k) \quad \text{and} \quad \lambda_{\mathbf{i},u} := \prod_{k \in u} \lambda_{i_k} \qquad \text{for} \qquad \mathbf{i} \in \mathbb{N}_+^{|u|}.$$

It is easy to see that for any $u \neq \emptyset$, $\{\zeta_{\mathbf{i},u}\}_{\mathbf{i} \in N_+^{|u|}}$ is an ON system in $H(K_{d,u})$. Due to orthogonality between $H(K_{d,u})$ and $H(K_{d,v})$ for $u \neq v$, the set containing 1 and all $\zeta_{\mathbf{i},u}$'s forms an orthogonal system in the space F_d. Of course, the $\eta_{\mathbf{i},u}$'s given by

$$\eta_{\mathbf{i},u} := \zeta_{\mathbf{i},u} \sqrt{\gamma_{d,u}} \tag{43}$$

together with the normalized 1 yield an ON system in F_d. That is,

$$f = \langle f, 1 \rangle_{F_d} \cdot \sqrt{\gamma_{d,\emptyset}} + \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \in \mathbb{N}_+^{|u|}} \langle f, \eta_{\mathbf{i},u} \rangle_{F_d} \cdot \eta_{\mathbf{i},u}.$$

Consider now the operator $W_d : F_d \to F_d$ defined by

$$W_d := \mathrm{APP}_d^* \circ \mathrm{APP}_d$$

Then (40) implies that $(\gamma_{d,u}\, \lambda_{\mathbf{i},u}, \eta_{\mathbf{i},u})$ are the eigenpairs of W_d, see [42, Lemma 1].

We study the algorithms $A_{d,\varepsilon}$ defined by

$$A_{d,\varepsilon} f := \langle f, 1 \rangle_{F_d} \cdot \sqrt{\gamma_{d,\emptyset}} + \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \in P(u)} \langle f, \eta_{\mathbf{i},u} \rangle_{F_d} \cdot \eta_{\mathbf{i},u} \tag{44}$$

with the sets $P(u) = P(\varepsilon, u)$ given by

$$P(u) := \left\{ \mathbf{i} \in \mathbb{N}_+^{|u|} \ : \ \gamma_{d,u}\, \lambda_{\mathbf{i},u} \geq \varepsilon^2 \, \|\mathrm{APP}_d\|^2 \right\}. \tag{45}$$

As in Section 3, for the general problem $\{S_d\}$ we assume (9) and define the algorithms

$$\widehat{A}_{d,\varepsilon} := S_d \circ A_{d,\varepsilon^*} \qquad \text{with} \qquad \varepsilon^* := \frac{\varepsilon \, \|S_d\|}{C_d \, \|\mathrm{APP}_d\|},$$

We have the following lemma.

Lemma 3 *Let (40) hold. For every d and $\varepsilon \in (0,1)$, the algorithms $A_{d,\varepsilon}$ and $\widehat{A}_{d,\varepsilon}$ satisfy*

$$e^{\mathrm{wor}}(A_{d,\varepsilon}; F_d) \leq \varepsilon \, \|\mathrm{APP}_d\| \qquad \text{and} \qquad e^{\mathrm{wor}}(\widehat{A}_{d,\varepsilon}; F_d) \leq \varepsilon \, \|S_d\|$$

24

with

$$\mathrm{card}(\widehat{A}_{d,\varepsilon}) \leq \mathrm{card}(A_{d,\varepsilon}) = 1 + \sum_{u \in \mathcal{U}_d} |P(\varepsilon, u)|. \tag{46}$$

Moreover, if also (42) holds then

$$|P(\varepsilon, u)| \leq \begin{cases} \alpha \ln^{|u|-1}(2\alpha + 1) & \text{if } \alpha \geq 1 \\ 0 & \text{otherwise,} \end{cases} \quad \text{where} \quad \alpha := \left(\frac{\gamma_{d,u} B^{|u|}}{\varepsilon^2 \|\mathrm{APP}_d\|^2} \right)^{p/2}. \tag{47}$$

Proof: We begin by estimating the error. Of course,

$$f - A_{d,\varepsilon}f = \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \notin P(u)} \langle f, \eta_{\mathbf{i},u} \rangle_{H(K_d)} \cdot \eta_{\mathbf{i},u}.$$

Therefore

$$\|f - A_{d,\varepsilon}(f)\|^2_{L_{2,\rho_d}(D_d)}$$
$$= \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \notin P(u)} \sum_{v \in \mathcal{U}_d} \sum_{\mathbf{j} \notin P(v)} \langle f, \eta_{\mathbf{i},u} \rangle_{F_d} \cdot \langle f, \eta_{\mathbf{j},v} \rangle_{F_d} \cdot \int_{D_d} \rho(\mathbf{x}) \eta_{\mathbf{i},u}(\mathbf{x}_u) \eta_{\mathbf{j},v}(\mathbf{x}_v) \, \mathrm{d}\mathbf{x}$$
$$= \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \notin P(u)} \langle f, \eta_{\mathbf{i},u} \rangle^2_{F_d} \cdot \gamma_{d,u} \cdot \lambda_{\mathbf{i},u},$$

with the last equality due to (41) and (43). Since $\|f\|^2_{F_d} = \sum_{u,\mathbf{i}} \langle f, \eta_{\mathbf{i},u} \rangle^2_{F_d}$, this implies that

$$e^{\mathrm{wor}}(A_{d,\varepsilon}; F_d) = \max_{u \in \mathcal{U}_d} \max_{\mathbf{i} \notin P(u)} \sqrt{\gamma_{d,u} \cdot \lambda_{\mathbf{i},u}} \leq \varepsilon \|\mathrm{APP}_d\|,$$

as claimed. For the general $\{S_d\}$, note that for $\|f\|_{F_d} \leq 1$,

$$\|S_d f - \widehat{A}_{d,\varepsilon} f\|_{G_d} \leq C_d \, e^{\mathrm{wor}}(A_{d,\varepsilon^*}) \leq C_d \|\mathrm{APP}_d\| \, \varepsilon^* = \varepsilon \|S_d\|,$$

which completes the first part of the proof.

We now estimate the cardinalities of both algorithms. For that end, observe that (46) is obvious since the functionals $\langle f, \eta_{\mathbf{i},u} \rangle_{F_d}$ are linearly independent for all $u \in \mathcal{U}_d$ and $\mathbf{i} \in P(u)$. The sharp estimates on the cardinality of the set $P(u)$ can be found in [23, Lemma A.1] and [44, Lemma 4.1]. Here, we opt for simplicity and present a simple proof which gives a sharp estimate on ε but not sharp with respect to the asymptotic constant.

Let $p(a, k)$ denotes the cardinality of a set on multi-indices $\mathbf{i} \in \mathbb{N}^k_+$ with $\prod_{j=1}^k i_j \leq a$. Then

$$p(a, k) \leq a \ln^{k-1}(2a + 1) \text{ for } a \geq 1 \quad \text{and} \quad p(a, k) = 0 \text{ for } a \in [0, 1). \tag{48}$$

25

Indeed, (48) holds for $k = 1$, and by induction for $a \geq 1$,

$$p(a, k+1) = \sum_{i=1}^{\lfloor a \rfloor} p(a/i, k) \leq a \ln^{k-1}(2a+1) \sum_{i=1}^{\lfloor a \rfloor} i^{-1}.$$

Let $f(x) = 1/x$. Then for every i and $x \in [i - 1/2, i + 1/2]$ we have $f(x) = i^{-1} - i^{-2}(x - i) + \zeta_i^{-3}(x - i)^2$ for some $\zeta_i \in [i - 1/2, i + 1/2]$. Hence $i^{-1} \leq \int_{i-1/2}^{i+1/2} f(x) \, dx$ and consequently

$$p(a, k+1) \leq a \ln^{k-1}(2a+1) \int_{1/2}^{\lfloor a \rfloor + 1/2} x^{-1} \, dx \leq a \ln^k(2a+1),$$

as claimed. Clearly, (48) implies (47) since $P(u) = p(\alpha, |u|)$. This completes the proof. $\square$

We are ready to state the following theorem which easily follows from Lemma 3 and the proof of part **(ii)** of Theorem 1.

Theorem 3 *Let (40) and (42) hold. Let $\gamma = \{\gamma_{d,u}\}$ be arbitrary finite-order weights with the order q^*.*

The weighted approximation problem $\{\mathrm{APP}_d\}$ is tractable, and the algorithms $\{A_{d,\varepsilon}\}$ defined by (44) and (45) are polynomial-time algorithms. More precisely, there exists a number a_1 such that

$$\mathrm{card}(A_{d,\varepsilon}) \leq a_1 \varepsilon^{-p} |\mathcal{U}_d| \left(\ln(|\mathcal{U}_d|/\varepsilon) \right)^{q^*-1} \leq 2 a_1 \varepsilon^{-p} d^{q^*} \left(\ln(2 d^{q^*}/\varepsilon) \right)^{q^*-1} \quad \forall d, \varepsilon.$$

The general problem $\{S_d\}$ is tractable if, in addition to (9), there exists a number k such that

$$\sup_{d=1,2\ldots} \frac{C_d \, \|\mathrm{APP}_d\|_{F_d \to L_{2,\rho_d}(D_d)}}{d^k \, \|S_d\|_{F_d \to G_d}} < \infty.$$

Then the algorithms $\{\widehat{A}_{d,\varepsilon}\}$ defined by (25) are polynomial-time algorithms. More precisely, there exists a number a_2 such that

$$\mathrm{card}(\widehat{A}_{d,\varepsilon}) \leq a_2 \varepsilon^{-p} d^{q^*+kp} \left(\ln(d/\varepsilon) \right)^{q^*-1} \quad \forall d, \varepsilon.$$

We now show that the upper bound for the approximation problem presented in Theorem 3 is essentially sharp. To see this, we consider

$$\lambda_i = i^{-2/p}$$

26

for some positive p. Then $\|W\| = 1$. We estimate from below the cardinalities of the sets $P(\varepsilon, u)$ given by (45) by counting only multi-indices of the form $\mathbf{i} = [i_1, 1, \ldots, 1]$ with $i_1^{-2/p} \geq \varepsilon^2 \|\text{APP}_d\|^2/\gamma_{d,u}$. Hence $|P(\varepsilon, u)| \geq \left\lfloor (\gamma_{d,u}/(\varepsilon^2\|\text{APP}_d\|^2))^{p/2} \right\rfloor$ and

$$\text{card}(A_{d,\varepsilon}) \; \geq \; \sum_{u \in \mathcal{U}_d} \left\lfloor \left(\frac{\gamma_{d,u}}{\varepsilon^2 \|\text{APP}_d\|^2} \right)^{p/2} \right\rfloor. \tag{49}$$

Consider now the finite-order weights with $\gamma_{d,u} = 1$ if $|u| \leq q^*$. Then $\|\text{APP}_d\| = \max(\gamma_{d,\emptyset}, \max_{u \in \mathcal{U}_d} \gamma_{d,u}\|W\|^{|u|}) = 1$ and

$$\text{card}(A_{d,\varepsilon}) \; \geq \; (\varepsilon^{-p} - 1) \cdot |\mathcal{U}_d|.$$

Since the cardinality of $\mathcal{U}_d$ is proportional to d^{q^*}, this proves that the bound from Theorem 3 is sharp modulo the logarithmic factor.

It follows from the general results, see e.g., [33], that the algorithms $A_{d,\varepsilon}$ are *complexity optimal*, i.e., they have the smallest cardinality among all algorithms that solve the approximation problem with error at most $\varepsilon \|\text{APP}_d\|$. Hence, whenever the bound on the cardinality of $A_{d,\varepsilon}$ is sharp, it becomes a sharp complexity bound.

5 Strong Tractability for Λ^{all}

In this section we show necessary and sufficient conditions on finite-order weights to guarantee strong tractability of the weighted approximation problem. Due to complexity optimality of the algorithms $A_{d,\varepsilon}$ these conditions are equivalent to $\{A_{d,\varepsilon}\}$ being strongly polynomial-time.

We need to assume the exact order of behavior of the eigenvalues λ_i for the basic case $d = 1$. Therefore we assume that p in (42) is best possible, or more precisely that

$$\lambda_i \; = \; \Theta\left(i^{-2/p} \right). \tag{50}$$

Theorem 4 *Let (40) and (50) hold. Then the approximation problem $\{\text{APP}_d\}$ equipped with finite-order weights is strongly tractable iff there exists a positive number s such that*

$$\sup_d \sum_{u \in \mathcal{U}_d} \widehat{\gamma}_{d,u}^{\,s} < \infty \quad \text{with} \quad \widehat{\gamma}_{d,u} := \frac{\gamma_{d,u}}{\max\limits_{v \in \{\emptyset\} \cup \mathcal{U}_d} \gamma_{d,v}}. \tag{51}$$

If this holds then the exponent of strong tractability is

$$p^* \; = \; \max(p, 2s^*),$$

where s^* is the infimum over s satisfying (51).

Moreover, if (51) holds then the algorithms $\{A_{d,\varepsilon}\}$ are strongly polynomial-time algorithms and for every $\delta > 0$ there exists a_δ such that

$$\text{card}(A_{d,\varepsilon}) \leq a_\delta \, \varepsilon^{-p^*(1+\delta)}.$$

Proof: Due to complexity optimality of $A_{d,\varepsilon}$ it is enough to verify when the cardinality of $A_{d,\varepsilon}$ does not depend on d. Suppose that (51) holds for some $s > s^*$. When $s > p/2$ then (46) and (47) imply

$$
\begin{aligned}
\text{card}(A_{d,\varepsilon}) \;\; &\leq \;\; 1 + c_s \sum_{u \in \mathcal{U}_d} \left(\frac{\gamma_{d,u} \, B^{\,|u|}}{\varepsilon^2 \|\text{APP}_d\|^2} \right)^s \\
&\leq \;\; 1 + \frac{c_s \max(1, B^{\,s\,q^*})}{\varepsilon^{2s}} \sum_{u \in \mathcal{U}^d} \left(\frac{\gamma_{d,u}}{\max(\gamma_{d,\emptyset}, \, \max_{v \in \mathcal{U}_d} \gamma_{d,v} \|W\|^{|v|})} \right)^s
\end{aligned}
$$

for some positive c_s. Since

$$\frac{\gamma_{d,u}}{\max(\gamma_{d,\emptyset}, \, \max_{v \in \mathcal{U}_d} \gamma_{d,v} \|W\|^{|v|})} \;\; \leq \;\; \frac{\widehat{\gamma}_{d,u}}{\min(1, \|W\|^{q^*})},$$

we conclude that

$$\text{card}(A_{d,\varepsilon}) \;\; \leq \;\; 1 + \frac{c_s}{\varepsilon^{2s}} \, \frac{\max(1, B^{sq^*})}{\min(1, \|W\|^{sq^*})} \sum_{u \in \mathcal{U}_d} \widehat{\gamma}^{\,s}_{d,u} \;\; \leq \;\; \widehat{c}_s \, \varepsilon^{-2s}$$

for some positive $\widehat{c}_s$ independent of d. Of course, when (51) holds for $s \leq p/2$ then it also holds for $s > p/2$, and the previous analysis applies. This proves that the problem is strongly tractable with the exponent at most $\max(p, 2s^*)$.

To complete the proof, assume that the problem is strongly tractable with the exponent of strong tractability p^*; hence the algorithms $A_{d,\varepsilon}$ are strongly polynomial-time. Then there exist positive α and $\widetilde{c}$ such that

$$\text{card}(A_{d,\varepsilon}) \leq \widetilde{c}\varepsilon^{-\alpha} \quad \forall \varepsilon, d. \tag{52}$$

Here $\alpha > p^*$, and can be made arbitrarily close to p^*. Obviously, $p^* \geq p$ since for $d = 1$ we have $\text{card}(A_{1,\varepsilon}) = \Theta(\varepsilon^{-p})$.

Recall that (49) states $\text{card}(A_{d,\varepsilon}) \geq \sum_{u \in \mathcal{U}_d} \left\lfloor (\gamma_{d,u}/(\varepsilon^2 \|\text{APP}_d\|^2))^{p/2} \right\rfloor$ for all ε and d. Since

$$\frac{\gamma_{d,u}}{\max(\gamma_{d,\emptyset}, \, \max_{v \in \mathcal{U}_d} \gamma_{d,v} \|W\|^{|v|})} \;\; \geq \;\; \frac{\widehat{\gamma}_{d,u}}{\max(1, \|W\|^{q^*})},$$

we conclude that
$$\sum_{u \in \mathcal{U}_d} \left\lfloor \left(\frac{\widehat{\gamma}_{d,u}}{\varepsilon^2 \max(1, \|W\|^{q^*})} \right)^{p/2} \right\rfloor \leq \widetilde{c} \varepsilon^{-\alpha} \quad \forall \varepsilon, d.$$

From this we obtain
$$\left| \left\{ u \in \mathcal{U}_d : \ \widehat{\gamma}_{d,u} \geq \varepsilon^2 \max(1, \|W\|^{q^*}) \right\} \right| \leq \widetilde{c} \varepsilon^{-\alpha} \quad \forall \varepsilon, d. \tag{53}$$

For $s > \alpha/2$ we have
$$\sum_{u \in \mathcal{U}_d} \widehat{\gamma}_{d,u}^s = \sum_{j=0}^{\infty} \sum_{u \in \mathcal{U}_d: \ 2^{-(j-1)} < \widehat{\gamma}_{d,u} \leq 2^{-j}} \widehat{\gamma}_{d,u}^s.$$

Setting $\varepsilon^2 = 2^{-j}/\max(1, \|W\|^{q^*})$ we obtain from (53)
$$\sup_d \sum_{u \in \mathcal{U}_d} \widehat{\gamma}_{d,u}^s \leq \widetilde{c} \max(1, \|W\|^{q^*})^{\alpha/2} \sum_{j=0}^{\infty} 2^{-j(s-\alpha/2)} < \infty.$$

Hence (51) holds with $2s > \alpha$. This implies that $2s^* \leq \alpha$, and since α can be arbitrarily close to p^*, we have $2s^* \leq p^*$, as needed. This completes the proof. □

It is interesting to compare conclusions of Theorems 3 and 4. Theorem 3 states that weighted approximation is tractable for *arbitrary* finite-order weights and
$$\operatorname{card}(A_{d,\varepsilon}) = O\left(\varepsilon^{-p(1+\delta)} d^{q^*} \right). \tag{54}$$

Theorem 4 states that for finite-order weights satisfying (51) with the smallest s denoted by s^*, the weighted approximation problem becomes strongly tractable and
$$\operatorname{card}(A_{d,\varepsilon}) = O\left(\varepsilon^{-p^*(1+\delta)} \right) \tag{55}$$

with the exponent of strong tractability $p^* = \max(p, 2s^*)$.

Consider two cases. First, assume that $s^* \leq p/2$. Then $p^* = p$ and, modulo δ, the weighted approximation problem for arbitrary d is roughly as hard as for $d = 1$. In this case, Theorem 4 is a clear improvement of Theorem 3 since the factor d^{q^*} present for arbitrary finite-order weights is eliminated for finite-order weights satisfying (51) with $s \approx s^*$, and with roughly the same dependence on ε^{-1}.

The second case, $s^* > p/2$, is more interesting. The exponent of strong tractability is now $p^* = 2s^* > p$, and the weighted approximation problem, although strongly tractable, is more difficult for $d \geq 2$ than for $d = 1$. Furthermore, we have a tradeoff:
$$\operatorname{card}(A_{d,\varepsilon}) = O\left(\min\left(\varepsilon^{-p(1+\delta)} d^{q^*}, \varepsilon^{-p^*(1+\delta)} \right) \right)$$

since both bounds (54) and (55) hold. Hence, for a fixed ε and varying d, we conclude that the bound (55) is better and we solve the weighted approximation problem with cost depending on the increased power of ε^{-1} independent of d. If, however, we fix d and vary ε, the opposite is true. The bound (54) is now better and we solve the weighted approximation problem with cost depending on the non-increased power of ε^{-1} at the expense of a polynomial dependence in d.

6 Extensions

In the previous sections we made some assumptions/restrictions to simplify the presentation. We now briefly comment how they can be relaxed.

6.1 Removing Assumption (3)

Suppose now that
$$1 \in H(K).$$

Due to separability of $H(K)$, there exist functions $\zeta_1, \zeta_2, \ldots \in H(K)$ that form an orthonormal system of the orthogonal complement of $\mathrm{span}(1)$ in $H(K)$. This means that

$$K(x,t) = c_\emptyset + \overline{K}(x,t) \quad \text{with} \quad \overline{K}(x,t) := \sum_{i=1}^{\dim(H(K))-1} \zeta_i(x)\zeta_i(t) \quad \text{and} \quad 1 \notin H(\overline{K}),$$

where $c_\emptyset = \|1\|_{H(K)}^{-2}$. Note that

$$K_d(\mathbf{x}, \mathbf{t}) = \gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u} K_{d,u}(\mathbf{x}, \mathbf{t}) = \gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u} \prod_{j \in u} (c_\emptyset + \overline{K}(x_j, t_j))$$

$$= \gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u} \left(c_\emptyset^{|u|} + \sum_{\emptyset \neq v \subseteq u} c_\emptyset^{|u|-|v|} \overline{K}_{d,v}(\mathbf{x}, \mathbf{t}) \right) = \overline{\gamma}_{d,\emptyset} + \sum_{v \in \mathcal{U}_d} \overline{\gamma}_{d,v} \overline{K}_{d,v}(\mathbf{x}, \mathbf{t}),$$

where

$$\overline{\gamma}_{d,\emptyset} := \gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u} c_\emptyset^{|u|}$$

and for $v \neq \emptyset$,

$$\overline{\gamma}_{d,v} := \sum_{u \in \mathcal{U}_{d,v}} \gamma_{d,u} c_\emptyset^{|u|-|v|} \quad \text{with} \quad \mathcal{U}_{d,v} = \{u \in \mathcal{U}_d : v \subseteq u\}.$$

30

We stress that if $\{\gamma_{d,u}\}$ are finite-order weights of order q^* then $\{\overline{\gamma}_{d,u}\}$ are also finite-order weights with the same order q^*. This means that tractability of the problem with K and $\{\gamma_{d,u}\}$ is equivalent to tractability of the problem with K replaced by $\overline{K}$ and $\gamma_{d,u}$ replaced by $\overline{\gamma}_{d,u}$. Moreover, using the construction presented in this paper for the kernel $\overline{K}$ and the weights $\overline{\gamma}_{d,u}$ we obtain linear algorithms which are polynomial-time for the kernel K and the weights $\gamma_{d,u}$. More precisely, the results of Section 3 hold if $\overline{K}(a,a) = 0$ for some $a \in D$, and the results of Sections 4 and 5 hold if $\int_{D^2} \rho(x)\,\rho(t)\,\overline{K}(x,t)\,\mathrm{d}(x,t) = 0$.

6.2   Assumption on Λ

In Section 3, we restricted the class Λ of permissible functionals to the class Λ^{std} or to the class Λ^{all}. However, all the results easily carry over for a more general class Λ as long as Λ enjoys the following tensor product property: if $L_1, \ldots L_d$ are from Λ for $d = 1$ then $L_1 \otimes \cdots \otimes L_d$ are from Λ for the d-dimensional case, and that evaluation of f at a is permissible for $d = 1$.

6.3   Replacing Assumption (12)

In Section 3, we assumed that there exists a point $a \in D$ for which $K(a,a) = 0$. Recall that this assumption is equivalent to assuming that all functions form $H(K)$ vanish at a. We now show how this assumption can be replaced by two others: (56) and (58) introduced below.

As in Section 3 we can assume, without loss of generality, the existence of algorithms $\{B_i\}_{i=1}^{\infty}$ satisfying (13)–(15). Additionally, we now assume that the algorithms B_i are exact for constant functions, i.e.,

$$B_i(1) \ = \ 1 \qquad \forall\, i \in \mathbb{N}_+. \tag{56}$$

Consider the algorithm $A_{d,\varepsilon}$ defined as in Section 3 with A_i's replaced by B_i's. That is,

$$A_{d,\varepsilon}f \ := \ f(\mathbf{a}) + \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \in Q(u)} \Delta_{u,\mathbf{i}}(f) \tag{57}$$

with

$$\Delta_{u,\mathbf{i}} \ := \ \bigotimes_{k=1}^{d} G_k(u,\mathbf{i}) \qquad \text{with} \qquad G_k(u,\mathbf{i}) := \left\{ \begin{array}{ll} B_1 & \text{if } k \notin u, \\ B_{i_k} - B_{i_k-1} & \text{if } k \in u, \end{array} \right.$$

for non-empty sets u, and $\Delta_{\emptyset} := \otimes_{k=1}^{d} B_1$. As before, the sets $Q(u) = Q(u,\varepsilon)$ are of the form

$$Q(u) = \left\{ \mathbf{i} \in \mathbb{N}^{|u|} \ : \ \mathbf{i} \geq \mathbf{2} \text{ and } |\mathbf{i}| \leq m(u) \right\};$$

however, now with a slightly changed $m(u) = m(u, \varepsilon)$ due to the fact that $A_i = B_i$ (and not B_{i-1} as before). That is, $m(u)$ is given by (23) and (24) with E_0 replaced by $E_0/2$ in the definition of y_u, i.e., now

$$y_u := \frac{\varepsilon^2 \, \|\mathrm{APP}_d\|^2}{c_{q^*}^2} \left(\frac{(4 \, D_0^{|u|})}{\gamma_{d,u} \, E_0^{2|u|}} \right)^{1/(1+p/2)} \frac{1}{\sum_{v \in \mathcal{U}_d} \left(\gamma_{d,v} (E_0/2)^{2|v|} \right)^{p/(2+p)} \, D_0^{|v|/(1+p/2)}} .$$

Due to representation $f = f_\emptyset(\mathbf{a}) + \sum_{v \in \mathcal{U}_d} f_v$ we have

$$f - A_{d,\varepsilon} f = \sum_{v \in \mathcal{U}_d} \sum_{u \in \mathcal{U}_d} \sum_{\mathbf{i} \in Q^c(u)} \Delta(u, \mathbf{i}) f_v .$$

Due to (56) we have

$$\Delta_{u, \mathbf{i}} f_v = 0 \quad \text{when} \quad u \setminus v \neq \emptyset .$$

This means that in the expression for $f - A_{d,\varepsilon} f$, we can restrict the sum over $u \in \mathcal{U}_d$ to the sum over $u \subseteq v$, i.e.,

$$f - A_{d,\varepsilon} f = \sum_{v \in \mathcal{U}_d} \sum_{\mathbf{i} \in Q^c(v)} \Delta_{v, \mathbf{i}} f + E(f),$$

where

$$E(f) := \sum_{v \in \mathcal{U}_d} \sum_{u \in \overline{\mathcal{U}}_{d,v}} \sum_{\mathbf{i} \in Q^c(u)} \Delta_{u, \mathbf{i}} f_v \quad \text{with} \quad \overline{\mathcal{U}}_{d,v} = \{ u \in \mathcal{U}_d : u \subset v, u \neq v \} .$$

We are ready to state the second assumption:

$$\forall d \in \mathbb{N}_+, \ \forall u, v \in \mathcal{U}_d, \quad \text{if} \quad u \subseteq v \quad \text{then} \quad u = v. \tag{58}$$

i.e.,

$$\overline{\mathcal{U}}_{d,v} = \emptyset \quad \forall d, \ \forall v \in \mathcal{U}_d .$$

Clearly, under this assumption, the term $E(f)$ above is zero and, therefore,

$$\|f - A_{d,\varepsilon} f\|_{L_{2,\rho_d}(D_d)} \leq \sum_{u \in \mathcal{U}_d} \sqrt{\gamma_{d,u}} \, \|f_u\|_{F_d} (2^{-1} \, E_0)^{|u|} \, 2^{-(m(u) - |u|)} \sum_{j=0}^{|u|-1} \binom{m(u) - |u|}{j} ,$$

and, as in Section 3, we conclude that for all d and $\varepsilon \in (0, 1)$ we have

$$e^{\mathrm{wor}}(A_{d,\varepsilon}; F_d) \leq \varepsilon \, \|\mathrm{APP}_d\| \quad \text{and} \quad e^{\mathrm{wor}}(\widehat{A}_{d,\varepsilon}; F_d) \leq \varepsilon \, \|S_d\| . \tag{59}$$

Of course, $\widehat{A}_{d,\varepsilon}$ is obtained from $A_{d,\varepsilon}$ as in Section 3. The rest of the analysis is very similar to the one in that section. This is why all the conclusions of Theorem 1 hold with the assumption (12) replaced by (56) and (58).

Note also that if $1 \in H(K)$, the modification discussed in Section 6.1 does not destroy the property (58) and gives new weights $\overline{\gamma}_{d,u}$ that are very simple:

$$\overline{\gamma}_{d,\emptyset} \;=\; \gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u}\, c_{\emptyset}^{|u|} \qquad \text{and} \qquad \overline{\gamma}_{d,u} \;=\; \gamma_{d,u} \quad \text{for} \quad u \neq \emptyset. \tag{60}$$

We stress that although the assumption (58) puts restrictions on the weights $\{\gamma_{d,u}\}$, it is satisfied by some interesting problems, including problems dealing with the Coulomb force and Coulomb potential functions, see Section 7.4.

6.4   More General Solution Operators than APP_d

The main results are obtained for the approximation problem $\{\mathrm{APP}_d\}$ and for problems $\{S_d\}$ that are related to the approximation via the assumption (9), i.e., that $\|S_d f\|_{G_d} \leq C_d \|f\|_{L_{2,\rho_d}(D_d)}$ for all f in F_d. Note that the embedding operators APP_d have a tensor product form, a property that was heavily used in all the proofs, whereas S_d does not need to have such a form.

We now show that the choice of APP_d as a *basic* problem is not necessary. The weighted approximation APP_d can be replaced by another tensor product problem as follows. As before, we assume without loss of generality that $1 \notin H(K)$. Consider a linear operator

$$T_1 \;:\; \mathrm{span}(1) \oplus H(K) \;\to\; G_1$$

with a Hilbert space G_1. For $d \geq 1$, define

$$T_d \;:=\; \bigotimes_{k=1}^{d} T_1 \qquad \text{and} \qquad G_d \;:=\; \bigotimes_{k=1}^{d} G_1 \tag{61}$$

with the norm of G_d given by

$$\left\| \bigotimes_{k=1}^{d} g_k \right\|_{G_d} \;=\; \prod_{k=1}^{d} \|g_k\|_{G_1} \quad \forall\, g_k \in G_1.$$

We can now modify (9) by assuming that the operators S_d are related to T_d by

$$\|S_d f\|_{G_d} \;\leq\; C_d\, \|T_d f\|_{G_d} \qquad \forall\, f \in F_d. \tag{62}$$

Observe that for $T_1 = \text{APP}_1$ the assumption (62) reduces to (9). In general, the choice of T_1 is arbitrary. Note, however, that if T_1 is a functional, e.g., $T_1 = \text{INT}_1$, then all T_d are functionals, and S_d satisfying the assumption (62) must be also a functional (possibly nonlinear).

We show how the assumption (62) can be used to obtain polynomial-time algorithms. As in Section 3, assume that $K(a, a) = 0$, and that there exist algorithms $\{B_i\}$ satisfying the assumptions (13), (14) and (15), of course with APP_1 replaced by T_1. Define the algorithms A_i as before with the only difference that now

$$A_1 f := f(a) \cdot T_1(1),$$

where 1 above stands for the constant function $f \equiv 1$. For finite-order weights of order q^*, we can assume without loss of generality that $T_1(1) \neq 0$, i.e.,

$$C_0 := \|T_1(1)\|_{G_1} > 0,$$

since otherwise $T_d(F_d) = \{0\}$ for $d > q^*$. Observe that now

$$\|\Delta_{u,\mathbf{i}}\| \leq C_0^{d-|u|} E_0^{|u|} 2^{-(|\mathbf{i}|-|u|)}.$$

Consider now the algorithm $A_{d,\varepsilon}$ given by (17) (18), and (23) with $\|T_d\|$ instead of $\|\text{APP}_d\|$ and $C_0^d (E_0/C_0)^{|u|}$ instead of $E_0^{|u|}$ in the definition of y_u. That is, $m(u)$ is given by (23) and (24) and

$$y_u := \frac{\varepsilon^2 \|T_d\|^2}{c_{q^*}^2 C_0^d} \left(\frac{D_0^{|u|}}{\gamma_{d,u} (E_0/C_0)^{2|u|}} \right)^{1/(1+p/2)} \frac{1}{\sum_{v \in \mathcal{U}_d} \left(\gamma_{d,v} (E_0/C_0)^{2|v|} \right)^{p/(2+p)} D_0^{|v|/(1+p/2)}}.$$

Then

$$e^{\text{wor}}(A_{d,\varepsilon}; F_d) \leq \varepsilon \|T_d\| \tag{63}$$

and

$$\text{card}(A_{d,\varepsilon}) \leq 1 + b_1 \left(b_2 \frac{|\mathcal{U}_d|}{\varepsilon^p} \left(\frac{\sum_{u \in \mathcal{U}_d} \gamma_{d,u}}{\|T_d\|^2} \right)^{p/2} + b_3 |\mathcal{U}_d| \right) \tag{64}$$

$$\times \left(\ln \left(\frac{b_2 |\mathcal{U}_d|}{\varepsilon^p} \left(\frac{\sum_{u \in \mathcal{U}_d} \gamma_{d,u}}{\|T_d\|^2} \right)^{p/2} + b_3 |\mathcal{U}_d| \right) + b_4 \right)^{(p+1)q^*}.$$

The constants b_i are given as in Section 3; however with E_0 replaced by E_0/C_0. Since

$$\|T_d\|^2 \geq \max_{u \in \mathcal{U} \cup \{\emptyset\}} \gamma_{d,u} C_0^{2(d-|u|)} \|T_1\|_{H(K) \to G_1}^{2|u|},$$

we conclude as in Theorem 1 (ii) that

$$\operatorname{card}(A_{d,\varepsilon}) \leq a_3 \, d^{q^*(1+p/2)} \, \varepsilon^{-p} \, (\ln(d/\varepsilon))^{(p+1)q^*} .$$

Moreover, if there exists a positive number C_1 such that

$$\|T_d\|^2 \geq C_1 \sum_{u \in \mathcal{U} \cup \{\emptyset\}} \gamma_{d,u} C_0^{2(d-|u|)} \|T_1\|_{H(K) \to G_1}^2, \tag{65}$$

then, as in Theorem 1 (i),

$$\operatorname{card}(A_{d,\varepsilon}) \leq a_1 \, |\mathcal{U}_d| \, \varepsilon^{-p} \, (\ln(|\mathcal{U}_d|/\varepsilon))^{(p+1)q^*} . \tag{66}$$

It can be shown that (65) holds with $C_1 = 1$ when T_1 is a functional. Hence (66) holds for all functionals T_1.

Of course, the modifications discussed in Sections 6.1-6.3 are also applicable to such general problems T_d.

Theorems 3 and 4 can also be extended for for $\Lambda = \Lambda^{\mathrm{all}}$ assuming that $T_1(1)$ is orthogonal to $T_1(H(K))$ which corresponds to and replaces the assumption (40). Then the eigenpairs of

$$W_d := T_d^* \circ T_d : F_d \to F_d$$

can be expressed in terms of the eigenpairs of $W = T_1^* \circ T_1 : H(K) \to H(K)$ similarly as it was done in Section 4. Indeed, let $\{\lambda_i, \zeta_i\}_i$ be the sequence of eigenpairs of W, with ordered λ_i and orthogonal ζ_i. Then

$$\zeta_{\mathbf{i},u} := \bigotimes_{k=1}^{d} \eta_k \quad \text{with} \quad \eta_k = \begin{cases} \zeta_{i_k} & \text{if } k \in u, \\ \\ T_1(1) & \text{if } k \notin u, \end{cases}$$

and $\zeta_{\emptyset} = \bigotimes_{k=1}^{d} T_1(1)$ are the eigenvectors of W_d corresponding to the eigenvalues

$$\lambda_{\mathbf{i},u} = \gamma_{d,u} \, C_0^{2(d-|u|)} \prod_{k \in u} \lambda_i.$$

Another extension of the results of this paper could be obtained by allowing the operators APP_d and/or T_d to be tensor products of operators that need not be identical, say

$$T_d = \bigotimes_{k=1}^{d} T_{d,k} \qquad \text{with} \qquad T_{d,k} \,:\, H(K_{d,k}) \,\rightarrow\, G_{d,k}.$$

Here $H(K_{d,k})$ for $k = 1, 2, \ldots, d$ are reproducing kernel Hilbert spaces, with possibly different kernels for different k, and $G_{d,k}$ are possibly different Hilbert spaces. Of course, the eigenvalues $\lambda_{\mathbf{i},u}$ are products of the corresponding eigenvalues of operators $T_{d,k}^* \circ T_{d,k}$. This extension does not make the analysis more difficult; however, it does result in a cumbersome notation. This *intractability* of the notation is the only reason why we have assumed the identical operators APP_1 (or T_1).

7   Applications

We illustrate results of this paper by a number of specific examples.

7.1   Weighted Approximation of Smooth Functions

Using the notation of Section 2, we now take $m = 1$ and $D = \mathbb{R}$. The kernel K is defined as in [39, 41]. That is, for a positive integer k, consider the following reproducing kernel

$$K(x,y) \;=\; \mathbf{1}_{\mathbb{R}_+}(x\,y) \int_0^{\infty} \frac{\left(|x|-t\right)_+^{k-1} \left(|y|-t\right)_+^{k-1}}{(\psi(t)\,(k-1)!)^2} \,\mathrm{d}t \qquad \forall\, x, y \in \mathbb{R},$$

where $\psi : \mathbb{R} \rightarrow \mathbb{R}_+$. Here $\mathbf{1}_{\mathbb{R}_+}$ denotes the characteristic function of $\mathbb{R}_+$ which implies that $K(x,y) = 0$ if $xy \leq 0$. Clearly $K(0,0) = 0$. The corresponding reproducing kernel Hilbert space is

$$H(K) \;=\; \Big\{ f : \mathbb{R} \rightarrow \mathbb{R} \,:\, f(0) = \ldots = f^{(k-1)}(0) = 0,$$
$$f^{(k-1)} \text{ is abs. cont., } \text{ and } \ f^{(k)}\psi \in L_2(\mathbb{R}) \Big\}$$

with the inner product given by

$$\langle f, g \rangle_{H(K)} \;=\; \int_{\mathbb{R}} f^{(k)}(x)\, g^{(k)}(x)\, \psi^2(x) \,\mathrm{d}x \qquad \forall\, f, g \in H(K).$$

The function ψ allows to have general spaces $H(K)$. For instance, letting ψ be the characteristic function of $[0,1]$ corresponds to the classical Sobolev space $H(K) = W_2^k([0,1])$.

When $\psi(x)$ converges sufficiently fast to zero with $|x| \to \infty$ then $H(K)$ is a much bigger space than when $\psi(x)$ converges slowly or does not converge at all, see again [39, 41], for more details.

Let ρ be a probability density function defined as in Section 2 for which (5) holds. Clearly, $\int_{\mathbf{R}^2} \rho(x)\rho(t)K(x,t)\,\mathrm{d}(x,t) > 0$. Under some additional conditions on (k, ψ, ρ), see [41] for details[3], we know that algorithms B_i based on a piecewise polynomial interpolation of degree $k - 1$, satisfy the assumptions of Theorem 1 with $p = 1/k$. Hence, the corresponding algorithms $\{A_{d,\varepsilon}\}$ are polynomial-time algorithms and

$$\mathrm{card}(A_{d,\varepsilon}) \ \leq \ a\, d^{q^*}\, \varepsilon^{-1/k}\, (\ln(d/\varepsilon))^{q^*(k+1)/k}\, .$$

7.2 Weighted Integration with Isotropic Kernel K

We now present an example with an isotropic kernel K and $m \geq 2$. To illustrate the approach outlined in Section 6, we discuss a multivariate integration problem with $T_1 = \mathrm{INT}_1$.

More specifically, let $m \geq 2$,

$$D \ = \ [0,1]^m \qquad \text{and} \qquad K(x,y) \ = \ \frac{\|x\|_2 + \|y\|_2 - \|x-y\|_2}{2},$$

where $\|\cdot\|_2$ is the ℓ_2 norm in $\mathbf{R}^m$. The corresponding reproducing kernel Hilbert space $H(K)$ was characterized in [5], see also [20]. Here we only mention that K is also a covariance kernel of the Brownian motion in Lévy's sense or, as it is sometimes called, of the isotropic Wiener stochastic process. Note that now $H(K)$ is an isotropic space of functions of m variables and it cannot be represented as a tensor product of m spaces of univariate functions.

Using the notation of the previous section, let

$$T_1(f) \ = \ \mathrm{INT}_1(f) \ := \ \int_D f(x)\,\mathrm{d}x.$$

Of course, $G_1 = \mathbf{R}$, and $\rho = 1$. Moreover, $\int_{D^2} K(t,x)\,\mathrm{d}(t,x) > 0$.

For $d \geq 2$, we want to approximate

$$T_d(f) \ = \ \mathrm{INT}_d(f) \ := \ \int_{D_d} f(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

with $f \in F_d$.

---

[3]For instance, this holds when $\rho(x) = \rho(-x)$, $\psi(x) = \psi(-x)$, $\psi(0) > 0$, ψ is non-increasing on $\mathbf{R}_+$ and continuous on it's support, as well as $\int_0^\infty \left(\sqrt{\rho(x)}/\psi(x)\right)^{1/k}\,\mathrm{d}x < \infty$.

Since T_d is a functional, it makes sense to consider only Λ^{std}. It follows from [36] that there exists a family of algorithms B_i satisfying (14) and (15) with $p = 2m/(m+1)$. We know that (65) holds since

$$\|\mathrm{INT}_d\|^2 \; = \; \gamma_{d,\emptyset} + \sum_{u \in \mathcal{U}_d} \gamma_{d,u} \|T_1\|^{2|u|} \quad \text{with} \quad \|T_1\|^2 \; = \; \int_{D^2} K(t,x)\, \mathrm{d}(t,x).$$

Hence, as in Theorem 1 (i), we conclude that the corresponding algorithms $\{A_{d,\varepsilon}\}$ are polynomial-time and

$$\mathrm{card}(A_{d,\varepsilon}) \; \leq \; a\, d^{q^*}\, \varepsilon^{-2m/(m+1)}\, \left(\ln(d/\varepsilon)\right)^{q^*(3m+1)/(m+1)}.$$

7.3   Weighted Approximation in the Supremum Norm

The range spaces of the operators defining specific problems were so far Hilbert spaces. We now show how some of the results can be extended to the case when the range space G_d is the Banach space $C(D_d)$ of continuous functions.

More specifically, consider the problem of approximating $f \in F_d$ by an algorithm A with the error measured in the following weighted supremum semi-norm,

$$\|f - Af\|_{G_d} \; := \; \sup_{\mathbf{x} \in D_d} |f(\mathbf{x}) - Af(\mathbf{x})|\, \rho_d(\mathbf{x}).$$

The space F_d is as before, $F_d = H(K_d)$ with K_d given by (8) and satisfying (12). It is easy to prove that the initial error $\|\mathrm{APP}_d\|$ is given by

$$\|\mathrm{APP}_d\| \; = \; \sup_{\mathbf{x} \in D_d} \rho_d(\mathbf{x}) \sqrt{K_d(\mathbf{x}, \mathbf{x})}.$$

This is why we assume that the supremum above is finite. We only consider algorithms using information from Λ^{std}.

Suppose now that $H(K)$ and ρ are such that they admit a sequence of algorithms B_i satisfying (15) and the following modified (14):

$$\lim_{i \to \infty} \|\mathrm{APP}_1 - B_i\|_{H(K) \to G_1} \; = \; 0 \qquad \text{and} \qquad \|B_i - B_{i-1}\|_{H(K) \to G_1} \; \leq \; E_0\, 2^{-i} \quad \forall\, i \geq 2, \quad (67)$$

There is a number of results that guarantee the existence of such algorithms, of course under some assumptions on K and ρ. For instance, for $m = 1$, K given in Subsection 6.1 with $D_d = \mathbb{R}^d$, ψ and ρ satisfying certain assumptions, see [39] for details[4], the conditions (67)

---

[4] This holds, for instance, if $o_\rho + \min(1/2, o_{1/\psi}) > k - 1/2$ and $o_\rho, o_{1/\psi} < \infty$, where $o_g := \sup\{\beta \in \mathbb{R} : \lim_{t \to \infty} g(t)\, t^\beta = 0\}$ measures the behavior of g at infinity.

and (15) hold with $p = 1/(k-1/2)$. Note also that the same bounds hold for an interpolation problem, where instead of the function f, one wants to approximate the value of f at a fixed point x. Of course, the interpolation problem is an instance of problems discussed in the previous section with the range space being a Hilbert space, more specifically just $\mathbb{R}$. Since this holds for an arbitrary x, we conclude that the corresponding algorithms $\{A_{d,\varepsilon}\}$ are polynomial-time and

$$\text{card}(A_{d,\varepsilon}) \leq a\, d^{q^*} \varepsilon^{-1/(k-1/2)}\, \left(\ln(d/\varepsilon)\right)^{q^*(2k+1)/(2k-1)}$$

7.4 Perturbed Coulomb Potential

As mentioned in the introduction, we now consider the approximation problem for functions given as small perturbation of the sum of Coulomb pair potentials,

$$f_\alpha(\vec{x}_1, \ldots, \vec{x}_\ell) = \sum_{1 \leq i < j \leq \ell} \frac{1}{\sqrt{\|\vec{x}_i - \vec{x}_j\|^2 + \alpha}} \quad \text{for} \quad \vec{x}_i \in \Omega \subseteq \mathbb{R}^3.$$

That is, f_α only depends on groups of two variables each being a 3-dimensional vector $x_i \in \mathbb{R}^3$. Letting $\mathbf{x} = [\vec{x}_1, \ldots, \vec{x}_\ell]$, we can view f_α as a function of $3 \cdot \ell$ scalar variables, i.e.,

$$d = 3\,\ell.$$

(Note that the perturbed Coulomb force $g_\alpha(\mathbf{x}) = \sum_{1 \leq i < j \leq \ell}(\|\vec{x}_i - \vec{x}_j\|^2 + \alpha)^{-1}$ is even simpler than f_α, and the analysis of this section may also be applied to g_α.)

For simplicity of presentation, we consider only bounded domain Ω and constant ρ. Since the function $g(t_1, \ldots, t_6) := (\alpha + \sum_{i=1}^{3}(t_i - t_{i+3})^2)^{-1/2}$, $t_i \in \mathbb{R}$, is infinitely many times differentiable, so is f_α. Therefore, the (perturbed) Coulomb potential function can be viewed as a member of spaces for many different reproducing kernels of various smoothness.

In what follows, for the sake of brevity, we will illustrate the results of this paper using a relatively low degree smoothness, although extensions to higher smoothness is straightforward. We will focus on two different approaches depending on the form of the domain Ω. The first is applicable if Ω is a Cartesian product of three identical subsets of $\mathbb{R}$, e.g., Ω is a cube in $\mathbb{R}^3$. The second approach is applicable for general domains Ω. As we will see, approximation of the Coulomb potential function is easier when Ω is a cube.

7.4.1 Approach 1: $\Omega = D^3$

Let $\Omega = D \times D \times D$, where, for simplicity of presentation, we assume that $D = [0,1]$. As for the kernel K we choose

$$K(x,t) = 1 + \min(x,t).$$

Recall that then $H(K) = W_2^1[0,1]$ is the Sobolev space and $\langle f, g \rangle_{H(K)} = f(0)\,g(0) + \int_0^1 f'(t)\,g'(t)\,\mathrm{d}t$.

The (perturbed) Coulomb potential function f_α belongs to $F_d = H(K_d)$ for

$$K_d(\mathbf{x}, \mathbf{y}) \;=\; \sum_{u \in \mathcal{U}_d} K_{d,u}(\mathbf{x}, \mathbf{y})$$

with $d = 3\,\ell$, $m = 1$, and the set $\mathcal{U}_d$ consisting of subsets

$$u \;=\; \{3i-2, 3i-1, 3i, 3j-2, 3j-1, 3j\} \quad \text{for} \quad i < j,$$

i.e., u contains indices of coefficients corresponding to two different vectors $\vec{x}_i$ and $\vec{x}_j$. More-over, we can set $\gamma_{d,u} = 1$ for $u \in \mathcal{U}_d$, and $\gamma_{d,u} = 0$ for $u \notin \mathcal{U}_d$. Of course, such weights are finite-order with $q^* = 6$ and

$$|\mathcal{U}_d| \;=\; \frac{(\ell-1)\ell}{2} \;<\; \frac{d^2}{18}.$$

Note also that $\mathcal{U}_d$ satisfies (58).

It is well-known that for the $d = 1$ case, the algorithms B_i based on piecewise constant interpolation at equally spaced points satisfy the assumptions (56) and (13)–(15) with $p = 1$ and $D_0 = E_0 = 1$. Due to (60), the initial error is at least as large as the cardinality of $\mathcal{U}_d$. Therefore, due to (29), the corresponding algorithm $\mathcal{A}_{d,\varepsilon}$ approximates f_α with error not exceeding $\varepsilon \, \|\mathrm{APP}_d\| \, \|f_\alpha\|_{F_d}$ and cost bounded by

$$\mathrm{card}(\mathcal{A}_{d,\varepsilon}) \;\leq\; a_1 \, |\mathcal{U}_d| \, \varepsilon^{-1} \, \ln^{12}(|\mathcal{U}_d|/\varepsilon) \;\leq\; a_2 \, d^2 \, \varepsilon^{-1} \, \ln^{12}(d/\varepsilon)$$

for positive constants a_i

We now estimate the norm $\|f_\alpha\|_{F_d}$. For that end, consider $g(\vec{x}, \vec{y}) = (\|\vec{x} - \vec{y}\|^2 + \alpha)^{-1/2}$. Let $\vec{z} = [z_1, \ldots, z_6]$ with $z_i = x_i$ for $i \leq 3$ and $z_i = y_{i-3}$ for $i > 3$. It is easy to verify that for a non-empty subset u of $\{1, 2, \ldots, 6\}$, we have

$$\frac{\partial^{|u|} g}{\prod_{j \in u} \partial z_j}(\vec{z}) \;=\; (-1)^{|u|} \prod_{j=1}^{|u|-1} (j + 1/2) \, \frac{\prod_{j \in u:\, j \leq 3}(x_j - y_j) \, \prod_{j \in u:\, j > 3}(y_{j-3} - x_{j-3})}{(\|\vec{x} - \vec{y}\|^2 + \alpha)^{|u|+1/2}},$$

with the convention that the product over the empty set is taken as one. Consider now $u = \{1, \ldots, 6\}$ and

$$a := \int_{[0,1]^6} \left(\frac{\partial^6}{\prod_{j=1}^6 \partial z_j} g(\vec{z}) \right)^2 \mathrm{d}\vec{z} = \Theta \left(\int_{[0,1]^6} \frac{(z_1 - z_4)^4 (z_2 - z_5)^4 (z_3 - z_6)^4}{\left((z_1 - z_4)^2 + (z_2 - z_5)^2 + (z_3 - z_6)^2 + \alpha \right)^{13}} \right) \mathrm{d}\vec{z}.$$

40

By changing variables $z_i = t_i \sqrt{\alpha}$ we immediately conclude that $a = \Theta(\alpha^{-4})$ as $\alpha \to 0$. This implies that

$$\|f_\alpha\|_{F_d} = d \cdot \Theta(\alpha^{-2}) \quad \text{as} \quad \alpha \to 0$$

with the factor in the Θ-notation independent of d.

Hence, for $\hat{\varepsilon} = \varepsilon \|f_\alpha\|_{F_d}^{-1}$, we obtain $\|f_\alpha - A_{d,\hat{\varepsilon}} f_\alpha\| \leq \varepsilon \|\mathrm{APP}_d\|$ with

$$\mathrm{card}(A_{d,\hat{\varepsilon}}) \leq a_3 \frac{d^3}{\varepsilon \, \alpha^2} \ln^{12}(d/\varepsilon)$$

for a positive number a_3.

7.4.2   Approach 2: General Ω

Let $\Omega \subset \mathbf{R}^3$ be Lebesgue measurable and bounded. Without loss of generality, we can assume that its Lebesgue measure is one. For the kernel K we choose

$$K(\vec{x}, \vec{y}) = 1 + \frac{\|\vec{x}\|_2 + \|\vec{y}\|_2 - \|\vec{x} - \vec{y}\|_2}{2} \quad \text{for} \quad \vec{x}, \vec{y} \in \Omega.$$

Then f_α belongs to $F_d = H(K_d)$ for

$$K_d(\mathbf{x}, \mathbf{y}) = \sum_{u \in \mathcal{U}_d} K_{d,u}(\mathbf{x}, \mathbf{y})$$

with $d = \ell$ and $m = 3$. The set $\mathcal{U}_d$ consists of sets $u = \{i, j\}$ for $1 \leq i < j \leq \ell$, i.e.,

$$|\mathcal{U}_d| = \frac{(\ell - 1)\ell}{2} < \frac{d^2}{2}.$$

We again take the weights $\gamma_{d,u} = 1$ for $u \in \mathcal{U}_d$, and $\gamma_{d,u} = 0$ otherwise. They are finite-order weights with $q^* = 2$, and $\mathcal{U}_d$ satisfies (58).

As follows from Section 7.2 of [36], there exist piecewise constant algorithms B_i that satisfy the assumptions (56) and (13)–(15) with $p = 3/2$ and $D_0 = 1$, however now $E_0 = E_0(\Omega)$ depends on Ω. As in Section 7.4.1, the initial error is proportional to the cardinality of $\mathcal{U}_d$. Therefore, the corresponding algorithm $\mathcal{A}_{d,\varepsilon}$ approximates f_α with error not exceeding $\varepsilon \|\mathrm{APP}_d\| \|f_\alpha\|_{F_d}$ and cost bounded by, see (29),

$$\mathrm{card}(\mathcal{A}_{d,\varepsilon}) \leq a_1 |\mathcal{U}_d| \varepsilon^{-3/2} \ln^5(|\mathcal{U}_d|/\varepsilon) \leq a_2 d^2 \varepsilon^{-3/2} \ln^5(d/\varepsilon)$$

for a positive constants a_i.

We now estimate the norm $\|f_\alpha\|_{F_d}$. For that end, we need to recall the following result from [5, 20]. Let $\hat{H}$ be the Hilbert space of functions defined on $\mathbb{R}^m$ generated by the reproducing kernel $\hat{K}$, where $\hat{K}$ is the natural extension of K from Ω^2 to $\mathbb{R}^{2m}$. Then for $g \in H(K)$ we have

$$\|g\|_{H(K)} \;=\; c_m \inf \left\{ L(f) \;:\; f \in \hat{H}, f|_\Omega = g \right\}.$$

Here $L(f) = \|\Delta^{(m+1)/4} f\|_{L_2}$ with the Laplacian operator Δ and a positive constant $c_m = c_m(\Omega)$. Since in our case $m = 3$, the norm $\|g\|_{H(K)}$ of $g \in H(K)$ is proportional to the $L_2(\Omega)$ norm of Δg. Consider now $g(\vec{x}, \vec{y}) = (\|\vec{x} - \vec{y}\|_2^2 + \alpha)^{-1/2}$. It can be checked that

$$\Delta_{\vec{x}} \Delta_{\vec{y}} g(\vec{x}, \vec{y}) \;=\; 15\,\alpha \left(\frac{3\,\alpha}{(\|\vec{x} - \vec{y}\|_2^2 + \alpha)^{9/2}} - \frac{4}{(\|\vec{x} - \vec{y}\|_2^2 + \alpha)^{7/2}} \right).$$

Using a similar change of variables as in the previous section, it is easy to show that $\|\Delta_{\vec{x}} \Delta_{\vec{y}} g\|^2_{L_2(\Omega^2)}$ equals $\Theta(\alpha^{-2})$ as $\alpha \to 0$. Therefore

$$\|f_\alpha\|_{F_d} \;=\; d \cdot \Theta(\alpha^{-1}) \quad \text{as} \quad \alpha \to 0$$

with the factor in Θ-notation depending only on Ω.

Hence, for $\hat{\varepsilon} = \varepsilon \, \|f_\alpha\|_{F_d}^{-1}$, we obtain $\|f_\alpha - A_{d,\hat{\varepsilon}} f_\alpha\| \le \varepsilon \, \|\text{APP}_d\|$ with

$$\text{card}(A_{d,\hat{\varepsilon}}) \;\le\; a_3 \, \frac{d^{3.5}}{\varepsilon^{1.5} \, \alpha^{1.5}} \, \ln^5(d/\varepsilon)$$

with $a_3 = a_3(\Omega)$ depending on Ω.

Acknowledgments

References

[1] N. Aronszajn, Theory of reproducing kernels, *Trans. AMS* **68** (1950), 337-404.

[2] K. I. Babenko, Approximation by trigonometric polynomials in a certain class of periodic functions of several variables, *Soviet. Math. Dokl.* **1** (1960), 672-675.

[3] G. Baszenski, F. J. Delvos, and S. Jester, Blending approximations with sine functions, *Numerical Methods of Approximation* **9** (D. Braess and L. L. Schumaker, eds.) Int. Ser. Num. Math. **105** (1992), 1-19.

[4] R.E. Caflisch, W. Morokoff and A. Owen, Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *J. Comp. Finance* **1** (1997), 27-46.

[5] Z. Ciesielski, On Lévy's Brownian motion with several-dimensional time, *Lecture Notes in Mathematics* **472** (A. Dold and B. Eckman, eds.), 29-56, Springer-Verlag, New York. 1975.

[6] F. J. Delvos, *d*-variate Boolean approximation, *J. of Approximation Theory* **34** (1982), 99-114.

[7] R. A. DeVore, S. V. Konyagin, and V. N. Temlyakov, Hyperbolic wavelet approximation, *Constructive Approximation* **14** (1998), 1-26.

[8] J. Dick, I.H. Sloan, X. Wang and H. Woźniakowski, Good lattice rules in weighted Korobov spaces with general weights, submitted, 2003.

[9] B. Efron and C. Stein, The Jackknife estimate of variance, *Ann. Stat.* **9** (1981), 586-596.

[10] J. Glimm and A. Jaffe, Quantum Physics, Springer Verlag, New York, second edition, 1987.

[11] S. Gal and C.A. Micchelli, Optimal sequential and non-sequential procedures for evaluating a functional, *Appl. Anal.* **10** (1980), 105-120.

[12] T. Gerstner and M. Griebel. Dimension-adaptive tensor-product quadrature, *Computing* **71** (2003), 65-87.

[13] M. Griebel and W. Knapek, Optimized tensor-product approximation spaces, *Constructive Approximation* **16** (2000), 252-540.

[14] M. Griebel, P. Oswald, and T. Schhiekofer, Sparse grids fro boundary integral equations, *Numer. Math.* **83** (1999), 279-312.

[15] F. J. Hickernell, I. H. Sloan, and G. W. Wasilkowski, On tractability of weighted integration for certain Banach spaces of functions, *Monte Carlo and Quasi-Monte Carlo Methods 2002*, (H. Niederreiter, ed.), Springer, 51-71, 2004.

[16] F.Y. Kuo and I.H. Sloan, Quasi-Monte Carlo methods can be efficient for integration over product spheres, Submitted, 2003.

[17] M. Kwas, Complexity of multivariate Feynman-Kac path integration in randomized and quantum settings, in progress.

[18] M. Kwas and Y. Li, Worst case complexity of multivariate Feynman-Kac path integration, *J. Complexity* **19** (2003), 730-743.

[19] Y. Li, Applicability of Smolyak's algorithm to certain Banach spaces of functions, *J. Complexity* **18** (2002), 792-814.

[20] G.M. Molchan, On some problems concerning Brownian motion in Lévy's sense, *Theory Probab. Appl.* **12** (1967), 682-690.

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

[22] E. Novak and H. Woźniakowski, When are integration and discrepancy tractable?, Foundation of Computational Mathematics, Oxford, 1999, R. A. DeVore, A. Iserles and E. Süli eds., *Cambridge University Press*, Cambridge, 211-266, 2001.

[23] A. Papageorgiou and G.W. Wasilkowski, On the average complexity of multivariate problems, *J. Complexity* **6** (1990), 1-23.

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

[25] L. Plaskota and G. W. Wasilkowski, The exact exponent of sparse grid quadratures in the weighted case, *J. of Complexity* **17** (2001), 840-849.

[26] L. Plaskota and G.W. Wasilkowski, Smolyak's algorithm for integration and L_1-approximation of multivariate functions with bounded mixed derivatives of second order, *Numerical Algorithms* **36** (2004), 229-246.

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

[28] S.A. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, *Dokl. Akad. Nauk SSSR*, (1963), 240-243.

[29] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation* **55** (2001), 271-280.

[30] V. N. Temlyakov, *Approximation of Periodic Functions,* Nova Science Publishers, Inc., New York, 1993.

[31] V. N. Temlyakov, On approximate recovery of functions with bounded mixed derivative, functions of several variables, *Soviet. Math. Dokl.* **1** (1960), 672-675.

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

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

[34] X. Wang and K.T. Fang, Effective dimensions and quasi-Monte Carlo integration, *J. Complexity* **19** (2003), 101-124.

[35] X. Wang and I.H. Sloan, Why are high-dimensional finance problems often of low effective dimension? submitted, 2003.

[36] G.W. Wasilkowski, Integration and approximation of multivariate functions: average case complexity with isotropic Wiener measure, *J. Approximation Theory* **77** (1993), 212-227.

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

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

[39] G.W. Wasilkowski and H. Woźniakowski, Complexity of weighted approximation over $\mathbb{R}^1$, *J. Approximation Theory* **103** (2000), 223-251.

[40] G.W. Wasilkowski and H. Woźniakowski, On the power of standard information for weighted approximation, *Found. Comput. Math.* **1** (2001), 417-434.

[41] G.W. Wasilkowski and H. Woźniakowski, Tractability of approximation and integration for weighted tensor product problems over unbounded domains, *Monte Carlo and Quasi-Monte Carlo Methods 2000*, (K.-T.Fang, F.J.Hickernell, H. Niederreiter, eds.), pp.497-522, Springer, 2002.

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

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

[44] H. Woźniakowski, Tractability and Strong Tractability of Multivariate Tensor Product Problems, *J. of Computing and Information* **4** (1994), 1-19.