

Approximation of High-Dimensional Rank One Tensors *

Markus Bachmayr, Wolfgang Dahmen,
Ronald DeVore, and Lars Grasedyck

March 14, 2013

Abstract

Many real world problems are high-dimensional in that their solution is a function which depends on many variables or parameters. This presents a computational challenge since traditional numerical techniques are built on model classes for functions based solely on smoothness. It is known that the approximation of smoothness classes of functions suffers from the so-called ‘curse of dimensionality’. Avoiding this curse requires new model classes for real world functions that match applications. This has led to the introduction of notions such as sparsity, variable reduction, and reduced modeling. One theme that is particularly common is to assume a tensor structure for the target function. This paper investigates how well a rank one function $f(x_1, \dots, x_d) = f_1(x_1) \cdots f_d(x_d)$, defined on $\Omega = [0, 1]^d$, can be captured through point queries. It is shown that such a rank one function with component functions f_j in $W_\infty^r([0, 1])$, can be captured (in L_∞) to accuracy $O(C(d, r)N^{-r})$ from N well chosen point evaluations. The constant $C(d, r)$ scales like d^{dr} . The queries in our algorithms have two ingredients, a set of points built on the results from discrepancy theory and a second adaptive set of queries dependent on the information drawn from the first set. Under the assumption that a point $z \in \Omega$ with non-vanishing $f(z)$ is known, the accuracy improves to $O(dN^{-r})$.

Key words and phrases: query algorithms, high-dimensional approximation, separable functions, rate of approximation

AMS Subject Classification: 41A25, 65D15

*This research was supported by the Office of Naval Research Contracts ONR-N00014-08-1-1113, ONR N00014-09-1-0107, and ONR N00014-11-1-0712; the AFOSR Contract FA95500910500; the NSF Grants DMS-0810869, and DMS 0915231 and the DFG Special Priority Program SPP-1324. This research was done when R. D. was a visiting Professor at RWTH and the AICES Graduate Program. This publication is based in part on work supported by Award No. KUS-C1-016-04 made by King Abdullah University of Science and Technology (KAUST)

1 Introduction

A recurring model in certain high-dimensional application domains is that the target function is a low rank tensor, or can be approximated well by a linear combination of such tensors. For an overview of numerical methods based on this concept and their applications, we refer to [3] and the references therein. We consider a fundamental question concerning the computational complexity of such low rank tensors: If we know that a given function has such a tensor structure, to what accuracy can we approximate it using only a certain number of deterministically chosen point queries? In this paper, we treat this problem in the simplest setting where the tensors are of rank one.

Given an integer r , we denote by $W_\infty^r[0, 1]$ the set of all univariate functions on $[0, 1]$ which have r weak derivatives in L_∞ , with the semi-norm

$$|f|_{W_\infty^r[0,1]} := \|f^{(r)}\|_{L_\infty}. \quad (1.1)$$

We shall study the following classes of rank one tensor functions defined on $\Omega := [0, 1]^d$. If r is a positive integer and $M > 0$, we consider the class of functions

$$\mathcal{F}^r(M) := \left\{ f \in C(\Omega) : f(x) = \prod_{i=1}^d f_i(x_i) \text{ with } \|f_i\|_{L_\infty[0,1]} \leq 1, |f_i|_{W_\infty^r[0,1]} \leq M, i = 1, \dots, d \right\}.$$

Note that we could equally well replace the bound 1 appearing in the definition by an arbitrary positive value and arrive at the above class by simple rescaling. Note also that whenever $\|f\|_{L_\infty(\Omega)} \leq 1$, we can achieve the restriction on the $\|f_i\|_{L_\infty[0,1]}$ in this definition by choosing a scaling of the individual factors so that $\|f_i\|_{L_\infty[0,1]} \leq 1$ for all i .

Let us note at the outset that \mathcal{F}^r is closely related to a class of functions with bounded mixed derivatives. We use the notation $D^\nu = D_{x_1}^{\nu_1} \cdots D_{x_d}^{\nu_d}$ for multivariate derivatives. Then, the class of functions $MW^r(L_\infty)$ consists of all functions $f(x_1, \dots, x_d)$ for which

$$|f|_{MW^r(L_\infty(\Omega))} := \sum_{\nu \in \Lambda_r \setminus \{0\}} \|D^\nu f\|_{L_\infty(\Omega)} < \infty, \quad (1.2)$$

where $\Lambda_r := \{\nu = (\nu_1, \dots, \nu_d) : 0 \leq \nu_i \leq r, i = 1, \dots, d\}$. We define the norm on this space by adding $\|f\|_{C(\Omega)}$ to the above semi-norm. This is a well studied class of functions, especially for the analysis of cubature formulae. Clearly, we have that $\mathcal{F}^r(M)$ is contained in a finite ball of $MW^r(L_\infty(\Omega))$ (see Chapters III and V of [7]). It is known that [1, Lemma 4.9] one can sample functions in $MW^r(L_\infty(\Omega))$ on a set of points (called sparse grids) with cardinality N and use these point values to construct an approximation to f with accuracy $C(d, r)\|f\|_{MW^r(L_\infty)} N^{-r} [\log N]^{(r+1)(d-1)}$ in $L_\infty(\Omega)$.

The main result of the present paper is to present a query algorithm for functions $f \in \mathcal{F}^r$. The query algorithm works without knowledge of M , but would require a bound on r . We show that we can query such a function f at $O(N)$ suitably chosen points and from these queries we can construct an approximation \tilde{f}_N that approximates f to accuracy $C(r, d)N^{-r}$. Thus, for rank one tensors, the $[\log N]^{(r+1)(d-1)}$ appearing for mixed norm classes can be removed. Moreover, \tilde{f}_N is again separable, that is, the algorithm preserves this structural property of the original function f .

Given a budget N , our queries of f will have two stages. The first queries of f occur at a set of $O(N)$ points built from discrepancy theory. If $f(z) \neq 0$ for one of the points z of the initial query then we continue and sample f at $O(N)$ points built from z . We then show how to build an approximation \tilde{f}_N to f from these query values which will provide the required accuracy.

2 Univariate approximation

Our construction of approximations of multivariate functions in $\mathcal{F}^r(M)$ is based on the approximation of univariate functions. It is well known that for $g \in W_\infty^r[0, 1]$, given the values $g(i/N)$, we can construct an approximation $\mathcal{I}_N((g(i/N))_{i=1}^N)$ that satisfies

$$\|g - \mathcal{I}_N((g(i/N))_{i=1}^N)\|_{L_\infty[0,1]} \leq C_1(r) \min\{\|g\|_{L_\infty[0,1]}, |g|_{W_\infty^r[0,1]} N^{-r}\}, \quad N = 1, 2, \dots \quad (2.1)$$

There are many ways to construct such an approximation operator \mathcal{I}_N . One is to use a quasi-interpolation operator built on univariate splines of order r . Another is to simply take for each interval $I = [j - 1/N, j/N)$, $j = 1, \dots, N$, a set S_j of r consecutive integers $i + 1, \dots, i + r$ that contain $j - 1$ and j , and then define g on the interval I as the polynomial of order r that interpolates g at the points in S_j .

In going further, we use any such construction of an operator \mathcal{I}_N . We note that \mathcal{I}_N needs as input any vector $y = (y_0, \dots, y_N)$. The y_i are usually taken as function values such as $y_i = g(i/N)$ above.

We need a second result about univariate functions summarized in the following lemma.

Lemma 2.1 *Suppose $g \in W_\infty^r[0, 1]$ is a univariate function that vanishes at r points $t_1, \dots, t_r \in [0, 1]$. If J is the smallest interval that contains all of the t_j , $j = 1, \dots, r$, then*

$$|g(t)| \leq \|g^{(r)}\|_{L_\infty[0,1]} (|J| + \text{dist}(t, J))^r, \quad t \in [0, 1]. \quad (2.2)$$

Proof: Note that each weak derivative $g^{(k)}$ for $k = 0, \dots, r - 1$ is in $W_\infty^1[0, 1]$, and can thus be identified with a continuous function. From Rolle's theorem, for each $k = 0, \dots, r - 1$, there is a point ξ_k in J such that $g^{(k)}(\xi_k) = 0$. This gives the bound

$$|g^{(r-1)}(t)| \leq \|g^{(r)}\|_{L_\infty[0,1]} |t - \xi_{r-1}| \leq \|g^{(r)}\|_{L_\infty[0,1]} (|J| + \text{dist}(t, J)), \quad t \in [0, 1]. \quad (2.3)$$

From this, we obtain the bound

$$\begin{aligned} |g^{(r-2)}(t)| &\leq \|g^{(r)}\|_{L_\infty[0,1]} (|J| + \text{dist}(t, J)) |t - \xi_{r-2}| \\ &\leq \|g^{(r)}\|_{L_\infty[0,1]} (|J| + \text{dist}(t, J))^2, \quad t \in [0, 1]. \end{aligned} \quad (2.4)$$

Continuing in this way, we arrive at (2.2). ■

3 Low-discrepancy point sequences

The first set of query points that we shall employ is a low-discrepancy sequence that is commonly used in quasi-Monte Carlo methods for high-dimensional integration. Roughly speaking, stopping at any place in the sequence gives a well scattered set of points in Ω . The particular property we are interested in here is that no d -dimensional rectangle contained in Ω can have large measure without containing at least one of these points. We shall adopt a method for constructing such a sequence given in [4, 5] which rests on base q expansions. For any prime number q and any positive integer n , we have a unique base q representation

$$n = \sum_{j \geq 0} b_j q^j, \quad b_j = b_j(q, n) \in \{0, \dots, q-1\}.$$

The b_j are the ‘bits’ of n in base q . For any $n < q^m$, this sequence has all zero entries in positions $j \geq m$.

With the bit sequence $(b_j) = (b_j(n))$ in hand, we define

$$\gamma_q(n) := \sum_{j \geq 0} b_j q^{-j-1}.$$

If q is fixed, the set of points $\Gamma_q(m) := \{\gamma_q(n) : 1 \leq n < m\}$ are in $(0, 1)$, and any point $x \in (0, 1)$ satisfies

$$\text{dist}(x, \Gamma_q(m)) \leq q/m. \quad (3.1)$$

Indeed, if $m = q^k$ for some positive integer k , then $\Gamma_q(m)$ contains all points j/m , $j = 1, \dots, m-1$ and so the distance in (3.1) does not exceed $1/m$. The general result for arbitrary m follows from this.

Definition 3.1 (Halton sequence) *Given the space dimension $d \geq 1$, we choose the first d prime numbers p_1, \dots, p_d . The sequence of points $(\hat{x}_k)_{k \in \mathbb{N}}$ in $[0, 1]^d$ is then defined by*

$$\hat{x}_k := (\gamma_{p_1}(k), \dots, \gamma_{p_d}(k)). \quad (3.2)$$

The following theorem (see [6] and [2]) shows that this sequence of points is well scattered in the sense that we need.

Theorem 3.2 *Let \hat{x}_k , $k = 1, 2, \dots$, be defined as in (3.2). For any d -dimensional rectangle $R = (\alpha_1, \beta_1) \times \dots \times (\alpha_d, \beta_d)$ with $0 \leq \alpha_i < \beta_i \leq 1$ that does not contain any of the points \hat{x}_k , $k = 1, \dots, N$, we have the following bound for the measure $|R|$ of R :*

$$|R| \leq \frac{C_H(d)}{N}, \quad (3.3)$$

where $C_H(d) := 2^d \prod_{i=1}^d p_i$.

Proof: For completeness, we give the short proof of this lemma. We first consider any d -dimensional rectangle $R_0 \subset \Omega$ of the form

$$R_0 := I_1 \times \dots \times I_d, \quad I_i := p_i^{-\nu_i} [t_i, (t_i + 1)), \quad i = 1, \dots, d, \quad (3.4)$$

where the $\nu_i \in \mathbb{N}$ and satisfy $p_1^{\nu_1} \cdots p_d^{\nu_d} \leq N$ and the t_i are positive integers. Such a rectangle obviously has volume $\geq 1/N$. We shall show that such a rectangle always contains a point \hat{x}_k for some $1 \leq k \leq N$ and thus obtain the theorem for rectangles of this special type.

Since $R_0 \subset \Omega$, each t_i is in $\{0, \dots, p_i^{\nu_i} - 1\}$ and therefore has a unique expansion

$$t_i = \sum_{j=0}^{\nu_i-1} a_{i,j} p_i^j$$

with $a_{i,j} \in \{0, \dots, p_i - 1\}$. We introduce the integers

$$m_i := \sum_{j=0}^{\nu_i-1} a_{i,\nu_i-j-1} p_i^j, \quad i = 1, \dots, d, \quad (3.5)$$

which satisfy

$$\gamma_{p_i}(m_i) = t_i p_i^{-\nu_i}, \quad i = 1, \dots, d.$$

From the Chinese remainder theorem, there is an integer $k < p_1^{\nu_1} \cdots p_d^{\nu_d} \leq N$ such that

$$k \equiv m_i \pmod{p_i^{\nu_i}}, \quad i = 1, \dots, d. \quad (3.6)$$

It follows that

$$\gamma_{p_i}(k) = t_i p_i^{-\nu_i} + \epsilon_i, \quad i = 1, \dots, d,$$

where $0 \leq \epsilon_i < p_i^{-\nu_i}$, $i = 1, \dots, d$. Therefore $\hat{x}_k = (\gamma_{p_1}(k), \dots, \gamma_{p_d}(k))$ is in R_0 and we have proven the theorem in this special case.

We now consider the general rectangle R in the statement of the theorem. We claim that R contains a special rectangle R_0 of the form (3.4) of volume larger than $C_H(d)^{-1}|R|$. Indeed, for the given $\alpha_i < \beta_i$, we define ν_i to be the smallest integer such that there exists an integer t_i with $[t_i p^{-\nu_i}, (t_i + 1) p^{-\nu_i}] \subset (\alpha_i, \beta_i)$. Then, $\beta_i - \alpha_i < 2p^{-\nu_i+1}$, since otherwise ν_i would not be minimal. This means that R contains a special rectangle R_0 with volume $|R_0| \geq C_H(d)^{-1}|R|$. Since R does not contain any of the \hat{x}_k , $k = 0, \dots, N$, the same is true of R_0 . Hence $|R_0| \leq N^{-1}$ and so $|R| \leq C_H(d)N^{-1}$. ■

4 Query points and the approximation

We now describe our query points. These will depend on r . If $r = 1$, then given our budget N of queries, it would be sufficient to simply query f at the points $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_N$ in succession. However, when $r > 1$, we will *occasionally* have to query f at a cloud of points near each \hat{x}_k in order to take advantage of the higher smoothness of f . We fix $r \geq 1$ in what follows. We next describe the cloud of points where we might query f . We define for each $k = 1, 2, \dots$, and each $n \geq k$,

$$\Gamma_n(\hat{x}_k) := \left\{ \hat{x}_k + \sum_{i=1}^d \frac{j_i}{r 2^n} e_i : j_i \in \{-r+1, \dots, 0, \dots, r-1\} \right\} \cap \Omega, \quad (4.1)$$

where e_i , $i = 1, \dots, d$, is the usual coordinate basis for \mathbb{R}^d . For each k, n , this set contains at most $(2r - 1)^d$ points and contains at least r^d points. When asked to query f at one of the sets $\Gamma_n(\hat{x}_k)$, we traverse these points in lexicographic order.

Our query algorithm given below will first sample f at point clouds $\Gamma_{n_k}(\hat{x}_k)$, $k = 1, \dots$. If we stipulate the budget N in advance, we can then fix the n_k . However, we would like this part of the sampling to be progressive where if the budget N changes, one still utilizes the previous samples. For this reason, we will occasionally update the assignment of n_k .

Given a function f and a budget index N , we proceed to query f as follows.

Query 1:

Step 1: We ask for the value of f at the points in $\Gamma_1(\hat{x}_1)$.

Step k : We sample f at the points in $\Gamma_k(\hat{x}_k)$. We also return to each $j < k$ where we have already sampled f at the point cloud $\Gamma_n(\hat{x}_j)$ for certain n . If the largest value $n = n_j$ where we have done this sampling satisfies $2^{n_j} < k$, then we sample f at the additional points in $\Gamma_k(\hat{x}_j)$.

Stopping criteria: If in the process of doing this sampling, we arrive at a point z for which $f(z) \neq 0$, then the querying stops and we go directly to **Query 2**. If it has not stopped earlier, we stop the querying when $k = N$.

Note that the updated queries in **Query 1** occur very infrequently.

Query 2: If $f(z) \neq 0$ for the stopping point z of **Query 1**, then for this z , we define z^j as the vector which agrees with z in all but the j -th coordinate and is zero in the j -th coordinate. We ask for the value of f at the points

$$\tilde{z}_{j,i} := z^j + \frac{i}{N}e_j, \quad i = 1, \dots, N, \quad j = 1, \dots, d. \quad (4.2)$$

We define $\Lambda_N(f)$ as the set of points where we have sampled f . We want next to bound the cardinality of $\Lambda_N(f)$. Since $\#(\Gamma_n(\hat{x}_k)) \leq (2r - 1)^d$, for all choices n, k , the only issue in bounding the number of samples in **Query 1** will be how many times we have resampled f near \hat{x}_j . Now, for a given \hat{x}_j , we originally sample f at the points $\Gamma_j(\hat{x}_j)$. This sampling will be updated to a sampling $\Gamma_{2^j}(\hat{x}_j)$ if $2^j < N$. It will be updated again if $2^{2^j} < N$ and so on. It follows that the only \hat{x}_j whose sampling is updated are those with $j \leq \log_2 N$ and the maximum number of times it is updated is bounded by $\log_2 N$. Thus, the total number of samples taken in **Query 1** does not exceed $(2r - 1)^d[N + (\log_2 N)^2] \leq 2 \cdot (2r - 1)^d N$. This gives that the total number of samples taken is

$$\#(\Lambda_N(f)) \leq C_1(d, r) N, \quad C_1(d, r) := 2(2r - 1)^d + d. \quad (4.3)$$

We now describe how we define the approximation \tilde{f}_N to f constructed from these query values.

Case 1: If f vanishes at each of the query points in **Query 1** then we define \tilde{f}_N to be identically zero on Ω .

Case 2: If $f(z) \neq 0$ for the stopping point z of **Query 1**, then we define

$$F_j := \mathcal{I}_N(f(\tilde{z}_{j,i})_{i=1}^N), \quad j = 1, \dots, d,$$

where \mathcal{I}_N is the operator of §2. Then, with $A := f(z)$, we define our approximation to f as

$$\tilde{f}_N(x) := A^{-d+1}F_1(x_1) \cdots F_d(x_d). \quad (4.4)$$

5 Error of approximation

We now analyze how well \tilde{f}_N approximates f .

Theorem 5.1 *If $f \in \mathcal{F}^r(M)$, then for each $N = 1, 2, \dots$, we have*

$$\|f - \tilde{f}_N\|_{L_\infty(\Omega)} \leq [C_H(d)]^r (2M)^d N^{-r}, \quad (5.1)$$

with $C_H(d)$ as in Theorem 3.2. If, however, **Query 1** stops at a point z where $f(z) \neq 0$, and N satisfies $C_1(r) M N^{-r} < 1/(2d)$, then

$$\|f - \tilde{f}_N\|_{L_\infty(\Omega)} \leq 2C_1(r) d M N^{-r}. \quad (5.2)$$

The remainder of this section is devoted to the proof of this theorem. We will consider the two cases used for the definition of \tilde{f}_N .

5.1 Proof of Theorem 5.1 in Case 1

We fix an arbitrary N . We begin with

Remark 5.2 *For each $k = 1, \dots, N$, there is a $j \in \{1, \dots, d\}$ such that f_j vanishes at r distinct points in $[0, 1]$ of the form $(\hat{x}_k)_j + t_{i,j}$, $i \in \{-r+1, \dots, 0, \dots, r-1\}$ with $|t_{i,j}| \leq N^{-1}$.*

Proof of Remark 5.2 We know that f vanishes at all points in $\Gamma_{n_k}(\hat{x}_k)$ where n_k is the last update associated to \hat{x}_k . We also know that $2^{-n_k} \leq 1/N$. We now prove the remark for $t_{i,j} = \frac{i}{r2^{n_k}}$. Suppose that the statement does not hold, then for this value of k and for each $j = 1, \dots, d$ there is an $i_j \in \{-r+1, \dots, 0, \dots, r-1\}$ such that $z_j := (\hat{x}_k)_j + (r2^{n_k})^{-1}i_j \in [0, 1]$ and $f_j(z_j) \neq 0$. But then $z := (z_1, \dots, z_d) \in \Gamma_{n_k}(\hat{x}_k)$ and $f(z) \neq 0$, which is the desired contradiction. \square

For each k , we let \mathcal{C}_k be the set of all such integers $j \in \{1, \dots, d\}$ that satisfy the Remark. We refer to the integers j in \mathcal{C}_k as the colors of \hat{x}_k .

In the case we are considering, we know that f vanishes at each of the points of **Query 1** and that $\tilde{f}_N = 0$. Let $x = (x_1, \dots, x_d) \in \Omega$. Our goal is to bound the value $f(x)$. We define

$$\delta_j := \delta_j(x) := \inf\{ |(\hat{x}_k)_j - x_j| : k \in \{1, \dots, N\} \text{ such that } j \in \mathcal{C}_k \}, \quad j = 1, \dots, d. \quad (5.3)$$

In other words, $\delta_j(x)$ tells us how well we can approximate x_j by the numbers $(\hat{x}_k)_j$ using those k for which j is in \mathcal{C}_k .

It follows that the rectangle $R := \Omega \cap \prod_{j=1}^d (x_j - \delta_j, x_j + \delta_j)$ does not contain any points \hat{x}_k which have color j and this is true for each $j = 1, \dots, d$. Since, as we have already observed in the Remark, every \hat{x}_k has some colors, it follows that R does not contain any of the points

\hat{x}_k , $k = 1, \dots, N$. From Theorem 3.2, we have that $|R| \leq C_H(d)/N$. Since $|R| \geq \prod_{j=1}^d \delta_j$, we obtain

$$\prod_{j=1}^d \delta_j(x) \leq C_H(d)/N. \quad (5.4)$$

Now fix any $1 \leq j \leq d$. We know from the definition of coloring and the definition of δ_j that there exist r points $t_1, \dots, t_r \in [0, 1]$ contained in an interval J of length $1/N$ such that $\text{dist}(x_j, J) \leq \delta_j$ and f_j vanishes at each of these points. Hence, from Lemma 2.1, we obtain

$$|f_j(x_j)| \leq \|f^{(r)}\|_{L_\infty[0,1]} (|J| + \delta_j)^r \leq M(N^{-1} + \delta_j)^r \leq 2M \max\{N^{-r}, \delta_j^r\}. \quad (5.5)$$

It follows that

$$|f(x)| = \prod_{j=1}^d |f_j(x_j)| \leq 2^d M^d \prod_{j=1}^d \max\{N^{-r}, \delta_j^r\} \leq 2^d M^d [C_H(d)]^r N^{-r}. \quad (5.6)$$

Here in the derivation of the last inequality we used (5.4) and the fact that all the δ_j , $j = 1, \dots, d$ are no greater than one. This completes the proof of the theorem in this case.

5.2 Proof of Theorem 5.1 in Case 2

We now consider the second possibility where $f(z) =: A \neq 0$ for some $z = (z_1, \dots, z_d)$ used in **Query 1**. Let $A_j := \prod_{i \neq j} f_i(z_i)$ for $j = 1, \dots, d$. Sampling f at the points $\tilde{z}_{j,i}$ of (4.2) thus yields the values $f(\tilde{z}_{j,i}) = A_j f_j(i/N)$, $i = 1, 2, \dots, N$. Hence, from (2.1) we obtain

$$\|A_j f_j(x) - F_j(x)\|_{L_\infty[0,1]} \leq C_1(r) A_j M N^{-r}, \quad j = 1, \dots, N.$$

In other words,

$$\|f_j - A_j^{-1} F_j\|_{L_\infty[0,1]} \leq C_1(r) M N^{-r}, \quad j = 1, \dots, N. \quad (5.7)$$

Since $\prod_{j=1}^d A_j = A^{d-1}$, we can write our approximation in the form $\tilde{f}_N(x) = \prod_{j=1}^d A_j^{-1} F_j(x_j)$. Hence, the approximation error can be rewritten as

$$f(x) - \tilde{f}_N(x) = \prod_{j=1}^d f_j(x_j) - \prod_{j=1}^d A_j^{-1} F_j(x_j). \quad (5.8)$$

Now, for any numbers $y_j, y'_j \in [-L, L]$, $j = 1, \dots, d$, we have

$$|y_1 \cdots y_d - y'_1 \cdots y'_d| = \left| \sum_{j=1}^d y_1 \cdots y_{j-1} y'_{j+1} \cdots y'_d (y_j - y'_j) \right| \leq dL^{d-1} \max_{1 \leq j \leq d} |y_j - y'_j|. \quad (5.9)$$

We use this inequality with $y_j := f_j(x_j)$ and $y'_j := A_j^{-1} F_j(x_j)$, in which case we can take $L := 1 + C_1(r) M N^{-r}$ to obtain

$$\|f - \tilde{f}_N\|_{L_\infty(\Omega)} \leq d(1 + C_1(r) M N^{-r})^{d-1} C_1(r) M N^{-r}, \quad (5.10)$$

where we have used (5.7).

For $\varepsilon := C_1(r) M N^{-r}$ we have $\varepsilon < 1/(2d)$ by our assumption, and hence

$$(1 + \varepsilon)^{d-1} = \sum_{\ell=0}^{d-1} \binom{d-1}{\ell} \varepsilon^\ell \leq 1 + (d-1)\varepsilon + \sum_{\ell=2}^{d-1} ((d-1)\varepsilon)^\ell \leq 1 + 2(d-1)\varepsilon.$$

Using this in (5.10), we obtain $\|f - \tilde{f}_N\|_{L^\infty(\Omega)} \leq d\varepsilon + 2d(d-1)\varepsilon^2 \leq 2d\varepsilon$, completing the proof of the theorem. \square

6 Optimality of the Algorithm

It is quite easy to see that our algorithm has asymptotically optimal performance, in terms of N , on the class $\mathcal{F}^r(M)$.

Theorem 6.1 *Given positive integers r and d , there is an absolute constant $c(d, r)$ such that the following holds: Given any algorithm which uses N point queries to approximate f by $A_N(f)$, there is a function $f \in \mathcal{F}^r(M)$ such that*

$$\|f - A_N(f)\|_{L^\infty(\Omega)} \geq c(r, d)M^d N^{-r}. \quad (6.1)$$

Proof: We can assume without loss of generality that $N = m^d - 1$ for some positive integer m . We divide Ω into $N + 1$ cubes of sidelength $1/m$. To the proposed query algorithm we return the value zero to each of the N query points. Now we can choose a cube Q of sidelength $1/m$ which contains none of the N query points. There is a function $g \in \mathcal{F}^r(M)$ which is supported in Q and has maximum value $[c(r)Mm^{-r}]^d$. Since the proposed algorithm gives $A_N(g) = A_N(0)$, for one of the two functions $f = 0$ or $f = g$, (6.1) follows. \square

Let us finally note that our estimate for the computational work in our algorithm is dominated by **Query 1**. Under additional assumptions on f , **Query 1** can have much lower complexity. For example, if each component function f_j is a polynomial of a fixed degree p , or more generally if each component has at most a fixed number p of zeros, then **Query 1** will terminate after at most p steps. Indeed, the Halton sequence never repeats a coordinate value.

References

- [1] H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numerica **13**(2004), 147–269.
- [2] A. Dumitrescu and M. Jiang, *On the largest empty axis-parallel box amidst n points*, Algorithmica, doi:10.1007/s00453-012-9635-5 (2012).
- [3] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, Springer Series in Computational Mathematics 42, Springer-Verlag Berlin Heidelberg (2012).
- [4] J.H. Halton, *On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals*, Numer. Math., **2**(1960), 84–90.

- [5] J.M. Hammersley, *Monte Carlo methods for solving multivariable problems*, Ann. New York Acad. Sci., **86**(1960), 844–874.
- [6] G. Rote and R.F. Tichy, *Quasi-Monte-Carlo methods and the dispersion of point sequences*, Mathl. Comput. Modelling, **23**(1996), 9–23.
- [7] V. Temlyakov, *Approximation of Periodic Functions*, Nova Science Publishers, Inc. (1993).

Markus Bachmayr,

Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, D-52056 Aachen Germany

bachmayr@igpm.rwth-aachen.de

Wolfgang Dahmen

Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, D-52056 Aachen Germany

dahmen@igpm.rwth-aachen.de

Ronald DeVore

Department of Mathematics, Texas A&M University, College Station, TX

rdevore@math.tamu.edu

Lars Grasedyck,

Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, D-52056 Aachen Germany

lgr@igpm.rwth-aachen.de