# The Randomized Complexity of Indefinite Integration

Stefan Heinrich, Bernhard Milla
Department of Computer Science
University of Kaiserslautern
D-67653 Kaiserslautern, Germany
e-mail: heinrich[milla]@informatik.uni-kl.de

**Abstract**

We show that for functions $f \in L_p([0,1]^d)$, where $1 \le p \le \infty$, the family of integrals

$$\int_{[0,x]} f(t)dt \qquad (x = (x_1, \dots, x_d) \in [0,1]^d)$$

can be approximated by a randomized algorithm uniformly over $x \in [0,1]^d$ with the same rate $n^{-1+1/\min(p,2)}$ as the optimal rate for a single integral, where $n$ is the number of samples. We present two algorithms, one being of optimal order, the other up to logarithmic factors. We also prove lower bounds and discuss the dependence of the constants in the error estimates on the dimension.

## 1 Introduction

It is well-known that optimal randomized algorithms for integration of $L_p([0,1]^d)$ functions with $n$ samples have error rate $n^{-1+1/\min(p,2)}$ [14, 4]. In this paper we show that the same rate can be obtained for the simultaneous computation of all integrals

$$\int_{[0,x]} f(t)dt$$

uniformly over $x \in [0,1]^d$. Thus, we want to approximate the indefinite integral, the anti-derivative. While numerous papers study the complexity of definite integrals, the case of indefinite integration has not been considered so far.

We propose and analyze two algorithms and prove lower bounds. The first algorithm is the simple sampling algorithm – a function valued version of the standard Monte Carlo method. The second one is a combination of the Smolyak algorithm with simultaneous Monte Carlo sampling. Both algorithms need $\mathcal{O}(n)$

function values and produce an approximation which is a linear combination of $\mathcal{O}(n)$ functions.

The first one is of optimal order for all $1 \leq p \leq \infty$ and, moreover, the constants in the error estimates depend polynomially on the dimension. Thus, it proves polynomial tractability of the problem in the randomized setting. This is noteworthy since so far only very few unweighted problems (i.e., all dimensions play the same role) are known to share polynomial tractability (see, e.g., the comment at the top of page 39 of [17]).

The second algorithm is of optimal order for $2 < p \leq \infty$, while for $1 \leq p \leq 2$ additional logarithmic factors occur. The second algorithm, however, has the advantage that once the approximation is established, any value of it can be computed in only $\mathcal{O}(1)$ operations for $2 < p \leq \infty$ and in $\mathcal{O}((\log n)^{d-1})$ for $1 \leq p \leq 2$, while in the case of the first algorithm this takes $\Theta(n)$. The simple sampling algorithm, on the other hand, can be made more efficient for $d$ fixed (and small), see Section 6.2. Still, for $2 < p \leq \infty$ the Smolyak-Monte Carlo algorithm has better efficiency estimates, see the discussion at the end of Section 6.2.

We also present a sharp in $n$ and dimension independent lower bound. Furthermore, for $p > 1$, we prove lower bounds which show that for fixed $\varepsilon > 0$ the dependence of the minimal number of samples of an algorithm with error $\leq \varepsilon$ on the dimension is linear.

Let us note that the rate of deterministic algorithms is $\Theta(1)$ for all $p$ with $1 \leq p \leq \infty$, thus there is no convergence to zero at all, see Section 6.1. For comparison, the optimal rate for randomized algorithms, is $n^{-1+1/\min(p,2)}$, so it is $n^{-1/2}$ for $2 \leq p \leq \infty$, but in the interval $1 < p < 2$ the exponent goes to zero as $p$ approaches 1. Finally, for $p = 1$ the rate of convergence of randomized algorithms is $\Theta(1)$, as well.

The paper is organized in the following way: Section 2 contains notation and preliminaries, the simple sampling algorithm is described and analyzed in Section 3, the Smolyak-Monte Carlo algorithm in Section 4. Lower bounds are presented in Section 5, and in Section 6 we comment on the deterministic setting, present an efficient way of computing point evaluations for the simple sampling algorithm, and discuss measurability issues.

# 2 Notation and Preliminaries

We write $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. The logarithm log is always meant as $\log_2$. All functions and Banach spaces considered in this paper are assumed to be defined over the same field of scalars $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. For a Banach space X we denote the unit ball by $\mathcal{B}_X$ and the dual space by $X^*$. Given Banach spaces $X, Y$, the space of all bounded linear operators from $X$ to $Y$ is denoted by $\mathscr{L}(X, Y)$, and, if $X = Y$, by $\mathscr{L}(X)$.

Let $d \in \mathbb{N}$, $Q = [0,1]^d$, let $C(Q)$ denote the space of continuous functions on $Q$ and, for $1 \le p \le \infty$, let $L_p(Q)$ be the space of (equivalence classes of) $p$-th power integrable with respect to the Lebesgue measure functions, both equipped with their usual norm. Let $\mathcal{F}(Q)$ denote the linear space of all functions on $Q$ and let $B_0(Q)$ be the space of all bounded Lebesgue measurable functions (not equivalence classes) on $Q$ with supremum norm.

Let $1 \le p \le \infty$. We study $S^{(d)} \in \mathscr{L}(L_p(Q), C(Q))$ given by

$$(S^{(d)} f)(x) = \int_{[0,x]} f(t)dt, \qquad (x = (x_1, \dots, x_d) \in Q),$$

where $[0,x] = [0,x_1] \times ... \times [0,x_d]$. Note that

$$\|S^{(d)} : L_p(Q) \to C(Q)\| = 1 \tag{1}$$

(the problem is normalized).

Throughout the paper the symbols $c, c_0, c_1, \dots$ denote positive constants which are either absolute or may depend only on $p$ and $p_1$. Constants which may also depend on $d$ are denoted by $c(d), c_0(d)$, etc. The same symbol may denote different constants (also when they appear in a sequence of relations).

# 3 The simple sampling algorithm

We have

$$(S^{(d)} f)(x) = \int_{[0,1]^d} \chi_{[0,x]}(t) f(t)dt.$$

We introduce the simple sampling algorithm as follows: Let $n \in \mathbb{N}$ and let $(\xi_i)_{i=1}^n$ be independent, uniformly distributed on $Q = [0,1]^d$ random variables on some probability space $(\Omega, \Sigma, \mathbb{P})$. We assume without loss of generality that $(\Omega, \Sigma, \mathbb{P})$ is complete, meaning that $D \subseteq D_1 \in \Sigma$ and $\mathbb{P}(D_1) = 0$ imply $D \in \Sigma$ (if $(\Omega, \Sigma, \mathbb{P})$ is not complete, we replace it by its completion). Then we approximate for $f \in L_p(Q)$

$$S^{(d)} f \approx A_n^1 f,$$

with $A_n^1 f$ given by

$$(A_n^1 f)(x) = \frac{1}{n} \sum_{i=1}^n \chi_{[0,x]}(\xi_i) f(\xi_i) \quad (x \in Q). \tag{2}$$

We have

$$A_n^1 f = \frac{1}{n} \sum_{i=1}^n f(\xi_i)\, \chi_{[\xi_i, \bar{1}]}, \tag{3}$$

where

$$\bar{1} = (\underbrace{1, 1, \dots, 1}_{d}).$$

Since

$$(S^{(d)}f)(x) = \int_{[0,1]^d} f(t)\chi_{[t,\bar{1}]}(x)dt \quad (x \in [0,1]^d),$$

the algorithm can be considered as a function-valued version of the standard Monte Carlo method for integration. Let us mention that the simple sampling algorithm produces discontinuous in $x$ functions, so we consider it as mapping into $B_0(Q)$ and as an approximation to $S^{(d)} : L_p(Q) \to B_0(Q)$, where we identify $C(Q)$ in the canonical way with a subspace of $B_0(Q)$. Furthermore, note that $A_n^1$ is lacking certain measurability properties, see the beginning of Section 5 and Section 6.3 for details. Nevertheless the mapping

$$\omega \to \|S^{(d)}f - A_{n,\omega}^1 f\|_{B_0(Q)}$$

is $\Sigma$-measurable, where we write

$$A_{n,\omega}^1 f = \frac{1}{n}\sum_{i=1}^{n} f(\xi_i(\omega))\,\chi_{[\xi_i(\omega),\bar{1}]} \quad (\omega \in \Omega) \tag{4}$$

to emphasize the dependence on $\omega \in \Omega$. Indeed, this follows from

$$\|S^{(d)}f - A_{n,\omega}^1 f\|_{B_0(Q)} = \sup_{x\in[0,1]^d\cap\mathbb{Q}^d} \left|(S^{(d)}f)(x) - (A_{n,\omega}^1 f)(x)\right| \quad (\omega \in \Omega)$$

(with $\mathbb{Q}$ the rationals), which, in turn, is a simple consequence of (4). Thus it makes sense to consider the $p_1$-st moment $\mathbb{E}\,\|S^{(d)}f - A_n^1 f\|_{B_0(Q)}^{p_1}$ for suitable $1 \le p_1 < \infty$, as we will do below.

We also introduce a slight modification of this algorithm, which has values in $C([0,1]^d)$ and possesses the desired measurability properties. For this purpose we introduce for $l \in \mathbb{N}$ the function $\varphi_l^{(1)} \in C([0,1]^2)$ by

$$\varphi_l^{(1)}(x,t) = \begin{cases} 1 & \text{if} \quad t \le x \\ 1 - l(t-x) & \text{if} \quad x < t < x + \frac{1}{l} \\ 0 & \text{if} \quad x + \frac{1}{l} \le t. \end{cases} \tag{5}$$

Define $\varphi_l^{(d)} \in C([0,1]^{2d})$ by setting for $x = (x_1, \dots, x_d)$ and $t = (t_1, \dots, t_d)$

$$\varphi_l^{(d)}(x,t) = \prod_{j=1}^{d} \varphi_l^{(1)}(x_j, t_j). \tag{6}$$

Now we put

$$A_{n,l}^2 f = \frac{1}{n}\sum_{i=1}^{n} \varphi_l^{(d)}(\,\cdot\,, \xi_i) f(\xi_i). \tag{7}$$

Let us first consider the cost of computing $A_n^1 f$ and $A_{n,l}^2 f$. Each of them needs $dn$ independent uniformly distributed on $[0,1]$ random variables and $n$ function values to determine the respective representation (3) and (7). Next we have a look at computing $(A_n^1 f)(x)$ and $\left(A_{n,l}^2 f\right)(x)$ for any given $x \in Q$. Since the supports of the involved functions can overlap in an arbitrary way, we need $cdn$ operations to compute term after term in (3), and similarly in (7). More efficient approaches for fixed (small) $d$ are discussed in Section 6.2.

Now we pass to the error analysis. For $m \in \mathbb{N}$ let $\Gamma_m$ be the equidistant grid on $Q = [0,1]^d$ with mesh-size $1/m$. We need the following (bracketing) lemma.

**Lemma 3.1.** *Let $1 < p \leq \infty$, $m \in \mathbb{N}$, and $f \in L_p(Q)$ with $f \geq 0$. Let $\varepsilon_0 > 0$ and assume $\psi : [0,1]^{2d} \to \mathbb{R}$ is a measurable function satisfying the following: For each $x \in [0,1]^d$ there exist $y, z \in \Gamma_m$ with*

$$y \leq z, \tag{8}$$

$$|[0,z]| - |[0,y]| \leq \varepsilon_0, \tag{9}$$

$$\chi_{[0,y]}(t) \leq \psi(x,t) \leq \chi_{[0,z]}(t) \quad (t \in [0,1]^d). \tag{10}$$

*Then the following holds $\mathbb{P}$-almost surely:*

$$\begin{aligned}
&\sup_{x\in Q}\left|\int_{[0,x]} f(t)dt - \frac{1}{n}\sum_{i=1}^n \psi(x,\xi_i)f(\xi_i)\right| \\
&\leq \varepsilon_0^{1/p^*}\|f\|_{L_p(Q)} + \sup_{x\in\Gamma_m}\left|\int_{[0,x]} f(t)dt - \frac{1}{n}\sum_{i=1}^n \chi_{[0,x]}(\xi_i)f(\xi_i)\right|,
\end{aligned} \tag{11}$$

*where $p^*$ is given by $1/p + 1/p^* = 1$.*

*Proof.* We assume that the values $f(\xi_i)$, $1 \leq i \leq n$, are defined, which is the case $\mathbb{P}$-almost surely. Let $x \in Q$ and choose $y, z \in \Gamma_m$ satisfying (8–10). Then

$$\begin{aligned}
&\int_{[0,x]} f(t)dt - \frac{1}{n}\sum_{i=1}^n \psi(x,\xi_i)f(\xi_i) \\
&\leq \int_{[0,z]\setminus[0,y]} f(t)dt + \int_{[0,y]} f(t)dt - \frac{1}{n}\sum_{i=1}^n \chi_{[0,y]}(\xi_i)f(\xi_i).
\end{aligned}$$

Similarly

$$\begin{aligned}
&-\int_{[0,x]} f(t)dt + \frac{1}{n}\sum_{i=1}^n \psi(x,\xi_i)f(\xi_i) \\
&\leq \int_{[0,z]\setminus[0,y]} f(t)dt - \int_{[0,z]} f(t)dt + \frac{1}{n}\sum_{i=1}^n \chi_{[0,z]}(\xi_i)f(\xi_i).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\left|\int_{[0,x]} f(t)dt - \frac{1}{n}\sum_{i=1}^n \psi(x,\xi_i)f(\xi_i)\right| \\
&\leq \int_{[0,z]\setminus[0,y]} f(t)dt + \max_{u\in\{y,z\}}\left|\int_{[0,u]} f(t)dt - \frac{1}{n}\sum_{i=1}^n \chi_{[0,u]}(\xi_i)f(\xi_i)\right|.
\end{aligned} \tag{12}$$

Moreover,

$$\int_{[0,z]\setminus[0,y]} f(t)dt \le (|[0,z]| - |[0,y]|)^{1/p^*} \|f\|_{L_p(Q)} \le \varepsilon_0^{1/p^*} \|f\|_{L_p(Q)}. \tag{13}$$

Combining (12) and (13) yields (11). □

Next we recall some facts on Banach space valued random variables which will be needed in the subsequent analysis. Let $X$ and $Y$ be Banach spaces and let $T \in \mathscr{L}(X,Y)$. Given $p$ with $1 \le p \le 2$, the type $p$ constant $\tau_p(T)$ of the operator $T$ is the smallest $c$ with $0 < c \le +\infty$, such that for all $n$ and all sequences $(x_i)_{i=1}^n \subset X$,

$$\mathbb{E}\Big\|\sum_{i=1}^n \varepsilon_i T x_i\Big\|^p \le c^p \sum_{i=1}^n \|x_i\|^p, \tag{14}$$

where $(\varepsilon_i)$ denotes a sequence of independent symmetric Bernoulli random variables on some probability space $(\Omega, \Sigma, \mathbb{P})$, i.e., $\mathbb{P}\{\varepsilon_i = 1\} = \mathbb{P}\{\varepsilon_i = -1\} = \frac{1}{2}$. The operator $T$ is said to be of type $p$ if $\tau_p(T) < \infty$. Each operator is of type 1. A Banach space $X$ is of type $p$ iff the identity operator of $X$ is of type $p$. We refer to [11], ch. 9 for definitions and basic facts on the type of Banach spaces. The operator analogues are straightforward.

We will use the following result. The Banach space case of it with $p_1 = p$ is contained in Proposition 9.11 of [11]. The proof given there easily extends to the case of general $p_1$, as shown in Lemma 2.1 of [8]. Here we note that the operator version of this lemma has literally the same proof, so we omit it.

**Lemma 3.2.** *Let $1 \le p \le 2$, $p \le p_1 < \infty$. Then there is a constant $c > 0$ such that for all Banach spaces $X, Y$, each operator $T \in \mathscr{L}(X,Y)$ of type $p$, each $n \in \mathbb{N}$ and each sequence of independent, mean zero $X$-valued random variables $(\eta_i)_{i=1}^n$ with $\mathbb{E}\,\|\eta_i\|^{p_1} < \infty$ $(1 \le i \le n)$ the following holds:*

$$\left(\mathbb{E}\Big\|\sum_{i=1}^n T\eta_i\Big\|^{p_1}\right)^{1/p_1} \le c\tau_p(T)\left(\sum_{i=1}^n \left(\mathbb{E}\,\|\eta_i\|^{p_1}\right)^{p/p_1}\right)^{1/p}.$$

The next lemma provides the key estimate for the simple sampling algorithm.

**Lemma 3.3.** *Let $1 \le p \le \infty$, $1 \le p_1 < \infty$, $p_1 \le p$, and $\bar{p} = \min(p, 2)$. Then there is a constant $c > 0$ such that for all $d, m, n \in \mathbb{N}$, $f \in L_p(Q)$*

$$\left(\mathbb{E} \sup_{x \in \Gamma_m} \left| \int_{[0,x]} f(t)dt - \frac{1}{n}\sum_{i=1}^n \chi_{[0,x]}(\xi_i) f(\xi_i)\right|^{p_1}\right)^{1/p_1} \le c d^{1-1/\bar{p}} n^{-1+1/\bar{p}} \|f\|_{L_p(Q)}.$$

*Proof.* We can assume that $p_1 \geq \bar{p}$, for smaller $p_1$ the result follows from Hölder's inequality. Let $\Sigma_m$ be the $\sigma$-algebra generated by the collection of sets $\{[0,x] : x \in \Gamma_m\}$ and let $M(Q, \Sigma_m)$ be the Banach space of signed measures on $\Sigma_m$, with the total variation norm. Consider the operator

$$J_m : M(Q, \Sigma_m) \to \ell_\infty(\Gamma_m), \qquad J_m\mu = (\mu([0,x]))_{x \in \Gamma_m}.$$

By a result of Pisier, see Theorem 1 and Remark 6 of [19], there is a constant $c > 0$ depending only on $\bar{p}$ such that the type $\bar{p}$ constant of $J_m$, see (14), satisfies

$$\tau_{\bar{p}}(J_m) \leq c d^{1-1/\bar{p}} \tag{15}$$

(this uses the fact that the Vapnik-Červonenkis dimension of the family of sets $\{[0,x] : x \in [0,1]^d\}$ is $d$, see, e.g., [3], Cor. 9.2.15). Now let $f \in L_p(Q)$. We define $M(Q, \Sigma_m)$-valued random variables $(\eta_i)_{i=1}^n$ by setting

$$\eta_i(B) = \int_B f(t)dt - \chi_B(\xi_i) f(\xi_i) \qquad (B \in \Sigma_m).$$

The $\eta_i$ are independent and of zero mean. Moreover,

$$\left(\mathbb{E}\, \|\eta_i\|_{M(Q,\Sigma_m)}^{p_1}\right)^{1/p_1} \leq \left(\mathbb{E}\left(\int_Q |f(t)|dt + |f(\xi_i)|\right)^{p_1}\right)^{1/p_1} \leq 2\|f\|_{L_{p_1}(Q)}.$$

By Lemma 3.2 and relation (15) we get

$$\begin{aligned}
&\left(\mathbb{E} \sup_{x \in \Gamma_m} \left| \int_{[0,x]} f(t)dt - \frac{1}{n}\sum_{i=1}^n \chi_{[0,x]}(\xi_i) f(\xi_i) \right|^{p_1}\right)^{1/p_1} \\
&= n^{-1} \left(\mathbb{E} \left\| \sum_{i=1}^n J_m \eta_i \right\|_{\ell_\infty(\Gamma_m)}^{p_1}\right)^{1/p_1} \\
&\leq c d^{1-1/\bar{p}} n^{-1} \left(\sum_{i=1}^n \left(\mathbb{E}\, \|\eta_i\|_{M(Q,\Sigma_m)}^{p_1}\right)^{\bar{p}/p_1}\right)^{1/\bar{p}} \\
&\leq c d^{1-1/\bar{p}} n^{-1+1/\bar{p}} \|f\|_{L_{p_1}(Q)}.
\end{aligned}$$

□

**Theorem 3.4.** *Let $1 \leq p \leq \infty$, $1 \leq p_1 < \infty$, $p_1 \leq p$, and $\bar{p} = \min(p, 2)$. Then there is a constant $c > 0$ such that for all $d, n, l \in \mathbb{N}$, $l \geq 2dn$, $f \in L_p(Q)$,*

$$\left.\begin{array}{l}
\left(\mathbb{E}\, \|S^{(d)} f - A_n^1 f\|_{B_0(Q)}^{p_1}\right)^{1/p_1} \\
\left(\mathbb{E}\, \|S^{(d)} f - A_{n,l}^2 f\|_{C(Q)}^{p_1}\right)^{1/p_1}
\end{array}\right\} \leq c d^{1-1/\bar{p}} n^{-1+1/\bar{p}} \|f\|_{L_p(Q)}.$$

*Proof.* For $p = 1$ the result follows trivially from the definitions (2) and (7) of $A_n^1$ and $A_{n,l}^2$. So let $p > 1$. We can assume $f \geq 0$, otherwise we consider positive and negative part of $f$ separately. Put $m = 2dn$ and observe first that the choice

$$\psi(x,t) = \chi_{[0,x]}(t)$$

needed for $A^1$ satisfies the assumptions (8–10) of Lemma 3.1 with $\varepsilon_0 = d/m$. Indeed, given $x \in [0,1]^d$ we can choose $y = (y_1, \dots, y_d) \in \Gamma_m$ and $z = (z_1, \dots, z_d) \in \Gamma_m$ so that (10) holds and

$$y_j + \frac{1}{m} = z_j \quad (j = 1, \dots, d).$$

We have

$$\begin{aligned} &||[0,z]| - |[0,y]|| \\ &\leq \sum_{j=1}^{d} |y_1 \dots y_{j-1} z_j z_{j+1} \dots z_d - y_1 \dots y_{j-1} y_j z_{j+1} \dots z_d| \leq \frac{d}{m}. \end{aligned} \tag{16}$$

Similarly, for

$$\psi(x,t) = \varphi_l^{(d)}(x,t),$$

see (5–6), with $l \geq m$, we can choose appropriate $y$ and $z$ with

$$y_j + \frac{2}{m} = z_j \quad (j = 1, \dots, d),$$

implying $||[0,z]| - |[0,y]|| \leq 2d/m$. We obtain from Lemmas 3.1 and 3.3

$$\left.\begin{array}{l} \left(\mathbb{E}\,\|S^{(d)} f - A_n^1 f\|_{B_0(Q)}^{p_1}\right)^{1/p_1} \\ \left(\mathbb{E}\,\|S^{(d)} f - A_{n,l}^2 f\|_{C(Q)}^{p_1}\right)^{1/p_1} \end{array}\right\}$$

$$\begin{aligned} &\leq n^{-1+1/p} \|f\|_{L_p(Q)} + c d^{1-1/\bar{p}} n^{-1+1/\bar{p}} \|f\|_{L_p(Q)} \\ &\leq c d^{1-1/\bar{p}} n^{-1+1/\bar{p}} \|f\|_{L_p(Q)}. \end{aligned}$$

□

It follows that for $1 < p \leq \infty$ the family of problems

$$\left(S^{(d)} : \mathcal{B}_{L_p([0,1]^d)} \to C([0,1]^d)\right)_{d\in\mathbb{N}}$$

is polynomially tractable in the randomized setting, for the absolute and the normalized error criterion (which in this case is the same, because of (1)), see [17] for the definitions. We note that most of the polynomially tractable problems considered in [17, 18] are weighted problems (i.e., with decreasing dependence on subsequent dimensions). This way we obtained a new family of unweighted polynomially tractable problems. Furthermore, most problems analyzed in [17, 18] are defined between Hilbert spaces, while here we study a Banach space situation.

# 4 The Smolyak-Monte Carlo algorithm

First we introduce the Smolyak algorithm in a form needed for our later purposes. The Smolyak algorithm is by now a standard technique of treating high-dimensional problems, in particular those of tensor product form. The basic idea of the algorithm is the balancing of fine approximation in certain dimensions with rough approximation in others. For further background we refer to [17, 18] and the references therein. For each $m \in \mathbb{N}$ with $m \geq 2$ let

$$(P_{m,l})_{l=0}^{\infty} \subset \mathscr{L}(C([0,1]))$$

be a sequence of operators of the form

$$P_{m,l} f = \sum_{i=1}^{n_{m,l}} f(x_{m,l,i}) \psi_{m,l,i} \tag{17}$$

with $x_{m,l,i} \in [0,1]$ and $\psi_{m,l,i} \in C([0,1])$, $\psi_{m,l,i} \neq 0$ $(i = 1, \ldots, n_{m,l},\ l \in \mathbb{N}_0)$. We assume w.l.o.g. that the points $\{x_{m,l,i} : i = 1, \ldots, n_{m,l}\}$ are pairwise different and ordered increasingly,

$$x_{m,l,1} < x_{m,l,2} < \ldots < x_{m,l,n_{m,l}}.$$

Furthermore, we define $x_{m,l,0} = 0$ and $x_{m,l,n_{m,l}+1} = 1$.

We assume the following: There are constants $c_{1-4} > 0$ such that for all $m \in \mathbb{N}$ with $m \geq 2$ and for all $l \in \mathbb{N}_0$

$$n_{m,l} \leq c_1 m^l \tag{18}$$

$$\max_{1 \leq i \leq n_{m,l}+1} (x_{m,l,i} - x_{m,l,i-1}) \leq c_2 m^{-l} \tag{19}$$

$$\|P_{m,l}\|_{\mathscr{L}(C([0,1]))} \leq c_3 \tag{20}$$

$$\sup_{f \in \mathcal{B}_{W_p^1([0,1])}} \|f - P_{m,l} f\|_{C([0,1])} \leq c_4 m^{-(1-1/p)l}. \tag{21}$$

Here $W_p^1([0,1])$ stands for the space of all functions in $L_p([0,1])$ whose first derivative, in the distributional sense, also belongs to $L_p([0,1])$, endowed with the norm

$$\|f\|_{W_p^1([0,1])} = \left( \|f\|_{L_p([0,1])}^p + \|f'\|_{L_p([0,1])}^p \right)^{1/p}$$

(and the usual modification for $p = \infty$).

Operators with these properties are easily constructed. For example, given $m$, we let $P_{m,l}$ be piecewise linear interpolation, applied to the subdivision of $[0,1]$ into $m^l$ equal length subintervals. For this choice it is well-known that (18–21) hold.

We fix any $m \in \mathbb{N}$, $m \geq 2$. In the sequel $m$ will be an algorithm parameter, and for convenience of notation we drop the subscript $m$ and write $P_l$, $n_l$, $x_{l,i}$, $\psi_{l,i}$.

For the definition of the Smolyak algorithm in the case $d > 1$ and for the subsequent analysis of the algorithm we use tensor products. Such an approach is usually applied in the case that both the source and the target space are Hilbert spaces. Here we study a Banach space situation, the source space being $L_p(Q)$ ($1 \leq p \leq \infty$), the target space $C(Q)$. For this purpose we use Banach space tensor norms, as recently done in [20].

The tensor product structure of $S^{(d)}$ in the Banach space case is more subtle than in the Hilbert case. In particular, we have to consider appropriate tensor norms to relate the spaces $C([0,1]^d)$ and $L_p([0,1]^d)$ on the $d$-dimensional cube to tensor products of the corresponding spaces on the unit interval. Moreover, these tensor products should have the property that the norm of the tensor product of operators is equal to the product of the norms of the operators. We present the needed notation and facts below. Further details and proofs can be found in [2] and [12].

Let $X \otimes Y$ be the algebraic tensor product of Banach spaces $X$ and $Y$. For $z = \sum_{i=1}^n x_i \otimes y_i \in X \otimes Y$ define

$$\lambda(z) = \sup_{u \in \mathcal{B}_{X^*},\, v \in \mathcal{B}_{Y^*}} \Big| \sum_{i=1}^n \langle x_i, u \rangle \langle y_i, v \rangle \Big|$$

and for $1 \leq p < \infty$, with $p^*$ satisfying $1/p + 1/p^* = 1$,

$$\alpha_p(z) = \inf \left\{ \Big( \sum_{i=1}^n \|x_i\|^p \Big)^{1/p} \sup_{v \in \mathcal{B}_{Y^*}} \Big( \sum_{i=1}^n | \langle y_i, v \rangle |^{p^*} \Big)^{1/p^*} \right\}$$

(with the usual modification for $p^* = \infty$), where the infimum is taken over all representations $z = \sum_{i=1}^n x_i \otimes y_i$. We have for $1 \leq p_1 \leq p < \infty$ and $z \in X \otimes Y$

$$\lambda(z) \leq \alpha_p(z) \leq \alpha_{p_1}(z). \tag{22}$$

For $\theta \in \{\lambda,\, \alpha_p\ (1 \leq p < \infty)\}$, the tensor product $X \otimes_\theta Y$ is defined as the completion of $X \otimes Y$ with respect to the norm $\theta$.

We use for $d > 1$ the canonical isometric identifications

$$C([0,1]) \otimes_\lambda C([0,1]^{d-1}) = C([0,1]^d), \tag{23}$$

for $1 \leq p < \infty$

$$L_p([0,1]) \otimes_{\alpha_p} L_p([0,1]^{d-1}) = L_p([0,1]^d), \tag{24}$$

and the canonical isometric embedding

$$L_\infty([0,1]) \otimes_\lambda L_\infty([0,1]^{d-1}) \subset L_\infty([0,1]^d) \tag{25}$$

(which is a proper embedding).

Given Banach spaces $X_1, X_2, Y_1, Y_2$, operators $T_1 \in \mathscr{L}(X_1, Y_1)$, $T_2 \in \mathscr{L}(X_2, Y_2)$, and two tensor norms

$$\theta_1, \theta_2 \in \{\lambda,\, \alpha_p\ (1 \le p < \infty)\}, \quad \theta_1 \ge \theta_2,$$

the algebraic tensor product

$$T_1 \otimes T_2 : X_1 \otimes X_2 \to Y_1 \otimes Y_2$$

extends to a bounded linear operator (we use the same symbol for the extension)

$$T_1 \otimes T_2 \in \mathscr{L}(X_1 \otimes_{\theta_1} X_2, Y_1 \otimes_{\theta_2} Y_2) \tag{26}$$

with

$$\|T_1 \otimes T_2 : X_1 \otimes_{\theta_1} X_2 \to Y_1 \otimes_{\theta_2} Y_2\| = \|T_1 : X_1 \to Y_1\|\, \|T_2 : X_2 \to Y_2\|. \tag{27}$$

Let $I^{(d)}$ denote the identity operator on $C([0,1]^d)$. In the sense of (23) and (26) we have

$$I^{(d)} = I^{(1)} \otimes I^{(d-1)},$$

furthermore, taking into account (24), we have for $1 \le p < \infty$,

$$S^{(d)} = S^{(1)} \otimes S^{(d-1)},$$

and finally, based on (25), for $p = \infty$,

$$S^{(d)}\big|_{L_\infty([0,1]) \otimes_\lambda L_\infty([0,1]^{d-1})} = S^{(1)} \otimes S^{(d-1)}.$$

Now we are ready to define operators $P_L^{(d)} \in \mathscr{L}(C([0,1]^d))$ for $L \in \mathbb{N}_0$ by induction over $d$. For $d = 1$ we put

$$P_L^{(1)} = P_L.$$

For $d > 1$ we use the identification (23) and set

$$P_L^{(d)} = \sum_{l=0}^{L} (P_l - P_{l-1}) \otimes P_{L-l}^{(d-1)}$$

with the convention that $P_{-1} := 0$. For the sequel we also fix $L$, which will be another algorithm parameter. The first step in the construction of our algorithm is the approximation of $S^{(d)} f$ by $P_L^{(d)} S^{(d)} f$.

Next we are going to approximate $P_L^{(d)} S^{(d)} f$. For this purpose let us take a closer look at the structure of the operator $P_L^{(d)}$. Let for $l \in \mathbb{N}_0$

$$\Gamma_l = \{x_{l,i} \,:\, 1 \le i \le n_l\}, \quad \hat{\Gamma}_l = \Gamma_{l-1} \cup \Gamma_l, \tag{28}$$

where we set $\Gamma_{-1} = \emptyset$. Let the points of $\hat{\Gamma}_l$ ($l \in \mathbb{N}_0$) be denoted in increasing order by

$$\hat{x}_{l,1} < \hat{x}_{l,2} < \ldots < \hat{x}_{l,\hat{n}_l}, \tag{29}$$

where $\hat{n}_l = |\hat{\Gamma}_l|$. Now the operator $P_l - P_{l-1}$ can be written as

$$\begin{aligned} (P_l - P_{l-1})f &= \sum_{j=1}^{n_l} f(x_{l,j})\psi_{l,j} - \sum_{j=1}^{n_{l-1}} f(x_{l-1,j})\psi_{l-1,j} \\ &= \sum_{i=1}^{\hat{n}_l} f(\hat{x}_{l,i})\hat{\psi}_{l,i} \end{aligned} \tag{30}$$

with $n_{-1} = 0$ and

$$\hat{\psi}_{l,i} = \begin{cases} \psi_{l,j} & \text{if} \quad \hat{x}_{l,i} = x_{l,j} \in \Gamma_l \setminus \Gamma_{l-1} \\ -\psi_{l-1,j} & \text{if} \quad \hat{x}_{l,i} = x_{l-1,j} \in \Gamma_{l-1} \setminus \Gamma_l \\ \psi_{l,j_1} - \psi_{l-1,j_2} & \text{if} \quad \hat{x}_{l,i} = x_{l,j_1} = x_{l-1,j_2} \in \Gamma_{l-1} \cap \Gamma_l. \end{cases}$$

We can split the operator $P_L^{(d)}$ as follows:

$$P_L^{(d)} = \sum_{\bar{l} \in \mathbb{N}_0^d,\, |\bar{l}| = L} U_{\bar{l}} \tag{31}$$

where for $\bar{l} = (l_1, \ldots, l_d)$ we set $|\bar{l}| = l_1 + \cdots + l_d$ and

$$U_{\bar{l}} = (P_{l_1} - P_{l_1-1}) \otimes \cdots \otimes (P_{l_{d-1}} - P_{l_{d-1}-1}) \otimes P_{l_d}. \tag{32}$$

Define for $\bar{l} = (l_1, \ldots, l_d) \in \mathbb{N}_0^d$

$$\begin{aligned} \bar{n}_{\bar{l}} &= (\hat{n}_{l_1}, \ldots, \hat{n}_{l_{d-1}}, n_{l_d}) \\ \mathcal{I}_{\bar{l}} &= \{\bar{i} \in \mathbb{N}^d \,:\, \bar{1} \leq \bar{i} \leq \bar{n}_{\bar{l}}\} \end{aligned}$$

with componentwise inequalities in the last line. Furthermore, for $0 \leq \bar{i} \leq \bar{n}_{\bar{l}}$ we set

$$x_{\bar{l},\bar{i}} = (\hat{x}_{l_1,i_1}, \ldots, \hat{x}_{l_{d-1},i_{d-1}}, x_{l_d,i_d}) \in [0,1]^d,$$

where we define $\hat{x}_{l,0} = 0$. Moreover, for $\bar{i} \in \mathcal{I}_{\bar{l}}$ we put

$$\begin{aligned} \psi_{\bar{l},\bar{i}} &= \hat{\psi}_{l_1,i_1} \otimes \cdots \otimes \hat{\psi}_{l_{d-1},i_{d-1}} \otimes \psi_{l_d,i_d} \in C([0,1]^d) \\ Q_{\bar{l},\bar{i}} &= [x_{\bar{l},\bar{i}-\bar{1}}, x_{\bar{l},\bar{i}}] \\ &= [\hat{x}_{l_1,i_1-1}, \hat{x}_{l_1,i_1}] \times \cdots \times [\hat{x}_{l_{d-1},i_{d-1}-1}, \hat{x}_{l_{d-1},i_{d-1}}] \times [x_{l_d,i_d-1}, x_{l_d,i_d}]. \end{aligned}$$

Combining (23), (17), (30), and (32), we obtain

$$U_{\bar{l}}\, f = \sum_{\bar{i} \in \mathcal{I}_{\bar{l}}} f(x_{\bar{l},\bar{i}})\psi_{\bar{l},\bar{i}} \qquad (f \in C([0,1]^d)), \tag{33}$$

hence

$$
\begin{aligned}
U_{\bar{l}} S^{(d)} f &= \sum_{\bar{i}\in\mathcal{I}_{\bar{l}}} \left( \int_{[0,x_{\bar{l},\bar{i}}]} f(t)dt \right) \psi_{\bar{l},\bar{i}} \\
&= \sum_{\bar{i}\in\mathcal{I}_{\bar{l}}} \left( \sum_{\bar{1}\le\bar{j}\le\bar{i}} \int_{Q_{\bar{l},\bar{j}}} f(t)dt \right) \psi_{\bar{l},\bar{i}} \,.
\end{aligned}
$$

We are ready to define the Smolyak-Monte Carlo algorithm. Let $\xi_{\bar{l},\bar{i}}$ $(|\bar{l}| = L,\ \bar{1} \le \bar{i} \le \bar{n}_{\bar{l}})$ be independent random variables on a complete probability space $(\Omega, \Sigma, \mathbb{P})$ such that $\xi_{\bar{l},\bar{i}}$ is uniformly distributed on $Q_{\bar{l},\bar{i}}$. Then we approximate

$$
U_{\bar{l}} S^{(d)} f \approx V_{\bar{l}} f := \sum_{\bar{i}\in\mathcal{I}_{\bar{l}}} \left( \sum_{\bar{1}\le\bar{j}\le\bar{i}} |Q_{\bar{l},\bar{j}}| f(\xi_{\bar{l},\bar{j}}) \right) \psi_{\bar{l},\bar{i}} \tag{34}
$$

and thus

$$
S^{(d)} f \approx A^3_{m,L} f := \sum_{\bar{l}\in\mathbb{N}_0^d,\, |\bar{l}|=L} V_{\bar{l}} f. \tag{35}
$$

Now we analyze the error, describe an efficient way to compute the needed quantities and estimate its cost. Let $1 \le p_1 < \infty$, $p_1 \le p$. We shall estimate the $p_1$-st moment of the error. By the triangle inequality, we have

$$
\begin{aligned}
&(\mathbb{E}\, \|S^{(d)} f - A^3_{m,L} f\|^{p_1})^{1/p_1} \\
&\le \|S^{(d)} f - P_L^{(d)} S^{(d)} f\| + (\mathbb{E}\, \|P_L^{(d)} S^{(d)} f - A^3_{m,L} f\|^{p_1})^{1/p_1}.
\end{aligned} \tag{36}
$$

In the following result we summarize the tensor product norm estimates which we will use below. The case $p = \infty$ is particularly important, since in this case, according to (25), the tensor product of the spaces $L_\infty([0,1])$ and $L_\infty([0,1]^{d-1})$ is only a subspace of $L_\infty([0,1]^d)$. The lemma ensures that we can still use product norm estimates.

**Lemma 4.1.** *For $1 \le p \le \infty$, $d > 1$, and any $T_1 \in \mathscr{L}(C([0,1]))$ and $T_2 \in \mathscr{L}(C([0,1]^{d-1}))$ we have*

$$
\begin{aligned}
&\|(T_1 \otimes T_2) S^{(d)} : L_p([0,1]^d) \to C([0,1]^d)\| \\
&= \|T_1 S^{(1)} : L_p([0,1]) \to C([0,1])\| \\
&\quad \times \|T_2 S^{(d-1)} : L_p([0,1]^{d-1}) \to C([0,1]^{d-1})\|.
\end{aligned} \tag{37}
$$

*Proof.* For $1 \le p < \infty$ this follows directly from (22), (27), (23), and (24). For the case $p = \infty$ we note that

$$
\mathcal{B}_{L_\infty([0,1]) \otimes_\lambda L_\infty([0,1]^{d-1})}
$$

is dense in $\mathcal{B}_{L_\infty([0,1]^d)}$ in the norm of $L_1([0,1]^d)$. This is easily deduced from the fact that the linear span of products of characteristic functions $\chi_{D_1} \otimes \chi_{D_2}$, with $D_1 \subseteq [0,1]$ and $D_2 \subseteq [0,1]^{d-1}$ measurable, is dense in $L_1([0,1]^d)$. Moreover, $S^{(d)}$ acts continuously from $L_1([0,1]^d)$ to $C([0,1]^d)$. Consequently,

$$\begin{aligned}
&\|(T_1 \otimes T_2)S^{(d)} : L_\infty([0,1]^d) \to C([0,1]^d)\| \\
&= \|(T_1 \otimes T_2)(S^{(1)} \otimes S^{(d-1)}) : L_\infty([0,1]) \otimes_\lambda L_\infty([0,1]^{d-1}) \to C([0,1]^d)\|,
\end{aligned}$$

from which (37) follows.

□

Now we are ready to estimate the first term on the right-hand side of (36).

**Lemma 4.2.** *Let $1 \le p \le \infty$ and $d \in \mathbb{N}$. Then there is a constant $c(d) > 0$ such that for all $m, L \in \mathbb{N}_0$, $m \ge 2$,*

$$\|S^{(d)} - P_L^{(d)} S^{(d)} : L_p([0,1]^d) \to C([0,1]^d)\| \le c(d)(L+1)^{d-1} m^{-(1-1/p)(L-d+1)}. \quad (38)$$

*Proof.* First note that

$$S^{(1)} \in \mathscr{L}(L_p([0,1]), W_p^1([0,1])),$$

which, by (21), implies

$$\|(I^{(1)} - P_l)S^{(1)} : L_p([0,1]) \to C([0,1])\| \le c m^{-(1-1/p)l} \quad (39)$$

and hence,

$$\|(P_l - P_{l-1})S^{(1)} : L_p([0,1]) \to C([0,1])\| \le c m^{-(1-1/p)(l-1)}. \quad (40)$$

To prove (38), we argue by induction over the dimension $d$. For $d = 1$ the result is just (39). Now let $d > 1$ and assume that (38) holds for $d - 1$. We have

$$\begin{aligned}
&\|S^{(d)} - P_L^{(d)} S^{(d)}\| \\
&\le \|S^{(d)} - (P_L \otimes I^{(d-1)})S^{(d)}\| + \|(P_L \otimes I^{(d-1)})S^{(d)} - P_L^{(d)} S^{(d)}\|. \quad (41)
\end{aligned}$$

Using Lemma 4.1, (39), and (1), the first term is estimated as

$$\begin{aligned}
&\|S^{(d)} - (P_L \otimes I^{(d-1)})S^{(d)}\| \\
&= \|\big((I^{(1)} - P_L) \otimes I^{(d-1)}\big)\big(S^{(1)} \otimes S^{(d-1)}\big)\| \\
&= \|(I^{(1)} - P_L)S^{(1)}\| \, \|S^{(d-1)}\| \le c m^{-(1-1/p)L}.
\end{aligned}$$

The second term of (41) is treated as follows.

$$\begin{aligned}
&\|\big(P_L \otimes I^{(d-1)}\big)S^{(d)} - P_L^{(d)} S^{(d)}\| \\
&= \Big\| \sum_{l=0}^{L} \big((P_l - P_{l-1}) \otimes I^{(d-1)}\big)S^{(d)} - \sum_{l=0}^{L} \big((P_l - P_{l-1}) \otimes P_{L-l}^{(d-1)}\big)S^{(d)} \Big\| \\
&= \Big\| \sum_{l=0}^{L} \big((P_l - P_{l-1})S^{(1)}\big) \otimes \big((I^{(d-1)} - P_{L-l}^{(d-1)})S^{(d-1)}\big) \Big\| \\
&\le \sum_{l=0}^{L} \|(P_l - P_{l-1})S^{(1)}\| \|(I^{(d-1)} - P_{L-l}^{(d-1)})S^{(d-1)}\| \\
&\le c\, c(d-1) \sum_{l=0}^{L} m^{-(1-1/p)(l-1)} (L-l+1)^{d-2} m^{-(1-1/p)(L-l-d+2)} \\
&\le c(d)(L+1)^{d-1} m^{-(1-1/p)(L-d+1)},
\end{aligned}$$

where we used Lemma 4.1, (40) and the induction hypothesis. This proves (38). □

For the further analysis we need the following direct consequence of the Kolmogorov-Doob inequality.

**Lemma 4.3.** *Let $1 < p_1 < \infty$, $\bar{k} \in \mathbb{N}^d$ and let $\{\varrho_{\bar{i}} : \bar{1} \le \bar{i} \le \bar{k}\}$ be independent, mean zero scalar-valued random variables with $\mathbb{E}\,|\varrho_{\bar{i}}|^{p_1} < \infty$ $(\bar{1} \le \bar{i} \le \bar{k})$. Then*

$$\Big(\mathbb{E} \max_{\bar{1} \le \bar{i} \le \bar{k}} \Big| \sum_{\bar{1} \le \bar{j} \le \bar{i}} \varrho_{\bar{j}} \Big|^{p_1}\Big)^{1/p_1} \le c_1^d \Big(\mathbb{E} \Big| \sum_{\bar{1} \le \bar{j} \le \bar{k}} \varrho_{\bar{j}} \Big|^{p_1}\Big)^{1/p_1}, \tag{42}$$

*where $c_1 = p_1/(p_1 - 1)$.*

*Proof.* For $d = 1$ this is just the Kolmogorov-Doob inequality. Now let $d \ge 2$ and assume that (42) holds for $d-1$. We write $\bar{i} = (i', i_d)$, $\bar{k} = (k', k_d)$, $\bar{1} = (1', 1)$, define $K' = \{i' : 1' \le i' \le k'\} \subset \mathbb{N}^{d-1}$ and

$$\zeta_{j_d} = \Bigg(\sum_{1' \le j' \le i'} \varrho_{j', j_d}\Bigg)_{i' \in K'} \in \ell_\infty(K') \qquad (1 \le j_d \le k_d).$$

Then

$$\mathbb{E} \max_{\bar{1} \le \bar{i} \le \bar{k}} \Big| \sum_{\bar{1} \le \bar{j} \le \bar{i}} \varrho_{\bar{j}} \Big|^{p_1} = \mathbb{E} \max_{1 \le i_d \le k_d} \Big\| \sum_{1 \le j_d \le i_d} \zeta_{j_d} \Big\|_{\ell_\infty(K')}^{p_1}. \tag{43}$$

Due to the assumptions,

$$\Big(\sum_{1 \le j_d \le i_d} \zeta_{j_d}\Big)_{1 \le i_d \le k_d}$$

is an $\ell_\infty(K')$-valued martingale, hence

$$\left(\Big\| \sum_{1\le j_d\le i_d} \zeta_{j_d}\Big\|_{\ell_\infty(K')}\right)_{1\le i_d\le k_d}$$

is a non-negative submartingale. Applying the Kolmogorov-Doob inequality we get

$$\begin{aligned}
&\mathbb{E} \max_{1\le i_d\le k_d} \Big\| \sum_{1\le j_d\le i_d} \zeta_{j_d}\Big\|^{p_1}_{\ell_\infty(K')} \le c_1^{p_1}\,\mathbb{E}\Big\| \sum_{1\le j_d\le k_d} \zeta_{j_d}\Big\|^{p_1}_{\ell_\infty(K')} \\
= \;& c_1^{p_1}\,\mathbb{E} \max_{1'\le i'\le k'} \Big| \sum_{1'\le j'\le i'} \Big( \sum_{1\le j_d\le k_d} \varrho_{j',j_d}\Big)\Big|^{p_1} \\
= \;& c_1^{p_1}\,\mathbb{E} \max_{1'\le i'\le k'} \Big| \sum_{1'\le j'\le i'} \eta_{j'}\Big|^{p_1}, \qquad (44)
\end{aligned}$$

with

$$\eta_{j'} = \sum_{1\le j_d\le k_d} \varrho_{j',j_d} \qquad (1'\le j'\le k'). \qquad (45)$$

Since $\{\eta_{j'} : 1'\le j'\le k'\}$ are independent, mean zero random variables with finite $p_1$-st moment, the induction hypothesis implies

$$\mathbb{E} \max_{1'\le i'\le k'} \Big| \sum_{1'\le j'\le i'} \eta_{j'}\Big|^{p_1} \le c_1^{(d-1)p_1}\,\mathbb{E}\Big| \sum_{1'\le j'\le k'} \eta_{j'}\Big|^{p_1}. \qquad (46)$$

Inserting (45) and combining (43), (44), and (46), the desired result follows. □

Now we consider the second term on the right-hand side of (36).

**Lemma 4.4.** *Let $d\in\mathbb{N}$, $1\le p\le\infty$, $\bar p=\min(p,2)$, $1\le p_1<\infty$, $p_1\le p$. Then there is a constant $c(d)>0$ such that for all $m,L\in\mathbb{N}_0$, $m\ge 2$, $f\in L_p(Q)$*

$$(\mathbb{E}\,\|P_L^{(d)}S^{(d)}f - A^3_{m,L}f\|^{p_1})^{1/p_1} \le c(d)(L+1)^{d-1}m^{-(1-1/\bar p)L}\|f\|_{L_p(Q)}.$$

*Proof.* We can assume $\bar p\le p_1$, the remaining cases follow by Hölder's inequality. We have

$$(\mathbb{E}\,\|P_L^{(d)}S^{(d)}f - A^3_{m,L}f\|^{p_1})^{1/p_1} \le \sum_{\bar l\in\mathbb{N}_0^d,\,|\bar l|=L} (\mathbb{E}\,\|U_{\bar l}S^{(d)}f - V_{\bar l}f\|^{p_1})^{1/p_1}.$$

For the further analysis we introduce

$$R_{\bar l}: C(Q)\to\ell_\infty(\mathcal{I}_{\bar l}), \qquad R_{\bar l}f = (f(x_{\bar l,\bar i}))_{\bar i\in\mathcal{I}_{\bar l}}$$

and

$$W_{\bar l}:\ell_\infty(\mathcal{I}_{\bar l})\to C(Q)$$

defined by

$$W_{\bar l}\, z = \sum_{\bar i \in \mathcal{I}_{\bar l}} z_{\bar i} \psi_{\bar l, \bar i} \quad \big(z = (z_{\bar i})_{\bar i \in \mathcal{I}_{\bar l}} \in \ell_\infty(\mathcal{I}_{\bar l})\big).$$

Using (33) and (34), we get

$$U_{\bar l} = W_{\bar l} R_{\bar l}, \tag{47}$$

$$V_{\bar l} f = W_{\bar l} \left( \sum_{\bar 1 \le \bar j \le \bar i} |Q_{\bar l, \bar j}| f(\xi_{\bar l, \bar j}) \right)_{\bar i \in \mathcal{I}_{\bar l}}. \tag{48}$$

We also note that

$$\|W_{\bar l}\| = \|U_{\bar l}\| \le c(d), \tag{49}$$

where the inequality is a consequence of (20) and (32). It follows from (47–49) that

$$\begin{aligned}
\|U_{\bar l} S^{(d)} f - V_{\bar l} f\| &= \Big\| W_{\bar l} \Big( (S^{(d)} f)(x_{\bar l, \bar i}) - \sum_{\bar 1 \le \bar j \le \bar i} |Q_{\bar l, \bar j}| f(\xi_{\bar l, \bar j}) \Big)_{\bar i \in \mathcal{I}_{\bar l}} \Big\| \\
&\le c(d) \max_{\bar i \in \mathcal{I}_{\bar l}} \left| \sum_{\bar 1 \le \bar j \le \bar i} \left( \int_{Q_{\bar l, \bar j}} f(t) dt - |Q_{\bar l, \bar j}| f(\xi_{\bar l, \bar j}) \right) \right| \\
&= c(d) \max_{\bar i \in \mathcal{I}_{\bar l}} \Big| \sum_{\bar 1 \le \bar j \le \bar i} \eta_{\bar l, \bar j} \Big|
\end{aligned} \tag{50}$$

with

$$\eta_{\bar l, \bar j} = \int_{Q_{\bar l, \bar j}} f(t) dt - |Q_{\bar l, \bar j}| f(\xi_{\bar l, \bar j}).$$

The random variables $\{\eta_{\bar l, \bar j} : \bar j \in \mathcal{I}_{\bar l}\}$ are independent, of mean zero, and satisfy

$$\begin{aligned}
(\mathbb{E}\, |\eta_{\bar l, \bar j}|^{p_1})^{1/p_1} &\le 2 |Q_{\bar l, \bar j}| (\mathbb{E}\, |f(\xi_{\bar l, \bar j})|^{p_1})^{1/p_1} \\
&= 2 |Q_{\bar l, \bar j}|^{1-1/p_1} \Big( \int_{Q_{\bar l, \bar j}} |f(t)|^{p_1} dt \Big)^{1/p_1}.
\end{aligned} \tag{51}$$

For $p_1 > 1$ we get from Lemma 4.3.

$$\Big( \mathbb{E} \max_{\bar i \in \mathcal{I}_{\bar l}} \Big| \sum_{\bar 1 \le \bar j \le \bar i} \eta_{\bar l, \bar j} \Big|^{p_1} \Big)^{1/p_1} \le c(d) \Big( \mathbb{E} \Big| \sum_{\bar j \in \mathcal{I}_{\bar l}} \eta_{\bar l, \bar j} \Big|^{p_1} \Big)^{1/p_1}. \tag{52}$$

Moreover, since $p_1 \ge \bar p$, Lemma 3.2 gives

$$\Big( \mathbb{E} \Big| \sum_{\bar j \in \mathcal{I}_{\bar l}} \eta_{\bar l, \bar j} \Big|^{p_1} \Big)^{1/p_1} \le c \Big( \sum_{\bar j \in \mathcal{I}_{\bar l}} (\mathbb{E}\, |\eta_{\bar l, \bar j}|^{p_1})^{\bar p / p_1} \Big)^{1/\bar p}. \tag{53}$$

From (52) and (53) we conclude for $p_1 > 1$

$$\Big(\mathbb{E}\max_{\bar{i}\in\mathcal{I}_{\bar{l}}}\Big|\sum_{\bar{1}\le\bar{j}\le\bar{i}}\eta_{\bar{l},\bar{j}}\Big|^{p_1}\Big)^{1/p_1} \le c(d)\Big(\sum_{\bar{j}\in\mathcal{I}_{\bar{l}}}(\mathbb{E}\,|\eta_{\bar{l},\bar{j}}|^{p_1})^{\bar{p}/p_1}\Big)^{1/\bar{p}}. \tag{54}$$

The same relation also holds for $p_1 = 1$ (implying $\bar{p} = 1$, by our assumption $\bar{p} \le p_1$), which follows with $c(d) = 1$ from the triangle inequality. Using (51) and, if $\bar{p} < p_1$, Hölder's inequality with exponent $p_1/\bar{p}$, we obtain

$$\begin{aligned}
&\sum_{\bar{j}\in\mathcal{I}_{\bar{l}}}(\mathbb{E}\,|\eta_{\bar{l},\bar{j}}|^{p_1})^{\bar{p}/p_1}\\
&\le\ 2^{\bar{p}}\max_{\bar{j}\in\mathcal{I}_{\bar{l}}}|Q_{\bar{l},\bar{j}}|^{\bar{p}-1}\sum_{\bar{j}\in\mathcal{I}_{\bar{l}}}|Q_{\bar{l},\bar{j}}|^{1-\frac{\bar{p}}{p_1}}\Big(\int_{Q_{\bar{l},\bar{j}}}|f(t)|^{p_1}dt\Big)^{\frac{\bar{p}}{p_1}}\\
&\le\ 2^{\bar{p}}\max_{\bar{j}\in\mathcal{I}_{\bar{l}}}|Q_{\bar{l},\bar{j}}|^{\bar{p}-1}\Big(\sum_{\bar{j}\in\mathcal{I}_{\bar{l}}}|Q_{\bar{l},\bar{j}}|\Big)^{1-\frac{\bar{p}}{p_1}}\Big(\sum_{\bar{j}\in\mathcal{I}_{\bar{l}}}\int_{Q_{\bar{l},\bar{j}}}|f(t)|^{p_1}dt\Big)^{\frac{\bar{p}}{p_1}}\\
&\le\ 2^{\bar{p}}\max_{\bar{j}\in\mathcal{I}_{\bar{l}}}|Q_{\bar{l},\bar{j}}|^{\bar{p}-1}\,\|f\|^{\bar{p}}_{L_{p_1}(Q)}.
\end{aligned} \tag{55}$$

Combining (50), (54), and (55), it follows that

$$\big(\mathbb{E}\,\|U_{\bar{l}}S^{(d)}f - V_{\bar{l}}f\|^{p_1}\big)^{1/p_1} \le c(d)\max_{\bar{j}\in\mathcal{I}_{\bar{l}}}|Q_{\bar{l},\bar{j}}|^{1-1/\bar{p}}\,\|f\|_{L_p(Q)}. \tag{56}$$

Taking into account that by (19), (28), and (29),

$$\max_{1\le i\le\hat{n}_l}(\hat{x}_{l,i}-\hat{x}_{l,i-1}) \le \max_{1\le i\le n_l+1}(x_{l,i}-x_{l,i-1}) \le cm^{-l},$$

we get

$$|Q_{\bar{l},\bar{i}}| = (x_{l_d,i_d}-x_{l_d,i_d-1})\prod_{k=1}^{d-1}(\hat{x}_{l_k,i_k}-\hat{x}_{l_k,i_k-1}) \le c(d)m^{-L} \qquad (\bar{i}\in\mathcal{I}_{\bar{l}},\ |\bar{l}| = L).$$

Together with (31), (35), and (56) we obtain

$$(\mathbb{E}\,\|P_L^{(d)}S^{(d)}f - A^3_{m,L}f\|^{p_1})^{1/p_1} \le c(d)(L+1)^{d-1}m^{-(1-1/\bar{p})L}\|f\|_{L_p(Q)},$$

which proves Lemma 4.4.

□

**Theorem 4.5.** *Let $d\in\mathbb{N}$, $1\le p\le\infty$, $\bar{p} = \min(p,2)$, $1\le p_1<\infty$, $p_1\le p$. Then there are constants $c_{1-4}(d) > 0$ such that for all $m\in\mathbb{N}$, $m\ge 2$, $L\in\mathbb{N}_0$ the algorithm $A^3_{m,L}$ uses not more than $c_1(d)(L+1)^{d-1}m^L$ function values and the error satisfies for each $f\in L_p(Q)$*

$$\begin{aligned}
&(\mathbb{E}\,\|S^{(d)}f - A^3_{m,L}f\|^{p_1})^{1/p_1}\\
&\le\ c_2(d)(L+1)^{d-1}\big(m^{-(1-1/p)(L-d+1)} + m^{-(1-1/\bar{p})L}\big)\|f\|_{L_p(Q)}.
\end{aligned} \tag{57}$$

*Moreover, for each $n \in \mathbb{N}$ with $n \geq 2$ there is a choice of the parameters $m$ and $L$ such that the algorithm uses not more than $c_3(d)n$ function values and the error can be estimated for $f \in L_p(Q)$ as*

$$(\mathbb{E}\,\|S^{(d)}f - A^3_{m,L}f\|^{p_1})^{1/p_1} \leq \begin{cases} c_4(d)n^{-1/2}\|f\|_{L_p(Q)} & \text{if} \quad 2 < p \leq \infty \\ c_4(d)(\log n)^{(2-1/p)(d-1)}n^{-1+1/p}\|f\|_{L_p(Q)} & \text{if} \quad 1 \leq p \leq 2. \end{cases} \tag{58}$$

*Proof.* Relation (57) follows readily from Lemmas 4.2 and 4.4. By (18), (28–29), and (34–35), the number of function values used in $A^3_{m,L}f$ is

$$\sum_{|\bar{l}|=L} \hat{n}_{l_1} \ldots \hat{n}_{l_{d-1}} n_{l_d} \leq c(d)(L+1)^{d-1}m^L. \tag{59}$$

To show the second part we first assume $2 < p \leq \infty$. Then $\bar{p} = 2$ and we put

$$L = \left\lceil \frac{2(p-1)(d-1)}{p-2} \right\rceil, \qquad m = \left\lceil n^{\frac{1}{L}} \right\rceil. \tag{60}$$

With this choice we have $n \leq m^L \leq 2^L n$ and

$$\begin{aligned} &\left(1 - \frac{1}{p}\right)(L - d + 1) \\ &\geq \frac{L}{2} + \left(\frac{1}{2} - \frac{1}{p}\right)\frac{2(p-1)(d-1)}{p-2} - \left(1 - \frac{1}{p}\right)(d-1) = \frac{L}{2}, \end{aligned}$$

which gives (58). Now let $1 \leq p \leq 2$, hence $\bar{p} = p$. For $n \leq (\log n)^{d-1}$, that is, $n \leq c(d)$ for some constant $c(d)$, the result follows trivially from (57) (with suitably chosen $c_3(d), c_4(d)$). If $n > (\log n)^{d-1}$, then the (standard) choice

$$m = 2, \qquad L = \lceil \log n - (d-1)\log\log n \rceil \tag{61}$$

implies

$$n(\log n)^{-(d-1)} \leq m^L \leq 2n(\log n)^{-(d-1)},$$

which yields (58). □

Note that in (58) of Theorem 4.5 we obtain for $p = 1$ no convergence to zero as $n \to \infty$. The lower bound in Proposition 5.1 shows that in this case no algorithm at all has an error converging to zero.

Let us comment on the arithmetic work required for the computation of $A^3_{m,L}f$ as given in (34–35) (we always assume the real number model, see [21, 17] and, for more details, [15]). Clearly, for the $\xi_{\bar{l},\bar{i}}$ ($|\bar{l}| = L$, $\bar{1} \leq \bar{i} \leq \bar{n}_{\bar{l}}$) we need

$$d \sum_{|\bar{l}|=L} \hat{n}_{l_1} \ldots \hat{n}_{l_{d-1}} n_{l_d}$$

independent random variables uniformly distributed on $[0,1]$. Taking into account (59), this number is

$$\leq c(d)(L+1)^{d-1}m^L \leq c(d)n$$

for each of the choices (60) and (61).

In order to compute the coefficients of the functions $\psi_{\bar{l},\bar{i}}$ in (34–35), for each $\bar{l}$ with $|\bar{l}| = L$ we have to carry out a task of the following type. Given $\bar{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$ and numbers $(a_{\bar{i}})_{\bar{1}\leq\bar{i}\leq\bar{k}}$, compute $(b_{\bar{i}})_{\bar{1}\leq\bar{i}\leq\bar{k}}$, where

$$b_{\bar{i}} = \sum_{\bar{1}\leq\bar{j}\leq\bar{i}} a_{\bar{j}}.$$

We show how this can be done with at most $c_0(d)k_1 \dots k_d$ arithmetic operations. For $d = 1$ with $c_0(1) = 1$ this is obvious. Now we use recursion. So let $d > 1$ and assume we have a suitable procedure for $d-1$. Let us write $\bar{k} = (k', k_d)$, and $\bar{i} = (i', i_d)$. We compute for each $j_d \in \{1, 2, \dots, k_d\}$

$$v_{i',j_d} = \sum_{1'\leq j'\leq i'} a_{j',j_d} \qquad (1' \leq i' \leq k')$$

by the procedure for dimension $d-1$ (thus, we compute the sums in the $j_d$-th 'layer'). Then for each $1' \leq i' \leq k'$ we determine

$$b_{i',i_d} = \sum_{1\leq j_d\leq i_d} v_{i',j_d} \quad (1 \leq i_d \leq k_d).$$

Clearly, this needs a total of

$$k_d \cdot c_0(d-1)k_1 \dots k_{d-1} + k_1 \dots k_{d-1} \cdot k_d = c_0(d)k_1 \dots k_d$$

operations, and we get $c_0(d) = c_0(d-1) + 1$, hence $c_0(d) = d$. Using again (59), the work of computing all coefficients in (34–35) is

$$d \sum_{|\bar{l}|=L} \hat{n}_{l_1} \dots \hat{n}_{l_{d-1}} n_{l_d} \leq c(d)(L+1)^{d-1}m^L \leq c(d)n.$$

Finally we consider the cost of computing the value $\left(A^3_{m,L}f\right)(x)$ for a given $x \in Q$, once the coefficients in (34–35) have been determined. For this purpose we assume that the functions $\psi_{m,l,i}$ $(i = 1, \dots, n_{m,l})$, see (17), have the following properties: There are constants $c_{1-3} > 0$ such that for all $m, l \in \mathbb{N}_0$, $m \geq 2$,

$$\sup_{t\in[0,1]} |\{i : \psi_{m,l,i}(t) \neq 0\}| \leq c_1, \tag{62}$$

furthermore, given $m, l, t$, the cost of identifying those $i \in \{1, \dots, n_{m,l}\}$ with $\psi_{m,l,i}(t) \neq 0$ is $\leq c_2$ and the cost of computing $\psi_{m,l,i}(t)$ for any such $i$ is $\leq c_3$. These properties hold, in particular, for piecewise linear interpolation as described after (18–21) (here we assume that our model of computation allows to take the integer part at cost $\le c$, which is needed to identify the indices $i$).

The assumptions imply that the corresponding statements also hold for the $\hat{\psi}_{m,l,i}$ $(i = 1, \dots, \hat{n}_{m,l})$ and therefore also for the $\psi_{\bar{l},\bar{i}}$ $(\bar{i} \in \mathcal{I}_{\bar{l}})$. Hence, the number of non-zero terms $\psi_{\bar{l},\bar{i}}(x)$ in (34–35) is

$$\left|\left\{(\bar{l}, \bar{i}) : |\bar{l}| = L, \, \bar{i} \in \mathcal{I}_{\bar{l}}, \, \psi_{\bar{l},\bar{i}}(x) \neq 0\right\}\right| \le c(d)L^{d-1}.$$

Moreover, the cost of identifying and computing them is $\le c(d)L^{d-1}$, as well. Thus, the cost of computing the value $\left(A^3_{m,L}f\right)(x)$ is $\le c(d)(L+1)^{d-1}$, therefore $\le c(d)$ for the choice (60) in the case $2 < p \le \infty$ and $\le c(d)(\log n)^{d-1}$ for the choice (61) in the case $1 \le p \le 2$.

# 5 Lower bounds and complexity

For basic notions concerning the randomized setting of information-based complexity – the framework we use – we refer to [14, 21, 4]. Here we consider the class of all randomized adaptive algorithms of varying cardinality. We refer to [5, 6] for this approach, the particular notation applied here, and more details.

First we introduce the respective deterministic class. An element

$$A \in \mathcal{A}^{\det}(\mathcal{F}(Q), Y)$$

is a tuple

$$A = ((L_i)_{i=1}^{\infty}, (\tau_i)_{i=0}^{\infty}, (\varphi_i)_{i=0}^{\infty})$$

such that

$$L_1 \in Q, \quad \tau_0 \in \{0, 1\}, \quad \varphi_0 \in Y,$$

and

$$\begin{aligned} L_i : \; & \mathbb{K}^{i-1} \to Q & (i = 2, 3, \dots) \\ \tau_i : \; & \mathbb{K}^{i} \to \{0, 1\} & (i = 1, 2, \dots) \\ \varphi_i : \; & \mathbb{K}^{i} \to Y & (i = 1, 2, \dots) \end{aligned}$$

are arbitrary mappings. Given $f \in \mathcal{F}(Q)$, we associate with it a sequence $(t_i)_{i=1}^{\infty}$ with $t_i \in Q$, defined as follows:

$$t_1 = L_1 \tag{63}$$

$$t_i = L_i(f(t_1), \dots, f(t_{i-1})) \quad (i \ge 2). \tag{64}$$

Define $\mathrm{card}(A, f)$, the cardinality of $A$ at input $f$, to be 0 if $\tau_0 = 1$. If $\tau_0 = 0$, let $\mathrm{card}(A, f)$ be the first integer $n \ge 1$ with

$$\tau_n(f(t_1), \dots, f(t_n)) = 1,$$

if there is such an $n$. If $\tau_0 = 0$ and no such $n \in \mathbb{N}$ exists, put $\operatorname{card}(A, f) = +\infty$. For $f \in \mathcal{F}(Q)$ with $\operatorname{card}(A, f) < \infty$ we define the output $Af$ of algorithm $A$ at input $f$ as

$$Af = \begin{cases} \varphi_0 & \text{if} \quad n = 0 \\ \varphi_n(f(t_1), \dots, f(t_n)) & \text{if} \quad n \geq 1. \end{cases}$$

Given $n \in \mathbb{N}_0$ and $F \subseteq \mathcal{F}(Q)$, we define $\mathcal{A}_n^{\det}(F, Y)$ as the set of those $A \in \mathcal{A}^{\det}(\mathcal{F}(Q), Y)$ for which

$$\max_{f \in F} \operatorname{card}(A, f) \leq n.$$

Given a mapping $S : F \to Y$, the error of $A \in \mathcal{A}_n^{\det}(F, Y)$ in approximating $S$ is defined as

$$e(S, A, F, Y) = \sup_{f \in F} \|Sf - Af\|_Y.$$

The deterministic $n$-th minimal error of $S$ is defined for $n \in \mathbb{N}_0$ as

$$e_n^{\det}(S, F, Y) = \inf_{A \in \mathcal{A}_n^{\det}(F,Y)} e(S, A, F, Y). \tag{65}$$

It follows that no deterministic algorithm that uses at most $n$ function values can have a smaller error than $e_n^{\det}(S, F, Y)$.

Next we introduce the class of randomized adaptive algorithms of varying cardinality. We do this for the case that $F$ consists of equivalence classes of functions, as needed for this paper, following the approach of [7]. The case of $F$ being a set of functions can be found in [5, 6]. Let $1 \leq p \leq \infty$ and let $F \subseteq L_p(Q)$. An element

$$A \in \mathcal{A}_n^{\text{ran}}(F, Y)$$

is a tuple

$$A = ((\Omega, \Sigma, \mathbb{P}), (A_\omega)_{\omega \in \Omega}),$$

where $(\Omega, \Sigma, \mathbb{P})$ is a probability space,

$$A_\omega \in \mathcal{A}^{\det}(\mathcal{F}(Q), Y) \quad (\omega \in \Omega), \tag{66}$$

and the following two properties are satisfied.

1. For each $f \in F$ and each representative $f_0$ of $f$ the mapping

   $$\omega \in \Omega \to \operatorname{card}(A_\omega, f_0)$$

   is $\Sigma$-measurable and satisfies

   $$\mathbb{E} \operatorname{card}(A_\omega, f_0) \leq n.$$

   Moreover, the mapping

   $$\omega \in \Omega \to A_\omega f_0 \in Y$$

   is $\Sigma$-to-Borel measurable and essentially separably valued, i.e., there is a separable subspace $Y_0 \subseteq Y$ such that

   $$A_\omega f \in Y_0 \quad \text{for } \mathbb{P}\text{-almost all } \omega \in \Omega.$$

2. If $f_0$ and $f_1$ are representatives of the same class $f \in F$, then $\mathbb{P}$-almost surely

$$\begin{aligned} \operatorname{card}(A_\omega, f_0) &= \operatorname{card}(A_\omega, f_1), \\ A_\omega f_0 &= A_\omega f_1. \end{aligned}$$

Consequently, we can define the output $Af$ of algorithm $A$ at input $f \in F \subseteq L_p(Q)$ as the $Y$-valued random variable $A_\omega f_0$ on $(\Omega, \Sigma, \mathbb{P})$, where $f_0$ is any representative of $f$. By the above, another choice of $f_0$ leads – up to equivalence – to the same random variable.

It is readily seen that

$$A^2_{n,l} \in \mathcal{A}^{\rm ran}_n(\mathcal{B}_{L_p(Q)}, C(Q))$$

and

$$A^3_{m,L} \in \mathcal{A}^{\rm ran}_n(\mathcal{B}_{L_p(Q)}, C(Q)) \quad \text{for} \quad n \geq c_1(d)(L+1)^{d-1} m^L \tag{67}$$

(see Theorem 4.5 for the estimate of the number of samples in (67)). Here we use the completeness of the measure $\mathbb{P}$ stated at the beginning of Section 3 and asssumed throughout the paper. Algorithm $A^1_n$ is of the required form (with $Y = B_0(Q)$), satisfies property 2, but not 1. The latter is discussed in Section 6.3.

Given a mapping $S : F \to Y$, the error of $A \in \mathcal{A}^{\rm ran}_n(\mathcal{F}(Q), Y)$ as an approximation of $S$ on $F$ is defined as

$$e(S, A, F, Y) = \sup_{f \in F} \mathbb{E} \, \|Sf - A_\omega f\|_Y. \tag{68}$$

The randomized $n$-th minimal error of $S$ is defined for $n \in \mathbb{N}_0$ as

$$e^{\rm ran}_n(S, F, Y) = \inf_{A \in \mathcal{A}^{\rm ran}_n(F,Y)} e(S, A, F, Y). \tag{69}$$

Consequently, no randomized linear algorithm that uses (on the average) at most $n$ function values has an error smaller than $e^{\rm ran}_n(S, F, Y)$. Note that the definition (68) involves the first moment. This way lower bounds have the strongest form, because respective bounds for higher moments follow by Hölder's inequality. In Sections 3 and 4 upper bounds for concrete algorithms were stated in a form which included possible estimates of higher moments.

Define for $\varepsilon > 0$ the information complexity as the inverse function of the $n$-th minimal error

$$n^{\rm ran}_\varepsilon(S, F, Y) = \min\{n \in \mathbb{N}_0 : \, e^{\rm ran}_n(S, F, Y) \leq \varepsilon\}, \tag{70}$$

if there is such an $n$, and

$$n^{\rm ran}_\varepsilon(S, F, Y) = +\infty, \tag{71}$$

if there is no such $n$. Thus, if $n_\varepsilon^{\rm ran}(S,F,Y)<\infty$, it follows that any algorithm with error $\le\varepsilon$ needs at least $n_\varepsilon^{\rm ran}(S,F,Y)$ samples, while (71) means that no algorithm at all has error $\le\varepsilon$.

Now let $\nu$ be a probability measure on $\mathcal{F}(Q)$ whose support, denoted by $\operatorname{supp}\nu$, is a finite set and satisfies $\operatorname{supp}\nu\subseteq F$ (meaning, more precisely, that each function from $\operatorname{supp}\nu$ belongs to a class from $F$). For $A\in\mathcal{A}^{\rm det}(\mathcal{F}(Q),Y)$ put

$$\begin{aligned}\operatorname{card}(A,\nu) &= \int_{\mathcal{F}(Q)} \operatorname{card}(A,f)\,d\nu(f),\\ e(S,A,\nu,Y) &= \int_{\mathcal{F}(Q)} \|Sf-Af\|_Y\,d\nu(f)\end{aligned}$$

and define the average $n$-th minimal error as

$$e_n^{\rm avg}(S,\nu,Y)=\inf\{e(S,A,\nu,Y)\,:\,A\in\mathcal{A}^{\rm det}(\mathcal{F}(Q),Y),\ \operatorname{card}(A,\nu)\le n\}.$$

Then the following holds

$$e_n^{\rm ran}(S,F,Y)\ge\frac{1}{2}\,e_{2n}^{\rm avg}(S,\nu,Y). \tag{72}$$

This follows from the usual relation between randomized and average case setting, going back to Bakhvalov, see [14, 4, 17].

We also consider two smaller classes of algorithms. The first one is the class of non-adaptive algorithms $\mathcal{A}_n^{\rm det,1}(\mathcal{F}(Q),Y)$. We define $A\in\mathcal{A}_n^{\rm det,1}(\mathcal{F}(Q),Y)$ if $A\in\mathcal{A}^{\rm det}(\mathcal{F}(Q),Y)$ and the respective functions $L_i$ and $\tau_i$ are constant and satisfy

$$\tau_0=\tau_1=\dots=\tau_{n-1}=0,\quad \tau_n=1.$$

Thus, an element $A$ of $\mathcal{A}_n^{\rm det,1}(\mathcal{F}(Q),Y)$ generates a mapping from $F(Q)$ to $Y$ of the form

$$Af=\begin{cases}\varphi_0 & \text{if}\quad n=0\\ \varphi_n(f(t_1),\dots,f(t_n)) & \text{if}\quad n\ge 1\end{cases}\qquad (f\in\mathcal{F}(Q))$$

with $\varphi_0\in Y$, $t_i\in Q$ $(i=1,\dots,n)$, not depending on $f$, and $\varphi_n:\mathbb{K}^n\to Y$ an arbitrary mapping.

The second class $\mathcal{A}_n^{\rm det,2}(\mathcal{F}(Q),Y)$ is the class of linear algorithms, that is, the set of all $A\in\mathcal{A}_n^{\rm det,1}(\mathcal{F}(Q),Y)$ with $\varphi_n$ linear. In other words, an element $A\in\mathcal{A}_n^{\rm det,2}(\mathcal{F}(Q),Y)$ has the form

$$Af=\begin{cases}0 & \text{if}\quad n=0\\ \displaystyle\sum_{i=1}^n f(t_i)\psi_i & \text{if}\quad n\ge 1\end{cases}\qquad (f\in\mathcal{F}(Q))$$

with $t_i\in Q$ and $\psi_i\in Y$ for $1\le i\le n$.

For $j=1,2$ we define $\mathcal{A}_n^{\mathrm{ran},j}(F,Y)$ as the set of all $A\in\mathcal{A}_n^{\mathrm{ran}}(F,Y)$ with

$$A_\omega \in \mathcal{A}_n^{\mathrm{det},j}(\mathcal{F}(Q),Y) \quad (\omega\in\Omega).$$

We note that the algorithms constructed in Sections 3 and 4 are linear in the sense that

$$\begin{aligned} A_{n,l}^2 &\in \mathcal{A}_n^{\mathrm{ran},2}(\mathcal{B}_{L_p(Q)},C(Q)) \quad (n\in\mathbb{N}),\\ A_{m,L}^3 &\in \mathcal{A}_n^{\mathrm{ran},2}(\mathcal{B}_{L_p(Q)},C(Q)) \quad (n\ge c_1(d)(L+1)^{d-1}m^L),\end{aligned}$$

and the operators $A_{n,\omega}^1$ constituting algorithm $A_n^1$, see (4), are linear, as well.

By analogy to the above, we define for $j=1,2$ the respective $n$-th minimal errors $e_n^{\mathrm{det},j}(S,F,Y)$, $e_n^{\mathrm{ran},j}(S,F,Y)$, the information complexities $n_\varepsilon^{\mathrm{ran}}(S,F,Y)$, and the average $n$-th minimal errors $e_n^{\mathrm{avg},j}(S,\nu,Y)$. The quantities $e_n^{\mathrm{det},2}(S,F,Y)$ were also called linear sampling numbers in [16], the $e_n^{\mathrm{det},1}(S,F,Y)$ nonlinear sampling numbers. Thus, the $e_n^{\mathrm{ran},j}(S,F,Y)$ $(j=1,2)$ could be viewed as the respective randomized counterparts.

In these cases slightly sharper lower bounds through the average case can be given:

$$e_n^{\mathrm{ran},j}(S,F,Y)\ge e_n^{\mathrm{avg},j}(S,\nu,Y) \quad (j=1,2). \tag{73}$$

We prove three lower bounds for the randomized $n$-th minimal error. The first one is standard and contains the sharp order in $n$. It has a constant independent of $d$, but it does not match the positive power of $d$ in the upper estimate.

**Proposition 5.1.** *Let $1\le p\le\infty$ and $\bar p=\min(p,2)$. Then there is a constant $c>0$ such that for all $d,n\in\mathbb{N}$*

$$e_n^{\mathrm{ran}}(S^{(d)},\mathcal{B}_{L_p([0,1]^d)},C([0,1]^d))\ge cn^{-1+1/\bar p}.$$

*Proof.* We write $t=(t_1,t')\in[0,1]^d$ with $t_1\in[0,1]$ and $t'\in[0,1]^{d-1}$. Let $0<\delta<1$ and let

$$R_\delta:\mathcal{F}([0,1])\to\mathcal{F}([0,1]^d)$$

be defined by

$$(R_\delta f)(t_1,t')=\begin{cases} f((1-\delta)^{-1}t_1) & \text{if} \quad 0\le t_1\le 1-\delta\\ 0 & \text{otherwise,}\end{cases}.$$

If $f_0,f_1\in\mathcal{F}([0,1])$ coincide except for a set of Lebesgue measure zero, the same is true for $R_\delta f_0, R_\delta f_1\in\mathcal{F}([0,1]^d)$. Moreover,

$$\|R_\delta: L_p([0,1])\to L_p([0,1]^d)\| = (1-\delta)^{1/p}\le 1. \tag{74}$$

Define

$$\Psi_\delta: C([0,1]^d)\to\mathbb{K}, \quad \Psi_\delta\, g=\delta^{-d}\int_{[1-\delta,1]^d} g(x)dx.$$

Then

$$\|\Psi_\delta : C([0,1]^d) \to \mathbb{K}\| = 1. \tag{75}$$

Finally, let

$$S_1 : L_p([0,1]) \to \mathbb{K}, \quad S_1 f = \int_0^1 f(t)dt$$

be the integration operator. Then for $x = (x_1, \dots, x_d)$ with $1-\delta \le x_1 \le 1$ and $f \in L_p([0,1])$

$$\left(S^{(d)} R_\delta f\right)(x) = x_2 \dots x_d (1-\delta) \int_0^1 f(t)dt,$$

and hence,

$$\Psi_\delta S^{(d)} R_\delta f = \delta^{-d} \left(\delta - \frac{\delta^2}{2}\right)^{d-1} \delta(1-\delta)\, S_1 f = \gamma(d,\delta)\, S_1 f, \tag{76}$$

with

$$\gamma(d,\delta) = \left(1 - \frac{\delta}{2}\right)^{d-1} (1-\delta)\,.$$

Now let

$$A = ((\Omega, \Sigma, \mathbb{P}), (A_\omega)_{\omega\in\Omega}) \in \mathcal{A}_n^{\mathrm{ran}}(\mathcal{B}_{L_p([0,1]^d)}, C([0,1]^d)).$$

By Lemma 2 of [6], for each $\omega \in \Omega$ there is an $A_{1,\omega} \in \mathcal{A}^{\mathrm{det}}(\mathcal{F}([0,1]), \mathbb{K})$ such that for all $f \in \mathcal{F}([0,1])$

$$\mathrm{card}(A_{1,\omega}, f) = \mathrm{card}(A_\omega, R_\delta f)$$

and, if $\mathrm{card}(A_{1,\omega}, f) < \infty$,

$$A_{1,\omega} f = \gamma(d,\delta)^{-1} \Psi_\delta A_\omega R_\delta f.$$

It follows that

$$A_1 = ((\Omega, \Sigma, \mathbb{P}), (A_{1,\omega})_{\omega\in\Omega}) \in \mathcal{A}_n^{\mathrm{ran}}(\mathcal{B}_{L_p([0,1])}, \mathbb{K}).$$

Moreover, because of (74–76),

$$\begin{aligned} e(S_1, A_1, \mathcal{B}_{L_p([0,1])}, \mathbb{K}) &= \sup_{f\in\mathcal{B}_{L_p([0,1])}} \mathbb{E}\,|S_1 f - A_{1,\omega} f| \\ &= \gamma(d,\delta)^{-1} \sup_{f\in\mathcal{B}_{L_p([0,1])}} \mathbb{E}\,|\Psi_\delta S^{(d)} R_\delta f - \Psi_\delta A_\omega R_\delta f| \\ &\le \gamma(d,\delta)^{-1} \sup_{g\in\mathcal{B}_{L_p([0,1]^d)}} \mathbb{E}\,\|S^{(d)} g - A_\omega g\| \\ &= \gamma(d,\delta)^{-1} e(S^{(d)}, A, \mathcal{B}_{L_p([0,1]^d)}, C([0,1]^d)). \end{aligned}$$

Consequently

$$\gamma(d,\delta)\, e_n^{\mathrm{ran}}(S_1, \mathcal{B}_{L_p([0,1])}, \mathbb{K}) \le e_n^{\mathrm{ran}}(S^{(d)}, \mathcal{B}_{L_p([0,1]^d)}, C([0,1]^d)).$$

Finally, the lower bound for integration is well-known, see [14, 4],

$$e_n^{\mathrm{ran}}(S_1, \mathcal{B}_{L_p([0,1])}, \mathbb{K}) \geq c_1 n^{-1+1/\bar{p}}.$$

With $\gamma(d,\delta) \to 1$ for $d$ fixed and $\delta \to 0$ the result follows. $\square$

The second lower bound is not sharp in $n$, but gives more information about the dependence on $d$. See also [9], the proof of Theorem 8, for a similar approach in the deterministic case.

**Proposition 5.2.** *Let $1 \leq p \leq \infty$. Then there is a constant $c > 0$ such that for all $d, n \in \mathbb{N}$*

$$e_n^{\mathrm{ran}}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \geq c\, 2^{-4n/d}.$$

*Proof.* Since $\mathcal{B}_{L_\infty(Q)} \subset \mathcal{B}_{L_p(Q)}$, it suffices to consider the case $p = \infty$. We use the following fact (see, e.g., [10], proof of Theorem 2): There is a constant $0 < c_0 \leq 1$ such that for each $d \in \mathbb{N}$ and each $0 < \varepsilon \leq 1$ there is a set $\mathscr{U} \subset [0,1]^d$ with

$$|\mathscr{U}| \geq \left(\frac{c_0}{\varepsilon}\right)^d, \tag{77}$$

$$|[0,u] \div [0,v]| \geq \varepsilon \qquad (u, v \in \mathscr{U},\ u \neq v), \tag{78}$$

where $\div$ denotes the symmetric difference. Let $u, v \in \mathscr{U}$, $u \neq v$. Then (78) gives

$$\begin{aligned} &\|S^{(d)}\chi_{[0,u]} - S^{(d)}\chi_{[0,v]}\|_{C(Q)} \\ &\geq \max_{x \in \{u,v\}} \left| \int_{[0,x]} \left(\chi_{[0,u]}(t) - \chi_{[0,v]}(t)\right) dt \right| \\ &= \max\Big(\big|[0,u] \setminus [0,v]\big|, \big|[0,v] \setminus [0,u]\big|\Big) \geq \varepsilon/2. \end{aligned} \tag{79}$$

Let $\nu$ be the uniform distribution on the set

$$\{\chi_{[0,u]} : u \in \mathscr{U}\} \subset \mathcal{F}(Q). \tag{80}$$

Given $n \in \mathbb{N}$, we put

$$\varepsilon = c_0\, 2^{-(4n+4)/d}. \tag{81}$$

Now we estimate $e_{2n}^{\mathrm{avg}}(S^{(d)}, \nu, C(Q))$ from below. So let $A \in \mathcal{A}^{\mathrm{det}}(\mathcal{F}(Q), C(Q))$, with

$$\mathrm{card}(A, \nu) = \int_{\mathcal{F}(Q)} \mathrm{card}(A, f)\ d\nu(f) \leq 2n. \tag{82}$$

Let

$$\mathscr{U}_0 = \{u \in \mathscr{U} : \mathrm{card}(A, \chi_{[0,u]}) \leq 4n\}.$$

It follows from (82) that

$$|\mathscr{U}_0| \geq \frac{1}{2}|\mathscr{U}| \geq \frac{1}{2}\left(\frac{c_0}{\varepsilon}\right)^d. \tag{83}$$

For $u \in \mathscr{U}_0$ let $(t_{u,i})_{i\in\mathbb{N}}$ be the respective sequence associated with $A$ and $\chi_{[0,u]}$ according to (63–64), and let $n_u = \operatorname{card}(A, \chi_{[0,u]})$. Define

$$\mathscr{T} = \left\{ (\chi_{[0,u]}(t_{u,i}))_{i=1}^{n_u} \,:\, u \in \mathscr{U}_0 \right\} \subseteq \bigcup_{k\le 4n} \{0,1\}^k. \tag{84}$$

This implies

$$|\{A\chi_{[0,u]} \,:\, u \in \mathscr{U}_0\}| \le |\mathscr{T}| \tag{85}$$

and

$$|\mathscr{T}| < 2^{4n+1}. \tag{86}$$

From (79) and (85) we get

$$|\{u \in \mathscr{U}_0 \,:\, \|S^{(d)}\chi_{[0,u]} - A\chi_{[0,u]}\| < \varepsilon/4\}| \le |\mathscr{T}|,$$

and therefore

$$e_{2n}^{\mathrm{avg}}(S^{(d)}, \nu, C(Q)) \ge \frac{|\mathscr{U}_0| - |\mathscr{T}|}{4|\mathscr{U}|}\,\varepsilon. \tag{87}$$

Using (81) and (83), we obtain

$$|\mathscr{U}_0| \ge \frac{1}{2}\left(\frac{c_0}{\varepsilon}\right)^d = 2^{4n+3}, \tag{88}$$

and with (83) and (86) it follows that

$$\frac{|\mathscr{U}_0| - |\mathscr{T}|}{|\mathscr{U}|} \ge \frac{|\mathscr{U}_0| - |\mathscr{T}|}{2|\mathscr{U}_0|} \ge \frac{1}{2}.$$

Now (87) and (81) imply

$$e_{2n}^{\mathrm{avg}}(S^{(d)}, \nu, C(Q)) \ge \frac{\varepsilon}{8} \ge c\, 2^{-4n/d}.$$

Since $\operatorname{supp}\nu \subseteq \mathcal{B}_{L_\infty(Q)}$, we apply (72), concluding the proof. $\square$

Combining Theorem 3.4 and Propositions 5.1 and 5.2, we obtain

**Theorem 5.3.** *Let $1 \le p \le \infty$ and $\bar p = \min(p, 2)$. Then there exist constants $c_{1-6} > 0$ such that for all $d, n \in \mathbb{N}$, $0 < \varepsilon \le c_1$,*

$$c_2 \max(n^{-1+1/\bar p}, 2^{-4n/d}) \le e_n^{\mathrm{ran}}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \le c_3 d^{1-1/\bar p} n^{-1+1/\bar p},$$

*moreover, for $p > 1$,*

$$\max\left(\frac{c_4}{\varepsilon^{\bar p/(\bar p-1)}}, \frac{d}{4}\log\left(\frac{c_5}{\varepsilon}\right)\right) \le n_\varepsilon^{\mathrm{ran}}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \le \frac{c_6 d}{\varepsilon^{\bar p/(\bar p-1)}}, \tag{89}$$

*and finally, for $p = 1$,*

$$n_\varepsilon^{\mathrm{ran}}(S^{(d)}, \mathcal{B}_{L_1(Q)}, C(Q)) = \infty.$$

As a consequence, we get the sharp order of the minimal error in $n$ for $d$ fixed.

**Corollary 5.4.** *Let $1 \le p \le \infty$, $\bar{p} = \min(p, 2)$, and $d \in \mathbb{N}$. There are constants $c_1(d), c_2(d) > 0$ such that for all $n \in \mathbb{N}$,*

$$c_1(d) n^{-1+1/\bar{p}} \le e_n^{\mathrm{ran}}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \le c_2(d) n^{-1+1/\bar{p}}.$$

So the algorithms constructed in Sections 3 and 4 are of optimal order (up to logarithmic factors for the Smolyak-Monte Carlo algorithm in the case $1 \le p \le 2$). Furthermore, we obtain for any fixed $0 < \varepsilon \le c_1$ the order of the information complexity (see relations (70) and (71)) as a function of $d$ — it is linear in $d$ for all $p > 1$.

**Corollary 5.5.** *Let $1 < p \le \infty$. Then there is a constant $c_1 > 0$ with the following property. For each $0 < \varepsilon \le c_1$ there exist constants $c_2(\varepsilon), c_3(\varepsilon) > 0$ such that for all $d \in \mathbb{N}$*

$$c_2(\varepsilon) d \le n_\varepsilon^{\mathrm{ran}}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \le c_3(\varepsilon) d.$$

Finally, as observed by an anonymous referee, the lower bound of (89) implies that the upper bound of the same relation is sharp among all estimates of the form

$$n_\varepsilon^{\mathrm{ran}}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \le \frac{c_1 d^{\sigma_1}}{\varepsilon^{\sigma_2}} \quad \text{for all} \quad d \in \mathbb{N},\ 0 < \varepsilon \le c_2, \tag{90}$$

in the sense that if $c_1, c_2 > 0$ and $\sigma_1, \sigma_2 \in \mathbb{R}$ are such that (90) holds, then

$$\sigma_1 \ge 1, \quad \sigma_2 \ge \bar{p}/(\bar{p} - 1). \tag{91}$$

This remark, as well as Corollaries 5.4 and 5.5, remain true with $e_n^{\mathrm{ran}}$ replaced by $e_n^{\mathrm{ran},j}$ and $n_\varepsilon^{\mathrm{ran}}$ by $n_\varepsilon^{\mathrm{ran},j}$ $(j = 1, 2)$, respectively, since the upper bounds were obtained by the help of a linear algorithm (Theorem 3.4).

For the class of nonadaptive algorithms the lower bounds of Proposition 5.2 and Theorem 5.3 can be strengthened as follows.

**Proposition 5.6.** *Let $1 \le p \le \infty$. Then there is a constant $c > 0$ such that for all $d, n \in \mathbb{N}$*

$$e_n^{\mathrm{ran},1}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \ge c \min(d/n, 1).$$

*Proof.* We argue in a way similar to the proof of Proposition 5.2. We use again the set $\mathscr{U}$, see relations (77) and (78), and the distribution $\nu$, see (80). Given $n \in \mathbb{N}$, we put here

$$\varepsilon = \frac{c_0}{2e} \min\left(\frac{d}{n}, 1\right). \tag{92}$$

We estimate $e_n^{\mathrm{avg},1}(S^{(d)}, \nu, C(Q))$ from below. Let $A \in \mathcal{A}_n^{\det,1}(\mathcal{F}(Q), C(Q))$,

$$Af = \varphi_n(f(t_1), \ldots, f(t_n)) \qquad (f \in \mathcal{F}(Q)),$$

and put

$$\mathscr{T} = \left\{(\chi_{[0,u]}(t_i))_{i=1}^n \,:\, u \in \mathscr{U}\right\} \subseteq \{0,1\}^n. \tag{93}$$

It follows that

$$|\{A\chi_{[0,u]} \,:\, u \in \mathscr{U}\}| \le |\mathscr{T}|. \tag{94}$$

Now we use an argument due to Hinrichs ([10], proof of Theorem 4). Since the Vapnik-Červonenkis dimension of the family $\{[0,u] \,:\, u \in \mathscr{U}\}$ is $\le d$ (referring again to [3], Cor. 9.2.15), we conclude from the shatter function lemma that

$$|\mathscr{T}| \le \left(e \max\left(\frac{n}{d}, 1\right)\right)^d \tag{95}$$

(see, e.g., [13], Lemma 5.9 and inequality (4.7), for the case $n \ge d$, the case $n < d$ is trivial). From (94) and (79) we get

$$|\{u \in \mathscr{U} \,:\, \|S^{(d)}\chi_{[0,u]} - A\chi_{[0,u]}\| < \varepsilon/4\}| \le |\mathscr{T}|,$$

hence

$$e_n^{\mathrm{avg},1}(S^{(d)}, \nu, C(Q)) \ge \frac{|\mathscr{U}| - |\mathscr{T}|}{4|\mathscr{U}|}\,\varepsilon. \tag{96}$$

On the other hand, by (77) and (92),

$$|\mathscr{U}| \ge \left(\frac{c_0}{\varepsilon}\right)^d = \left(2e \max\left(\frac{n}{d}, 1\right)\right)^d. \tag{97}$$

Together with (95) we obtain

$$\frac{|\mathscr{U}| - |\mathscr{T}|}{|\mathscr{U}|} \ge 1 - 2^{-d}, \tag{98}$$

consequently, from (96) and (92),

$$e_n^{\mathrm{avg},1}(S^{(d)}, \nu, C(Q)) \ge \frac{\varepsilon}{8} = \frac{c_0}{16e} \min\left(\frac{d}{n}, 1\right),$$

and the desired result follows from (73). $\square$

As a consequence of Theorem 3.4 and Proposition 5.6 we get

**Theorem 5.7.** *Let $1 \le p \le \infty$ and $\bar{p} = \min(p, 2)$. Then there exist constants $c_{1-5} > 0$ such that for all $d, n \in \mathbb{N}$, $0 < \varepsilon \le c_1$, $j = 1, 2$,*

$$c_2 \max(n^{-1+1/\bar{p}}, \min(d/n, 1)) \le e_n^{\mathrm{ran},j}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \le c_3 d^{1-1/\bar{p}} n^{-1+1/\bar{p}},$$

*furthermore, for $p > 1$,*

$$c_4 \max(1/\varepsilon^{\bar{p}/(\bar{p}-1)}, d/\varepsilon) \le n_\varepsilon^{\mathrm{ran},j}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \le c_5 d/\varepsilon^{\bar{p}/(\bar{p}-1)}.$$

*and for $p = 1$,*

$$n_\varepsilon^{\mathrm{ran},j}(S^{(d)}, \mathcal{B}_{L_1(Q)}, C(Q)) = \infty.$$

We do not know if Proposition 5.6 holds for $e_n^{\mathrm{ran}}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q))$. Its proof does not generalize directly to adaptive algorithms. An obvious obstacle is that we cannot apply the shatter function lemma since the point set $(t_i)_{i=1}^n$ may vary with the input $\chi_{[0,u]}$. But more than that, one can show that, in a certain sense, this proof cannot work for adaptive algorithms. Namely, observe that the proof operates on the smaller class

$$F_1 = \{\chi_{[0,u]} : u \in [0,1]^d\}$$

and yields the estimate

$$e_n^{\mathrm{ran},1}(S^{(d)}, F_1, C(Q)) \geq c \min(d/n, 1).$$

However, this estimate does not hold for $e_n^{\mathrm{ran}}(S^{(d)}, F_1, C(Q))$. Indeed, for the class $F_1$ adaptive algorithms can have a much better, an exponential rate, as the following result shows.

**Proposition 5.8.** *For all $d, n \in \mathbb{N}$,*

$$e_n^{\mathrm{ran}}(S^{(d)}, F_1, C(Q)) \leq d\, 2^{-\lfloor n/d \rfloor - 1}. \tag{99}$$

*Proof.* Use bisection to determine an approximation $v = (v_1, \dots, v_d)$ to the input $u = (u_1, \dots, u_n)$ with $n$ queries (in other words, with $n$ function values $\chi_{[0,u]}(t_i)$ with adaptively chosen $t_i$) and precision

$$\max_{1 \leq i \leq d} |u_i - v_i| \leq 2^{-\lfloor n/d \rfloor - 1}.$$

Then approximate

$$S^{(d)} \chi_{[0,u]} \approx S^{(d)} \chi_{[0,v]},$$

where for $x = (x_1, \dots, x_n) \in [0,1]^d$

$$\big(S^{(d)} \chi_{[0,v]}\big)(x) = \int_{[0,x]} \chi_{[0,v]}(t) dt = \prod_{i=1}^d \min(x_i, v_i).$$

Arguing similarly to (16), this leads to (99) (in fact, this is a deterministic approximation). □

Note that the results of this section remain true for the case that $S^{(d)}$ is considered as an operator into $L_\infty([0,1]^d)$.

For $p > 1$ the sharp order of

$$e_n^{\mathrm{ran}}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \quad \text{and} \quad e_n^{\mathrm{ran},j}(S^{(d)}, \mathcal{B}_{L_p(Q)}, C(Q)) \quad (j = 1, 2)$$

as a function of $n$ and $d$ simultaneously is an open problem.

# 6 Supplements, extensions, comments

## 6.1 Deterministic setting

We want to compare our results to the deterministic setting, which was defined in Section 5. The deterministic setting is not well-defined for $F = \mathcal{B}_{L_p(Q)}$, since the elements are classes of functions for which function values are not well-defined. Alternatively, we might consider the dense subset $F = \mathcal{B}_{L_p(Q)} \cap C(Q)$ . Then function valued are defined. However, we have the following essentially well-known result.

**Proposition 6.1.** *For all* $n \in \mathbb{N}$

$$e_n^{\det}(S^{(d)}, \mathcal{B}_{L_p(Q)} \cap C(Q), C(Q)) = 1.$$

*Proof.* The case $p < \infty$ of Proposition 6.1 follows from the case $p = \infty$, which says that

$$e_n^{\det}(S^{(d)}, \mathcal{B}_{C(Q)}, C(Q)) = 1.$$

Using the same argument as in the proof of Proposition 5.1, this is readily reduced to

$$e_n^{\det}(S_1, \mathcal{B}_{C([0,1])}, \mathbb{K}) = 1,$$

which is well-known and easily checked. □

Thus, we see that deterministic algorithms can have no convergence rate at all for the problem $S^{(d)} : \mathcal{B}_{L_p(Q)} \cap C(Q) \to C(Q)$.

## 6.2 Efficient function evaluation for simple sampling

It is interesting to consider the task that once the representation (3), (7), or (34–35) of the output of the respective algorithm has been obtained, we want to compute many function values of it. The case of (34–35) was discussed at the end of Section 4. Here we restrict the consideration to (3). It was mentioned in the discussion after (7) that a direct appoach leads to a cost of *cdn* for each value. In this case it might make sense to spend some extra effort in advance to make the subsequent computations more efficient. This is the topic of the present subsection.

We have the following task: given $n \in \mathbb{N}$ and any

$$z_i \in [0,1]^d, \quad \beta_i \in \mathbb{K} \quad (i = 1, \dots, n), \tag{100}$$

compute

$$s(x) = \sum_{i=1}^{n} \beta_i \chi_{[z_i, \bar{1}]}(x) \tag{101}$$

for a given $x \in [0,1]^d$ (or a number of such $x$). We assume that $n = 2^L$ for some $L \in \mathbb{N}_0$. (If this is not the case we put $L = \lceil \log n \rceil$ and add points $z_i = 0$ and numbers $\beta_i = 0$ for $i = n+1, \dots, 2^L$.)

We need some notation. Let $\mathcal{D}_L$ be the set of all integer intervals of the form

$$I = \{k2^l + 1, k2^l + 2, \dots, (k+1)2^l\} \tag{102}$$

contained in $\{1, \dots, 2^L\}$, i.e., all intervals (102) with $0 \le l \le L$ and $0 \le k < 2^{L-l}$. In a first step we provide the needed arrays of auxiliary numbers, that is, we compute a series of numbers which depend on the $z_i$ and $\beta_i$, which are then used for the subsequent computation of the value $s(x)$. Let us call this structure a $d$-dimensional sampling array of size $n$. It is defined recursively.

A one-dimensional sampling array of size $n$ is a pair of $n$-vectors

$$a = ((u_i)_{i=1}^n, (\gamma_i)_{i=1}^n) \tag{103}$$

with $\gamma_i \in \mathbb{K}$ $(i = 1, \dots, n)$ and

$$0 \le u_1 \le \ldots \le u_n \le 1. \tag{104}$$

For $d > 1$ a $d$-dimensional sampling array of size $n$ is a tuple

$$a = ((u_i)_{i=1}^n, (a_I)_{I \in \mathcal{D}_L}) \tag{105}$$

where $(u_i)_{i=1}^n$ satisfies (104) and the $a_I$ are $(d-1)$-dimensional sampling arrays of size $|I|$. Let $\mathscr{S}(d,n)$ denote the set of all $d$-dimensional sampling arrays of size $n$.

Let $\mathscr{P}_n$ denote the set of all permutations of $(1, 2, \dots, n)$ and let

$$\Pi_L : [0,1]^n \to \mathscr{P}_n$$

be such that for all $(u_i)_{i=1}^n \in [0,1]^n$ the following holds: if $\Pi_L(u_1, \dots, u_n) = \pi$, then

$$u_{\pi(i)} \le u_{\pi(j)} \quad (1 \le i < j \le n)$$

(i.e., $\pi$ induces a non-decreasing reordering of the $(u_i)_{i=1}^n$).

Now we define recursively

$$\Lambda_d(z_1, \dots, z_n, \beta_1, \dots, \beta_n) \in \mathscr{S}(d,n),$$

for all $(z_i)_{i=1}^n, (\beta_i)_{i=1}^n$ satisfying (100). We let

$$\pi = \Pi_L(z_{1,1}, \dots, z_{n,1}).$$

For $d = 1$ define

$$\Lambda_1(z_1, \dots, z_n, \beta_1, \dots, \beta_n) = ((z_{\pi(i)})_{i=1}^n, (\gamma_i)_{i=1}^n), \tag{106}$$

where

$$\gamma_i = \sum_{k=1}^{i} \beta_{\pi(k)} \quad (i = 1, \dots, n).$$

For $d > 1$, we write $z_i = (z_{i,1}, z_i')$ with $z_{i,1} \in [0,1]$ and $z_i' \in [0,1]^{d-1}$. Then we define

$$\Lambda_d(z_1, \dots, z_n, \beta_1, \dots, \beta_n) = \left((z_{\pi(i),1})_{i=1}^n, (a_I)_{I \in \mathcal{D}_L}\right) \tag{107}$$

with

$$a_I = \Lambda_{d-1}\left((z'_{\pi(i)})_{i \in I}, (\beta_{\pi(i)})_{i \in I}\right). \tag{108}$$

Given $x = (x_1, x') \in [0,1] \times [0,1]^{d-1}$ and a $d$-dimensional sampling array $a \in \mathscr{S}(d,n)$, we define the function

$$\Psi_d(x, a) \in \mathbb{K}$$

as follows. Let $a$ have the form (103) if $d = 1$ and the form (105) if $d > 1$. In both cases we determine the largest $j \le n$ with $u_j \le x_1$. If there is no such $j$, we set

$$\Psi_d(x, a) = 0. \tag{109}$$

Otherwise we put for $d = 1$

$$\Psi_1(x, a) = \gamma_j. \tag{110}$$

If $d > 1$, let

$$\{1, 2, \dots, j\} = \bigcup_{l=1}^{m} I_l \tag{111}$$

be the unique representation with $1 \le m \le L$, $I_l \in \mathcal{D}_L$ $(l = 1, \dots, m)$ and

$$I_{l_1} \cap I_{l_2} = \emptyset, \quad |I_{l_1}| > |I_{l_2}| \quad (l_1 < l_2). \tag{112}$$

Then we set

$$\Psi_d(x, a) = \sum_{l=1}^{m} \Psi_{d-1}(x', a_{I_l}). \tag{113}$$

Our first claim is the following

**Lemma 6.2.** *For all $d \in \mathbb{N}$, $L \in \mathbb{N}_0$, $n = 2^L$, $x \in [0,1]^d$, $z_1, \dots, z_n \in [0,1]^d$, $\beta_1, \dots, \beta_n \in \mathbb{K}$ the following holds*

$$s(x) = \Psi_d(x, \Lambda_d(z_1, \dots, z_n, \beta_1, \dots, \beta_n)). \tag{114}$$

*Proof.* We argue by induction over $d$. First let $d = 1$. If

$$\{i : z_i \le x\} = \emptyset,$$

then

$$s(x) = \sum_{i=1}^{n} \beta_i \chi_{[z_i,1]}(x) = 0 = \Psi_1(x, \Lambda_1(z_1, \dots, z_n, \beta_1, \dots, \beta_n))$$

by (109). Otherwise, with

$$j = \max\{i : \, z_{\pi(i)} \le x\},$$

we have

$$\begin{aligned} s(x) &= \sum_{i=1}^{n} \beta_i \chi_{[z_i,1]}(x) = \sum_{i=1}^{n} \beta_{\pi(i)} \chi_{[z_{\pi(i)},1]}(x) \\ &= \sum_{i=1}^{j} \beta_{\pi(i)} = \gamma_j = \Psi_1(x, \Lambda_1(z_1, \dots, z_n, \beta_1, \dots, \beta_n)). \end{aligned}$$

Now let $d > 1$ and assume the statement holds for $d-1$. Again, we first consider the case

$$\{i : \, z_{i,1} \le x_1\} = \emptyset. \tag{115}$$

Here we have

$$\begin{aligned} s(x) &= \sum_{i=1}^{n} \beta_i \chi_{[z_i,\bar{1}]}(x) = \sum_{i=1}^{n} \beta_i \chi_{[z_{i,1},1]}(x_1) \chi_{[z_i',1']}(x') \\ &= 0 = \Psi_d(x, \Lambda_d(z_1, \dots, z_n, \beta_1, \dots, \beta_n)), \end{aligned}$$

by (109). If (115) does not hold, we set

$$j = \max\{i : \, z_{\pi(i),1} \le x_1\}$$

and conclude

$$\begin{aligned} s(x) &= \sum_{i=1}^{n} \beta_i \chi_{[z_i,\bar{1}]}(x) = \sum_{i=1}^{n} \beta_{\pi(i)} \chi_{[z_{\pi(i)},\bar{1}]}(x) \\ &= \sum_{i=1}^{n} \beta_{\pi(i)} \chi_{[z_{\pi(i),1},1]}(x_1) \chi_{[z'_{\pi(i)},1']}(x') \\ &= \sum_{i=1}^{j} \beta_{\pi(i)} \chi_{[z'_{\pi(i)},1']}(x') = \sum_{l=1}^{m} \sum_{i \in I_l} \beta_{\pi(i)} \chi_{[z'_{\pi(i)},1']}(x') \\ &= \sum_{l=1}^{m} \Psi_{d-1}(x', \Lambda_{d-1} \left( (z'_{\pi(i)})_{i \in I_l}, (\beta_{\pi(i)})_{i \in I_l} \right) \\ &= \Psi_d(x, \Lambda_d(z_1, \dots, z_n, \beta_1, \dots, \beta_n)), \end{aligned}$$

by the induction hypothesis, (107), (108), and (113). □

Now we have a look at the number of arithmetic operations needed for the computation of $s(x)$ according to formula (114). Recall that we assume the real number model [15].

**Lemma 6.3.** *There is a choice of* $(\Pi_L)_{L\in\mathbb{N}_0}$ *such that the following holds: For all* $d \in \mathbb{N}$ *there is a constant* $c(d) > 0$ *such that for all* $L \in \mathbb{N}_0$, $n = 2^L$, $z_1, \ldots, z_n \in [0,1]^d$, $\beta_1, \ldots, \beta_n \in \mathbb{K}$ *the d-dimensional sampling array*

$$\Lambda_d(z_1, \ldots, z_n, \beta_1, \ldots, \beta_n),$$

*as defined in (106), (107), and (108), can be computed in*

$$\leq c(d)n(\log n)^{\max(d-1,1)}$$

*operations.*

*Proof.* Let $\Pi_L(u_1, \ldots, u_n)$ be the output supplied by merge sorting, which can be obtained in $\leq cn \log n$ operations (see [1], Ch. 2.7). For $d = 1$ we need a total of $\leq c_1 n \log n$ operations for sorting the $z_i$ and computing the sums. If $d = 2$, we first sort $z_{i,1}$ in $\leq cn \log n$, which gives $\pi$. Then for each $I \in \mathcal{D}_L$ we have to sort $z'_{\pi(i)} = z_{\pi(i),2}$ for $i \in I$. These results are obtained simultaneously for all $I$ as a by-product of merge sorting $(z_{\pi(i),2})_{i=1}^n$, which requires $\leq cn \log n$ operations. The remaining computations of the sums require

$$\leq c \sum_{I\in\mathcal{D}_L} |I| \leq cn \log n$$

operations.

For $d > 2$ we argue by induction. So assume the statement holds for $d - 1$. To compute

$$\Lambda_d(z_1, \ldots, z_n, \beta_1, \ldots, \beta_n)$$

according to (107) and (108), we need $cn \log n$ operations for sorting the first component. By the induction assumption, the computation of the $a_I$ requires not more than

$$\begin{aligned} c(d-1) \sum_{I\in\mathcal{D}_L} |I|(\log |I|)^{d-2} &\leq c(d-1)2^L \sum_{l=0}^{L} (L-l)^{d-2} \\ &\leq c(d)2^L L^{d-1} \leq c(d)n(\log n)^{d-1}. \end{aligned}$$

□

**Lemma 6.4.** *Let* $d \in \mathbb{N}$. *Then there is a constant* $c(d) > 0$ *such that for all* $L \in \mathbb{N}_0$, $n = 2^L$, $a \in \mathscr{S}(d,n)$, $x \in [0,1]^d$ *the function* $\Psi_d(x,a)$ *given by (109), (110), and (113), can be computed in* $\leq c(d)(\log n + 1)^d$ *operations.*

*Proof.* For $d = 1$ we apply the bisection algorithm to determine $j$ (or its non-existence) in $\leq c(\log n + 1)$ operations. For $d > 1$ we argue by induction. Assume the statement is true for $d - 1$. Again, we determine $j$ by bisection. The binary representation of $j$ yields $m \leq L$ and the sets $(I_l)_{l=1}^m$ so that (111) and (112) hold. By the induction assumption, the cost of computing $\Psi_{d-1}(x', a_{I_l})$ is $\leq c(d-1)(\log |I_l| + 1)^{d-1}$, so the total cost is

$$\begin{aligned} &\leq \quad c(\log n + 1) + c(d-1) \sum_{l=1}^{m} (\log |I_l| + 1)^{d-1} \\ &\leq \quad c(\log n + 1) + c(d-1) \sum_{l=1}^{L} (l+1)^{d-1} \\ &\leq \quad c(d)(\log n + 1)^d. \end{aligned}$$

□

**Corollary 6.5.** *Let $d \in \mathbb{N}$. Then there are constants $c_1(d), c_2(d) > 0$ such that for all $n \in \mathbb{N}$, $n \geq 2$, $z_1, \dots, z_n \in [0,1]^d$, $\beta_1, \dots, \beta_n \in \mathbb{K}$, $N \in \mathbb{N}$, $x_i \in [0,1]^d$ $(i = 1, \dots, N)$ the values*

$$s(x_i) = \sum_{i=1}^{n} \beta_i \chi_{[z_i, \bar{1}]}(x_i) \quad (i = 1, \dots, N)$$

*can be computed in*

$$\leq c_1(d) n (\log n)^{\max(d-1,1)} + c_2(d)(\log n)^d N$$

*operations.*

Let us summarize the total cost – including the computation of $N$ values of the output function – needed for algorithm $A_n^1$ to reach an error $\varepsilon > 0$. Combining Lemmas 6.2–6.4 with Theorem 3.4 we obtain the following

**Corollary 6.6.** *Let $d \in \mathbb{N}$, $1 < p \leq \infty$, $1 \leq p_1 < \infty$, $p_1 \leq p$, and $\bar{p} = \min(p, 2)$. Then there are constants $c_1, c_2 > 0$ not depending on $d$ and constants $c_3(d), c_4(d) > 0$ such that for each $0 < \varepsilon \leq 1/2$ there exists an $n \in \mathbb{N}$ with the following properties. The algorithm $A_n^1$ has error*

$$\sup_{f \in \mathcal{B}_{L_p(Q)}} \left( \mathbb{E} \sup_{x \in Q} \left| \left(S^{(d)} f\right)(x) - \left(A_n^1 f\right)(x) \right|^{p_1} \right)^{1/p_1} \leq \varepsilon$$

*and for each $f \in L_p(Q)$ and $\omega \in \Omega$ it uses not more than*

$$n \leq c_1 d \left( \frac{1}{\varepsilon} \right)^{\bar{p}/(\bar{p}-1)}$$

*function values of $f$ and needs*

$$\leq c_2 d^2 \left( \frac{1}{\varepsilon} \right)^{\bar{p}/(\bar{p}-1)}$$

*operations to set up the approximating function* $A^1_{n,\omega}f$.

*Moreover, having obtained* $A^1_n f$*, a* $d$*-dimensional sampling array of size* $n$ *can be computed in*

$$\le c_3(d)\left(\frac{1}{\varepsilon}\right)^{\bar p/(\bar p-1)}\left(\log\left(\frac{1}{\varepsilon}\right)\right)^{\max(d-1,1)}$$

*operations, with the property that for each* $N\in\mathbb{N}$, $x_i\in Q$ $(i=1,\dots,N)$*, the values* $\big(A^1_n f\big)(x_i)$ $(i=1,\dots,N)$ *can be computed in*

$$c_4(d)\left(\log\left(\frac{1}{\varepsilon}\right)\right)^{d} N$$

*operations.*

For comparison, let us formulate the analogous result for the Smolyak-Monte Carlo algorithm, which is a consequence of Theorem 4.5 and the cost analysis given after its proof.

**Corollary 6.7.** *Let* $d\in\mathbb{N}$, $1<p\le\infty$, $1\le p_1<\infty$, $p_1\le p$*, and* $\bar p=\min(p,2)$*. Then there are constants* $c_{1-3}(d)>0$ *such that for each* $0<\varepsilon\le 1/2$ *there exist* $m,L\in\mathbb{N}_0$, $m\ge 2$ *with the following properties. The algorithm* $A^3_{m,L}$ *has error*

$$\sup_{f\in\mathcal{B}_{L_p(Q)}}\left(\mathbb{E}\sup_{x\in Q}\left|\big(S^{(d)}f\big)(x)-\big(A^3_{m,L}f\big)(x)\right|^{p_1}\right)^{1/p_1}\le\varepsilon$$

*and for each* $f\in L_p(Q)$ *and* $\omega\in\Omega$ *it needs not more than*

$$\begin{aligned} & c_1(d)(L+1)^{d-1}m^L \\ \le\ & c_2(d)\begin{cases}\left(\frac{1}{\varepsilon}\right)^{\bar p/(\bar p-1)} & \text{if}\quad 2<p\le\infty\\ \left(\frac{1}{\varepsilon}\right)^{\bar p/(\bar p-1)}\left(\log\left(\frac{1}{\varepsilon}\right)\right)^{\left(1+\frac{p}{p-1}\right)(d-1)} & \text{if}\quad 1<p\le 2\end{cases}\end{aligned}$$

*function values of* $f$ *and (up to a constant factor depending on* $d$*) the same number of operations to set up the approximating function* $A^3_{m,L}f$.

*Furthermore, having obtained* $A^3_{m,L}f$*, for each* $N\in\mathbb{N}$, $x_i\in Q$ $(i=1,\dots,N)$*, the values* $\big(A^3_{m,L}f\big)(x_i)$ $(i=1,\dots,N)$ *can be computed in*

$$\le c_3(d)\begin{cases} N & \text{if}\quad 2<p\le\infty\\ \left(\log\left(\frac{1}{\varepsilon}\right)\right)^{d-1}N & \text{if}\quad 1<p\le 2\end{cases}$$

*operations.*

On the basis of these results let us compare the total cost of both algorithms, including the computation of $N$ values of the approximating function. We assume $d$ to be fixed and consider the behaviour as $\varepsilon \to 0$, $N \to \infty$. For $2 < p \leq \infty$ and, if $d = 1$, also for $1 < p \leq 2$, the cost of the Smolyak-Monte Carlo algorithm is (up to a constant factor depending on $d$) lower than that of the simple sampling algorithm. On the other hand, for $1 < p \leq 2$ and $d > 1$ the cost of the Smolyak-Monte Carlo algorithm can be anything from higher ($N$ small relative to $1/\varepsilon$) to slightly lower than simple sampling ($N$ large relative to $1/\varepsilon$).

If $d$ is large, the simple sampling algorithm with direct term-by-term computation of cost $\leq cdnN$ is obviously preferable to the version with sampling array computation and also to the Smolyak-Monte Carlo algorithm because of the exponential dependence of the cost on $d$ in the latter two.

## 6.3 Separability and measurability

In Section 5 we mentioned that the simple sampling algorithm

$$A_n^1 \notin \mathcal{A}^{\mathrm{ran}}(\mathcal{B}_{L_p(Q)}, B_0(Q)). \tag{116}$$

We show that it does not have property 1 introduced in Section 5 (see below (66)). Let $f_0(x) \equiv 1$ ($x \in [0,1]^d$). Then by (3)

$$A_{n,\omega}^1 f_0 = \frac{1}{n} \sum_{i=1}^{n} \chi_{[\xi_i(\omega), \bar{1}]} \quad (\omega \in \Omega).$$

Define

$$\begin{aligned} Q_i &= \left[\frac{i-1}{n}, \frac{i}{n}\right)^d \subset [0,1]^d = Q \quad (i = 1, \dots, n), \\ K &= \prod_{i=1}^{n} Q_i \subset Q^n. \end{aligned}$$

For $(x_1, \dots, x_n), (y_1, \dots, y_n) \in K$ we have

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \chi_{[x_i, \bar{1}]} - \frac{1}{n} \sum_{i=1}^{n} \chi_{[y_i, \bar{1}]} \right\|_{B_0(Q)} \geq \frac{1}{n} \tag{117}$$

whenever $(x_1, \dots, x_n) \neq (y_1, \dots, y_n)$. Put

$$\Omega_0 = \{\omega \in \Omega : \, (\xi_1(\omega), \dots, \xi_n(\omega)) \in K\}.$$

Clearly, $\mathbb{P}(\Omega_0) \neq 0$. Moreover, if $X$ is a separable subspace of $B_0(Q)$, then due to (117), the set

$$K_X = \left\{ (x_1, \dots, x_n) \in K : \, \frac{1}{n} \sum_{i=1}^{n} \chi_{[x_i, \bar{1}]} \in X \right\}$$

is at most countable, which implies

$$\begin{aligned}&\mathbb{P}(\{\omega \in \Omega_0 :\ A^1_{n,\omega} f_0 \in X\})\\ &\quad = \quad \mathbb{P}(\{\omega \in \Omega_0 :\ (\xi_1(\omega), \dots, \xi_n(\omega)) \in K_X\}) = 0 \neq \mathbb{P}(\Omega_0).\end{aligned}$$

Hence, the mapping $\Phi : \Omega \to B_0(Q)$ given by

$$\Phi(\omega) = A^1_{n,\omega} f_0 \quad (\omega \in \Omega)$$

is not essentially separably valued, and (116) follows.

Let us also mention that if we consider the canonical choice $\Omega = [0,1]^{nd}$, $\Sigma$ the $\sigma$-algebra of all Lebesgue measurable subsets, $\mathbb{P}$ the Lebesgue measure on $[0,1]^{nd}$ and, for $1 \le i \le n$,

$$\xi_i(\omega) = x_i \quad (\omega = (x_1, \dots, x_n) \in Q^n = \Omega),$$

then $\Phi$ is not $\Sigma$-to-Borel measurable. To see this, assume the contrary. We have $\Omega_0 = K$ and for $\omega = (x_1, \dots, x_n) \in Q^n$

$$\Phi(\omega) = \frac{1}{n} \sum_{i=1}^{n} \chi_{[x_i, \bar{1}]}.$$

Moreover, by (117), $\Phi$ is a one-to-one mapping of $K$ onto

$$Z = \left\{ \frac{1}{n} \sum_{i=1}^{n} \chi_{[x_i,\bar{1}]} :\ (x_1, \dots, x_n) \in K \right\} \subset B_0(Q).$$

Also by (117), each subset of $Z$ is a closed subset of $B_0(Q)$, hence Borel measurable, implying that each subset of $K$ is Lebesgue measurable, a contradiction.

The arguments above remain true when considering $A^1_{n,\omega}$ as a mapping into $L_\infty(Q)$.

# References


[1] A. Aho, J. Hopcroft, J. Ullman, The Desing and Analysis of Computer Algorithms, Addison-Wesley, Reading, Massachusetts, 1974.

[2] A. Defant, K. Floret, Tensor Norms and Operator Ideals, North Holland, Amsterdam, 1993.

[3] R. M. Dudley, A course on empirical processes (École d'Été de Probabilités de Saint-Flour XII-1982). Lecture Notes in Mathematics 1097, 2–141, Springer-Verlag, New York, 1984.

[4] S. Heinrich, Random approximation in numerical analysis, in: K. D. Bierstedt, A. Pietsch, W. M. Ruess, D. Vogt (Eds.), Functional Analysis, Marcel Dekker, New York, 1993, 123–171.

[5] S. Heinrich, Monte Carlo approximation of weakly singular integral operators, J. Complexity 22 (2006), 192–219.

[6] S. Heinrich, The randomized information complexity of elliptic PDE, J. Complexity 22 (2006), 220–249.

[7] S. Heinrich, Randomized approximation of Sobolev embeddings II, J. Complexity 25 (2009), 455–472.

[8] S. Heinrich, Randomized approximation of Sobolev embeddings III, J. Complexity 25 (2009), 473–507.

[9] S. Heinrich, E. Novak, G. W. Wasilkowski, H. Woźniakowski, The inverse of the star-discrepancy depends linearly on the dimension, Acta Arithmetica 96 (2001), 279-302.

[10] A. Hinrichs, Covering numbers, Vapnik-Červonenkis classes and bounds for the star-discrepancy, J. Complexity 20 (2004), 477–483.

[11] M. Ledoux, M. Talagrand, Probability in Banach Spaces, Springer, Berlin–Heidelberg–New York, 1991.

[12] W. A. Light, W. Cheney, Approximation Theory in Tensor Product Spaces, Lecture Notes in Mathematics 1169, Springer-Verlag, Berlin, 1985.

[13] J. Matoušek, Geometric Discrepancy. An Illustrated Guide, Springer, Berlin, 1999.

[14] E. Novak, Deterministic and Stochastic Error Bounds in Numerical Analysis, Lecture Notes in Mathematics 1349, Springer-Verlag, Berlin, 1988.

[15] E. Novak, The real number model in numerical analysis, J. Complexity 11 (1995), 57–73.

[16] E. Novak, H. Triebel, Function spaces in Lipschitz domains and optimal rates of convergence for sampling, Constr. Approx. 23 (2006), 325–350.

[17] E. Novak, H. Woźniakowski, Tractability of Multivariate Problems, Volume 1, Linear Information, European Math. Soc., Zürich, 2008.

[18] E. Novak, H. Woźniakowski, Tractability of Multivariate Problems, Volume 2, Standard Information for Linear Functionals (in preparation).

[19] G. Pisier, Remarques sur les classes de Vapnik-Červonenkis, Ann. Inst. Henri Poincaré, Probab. Stat. 20 (1984), 287–298.

[20] W. Sickel, T. Ullrich, Tensor products of Sobolev-Besov spaces and applications to approximation from the hyperbolic cross, J. Approx. Theory 161 (2009), 748–786.

[21] J. F. Traub, G. W. Wasilkowski, H. Woźniakowski, Information-Based Complexity, Academic Press, 1988.