Optimal confidence for Monte Carlo integration of smooth functions

Robert J. Kunsch[*] and Daniel Rudolf[†]

September 27, 2018

Abstract

We study the complexity of approximating integrals of smooth functions at absolute precision $\varepsilon > 0$ with confidence level $1 - \delta \in (0, 1)$. The optimal error rate for multivariate functions from classical isotropic Sobolev spaces $W_p^r(G)$ with sufficient smoothness on bounded Lipschitz domains $G \subset \mathbb{R}^d$ is determined. It turns out that the integrability index p has an effect on the influence of the uncertainty δ in the complexity. In the limiting case $p = 1$ we see that deterministic methods cannot be improved by randomization. In general, higher smoothness reduces the additional effort for diminishing the uncertainty. Finally, we add a discussion about this problem for function spaces with mixed smoothness.

Keywords. Monte Carlo integration; Sobolev functions; information-based complexity; standard information; asymptotic error; confidence intervals.

1 Introduction

We want to compute the integral

$$\mathrm{INT}(f) = \int_G f(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{1}$$

of $f\colon G \to \mathbb{R}$ from the unit ball $\mathcal{B}_{\mathcal{W}}$ of a (semi-)normed linear space $\mathcal{W}$ of functions defined on a domain $G \subset \mathbb{R}^d$ where we are only allowed to use function

[*]Institut für Mathematik, Universität Osnabrück, Albrechtstr. 28a, 49076 Osnabrück, Germany, Email: robert.kunsch@uni-osnabrueck.de

[†]Institute for Mathematical Stochastics, Universität Göttingen & Felix-Bernstein-Institute for Mathematical Statistics, Goldschmidtstraße 7, 37077 Göttingen, Germany, Email: daniel.rudolf@uni-goettingen.de

values as information within randomized algorithms. The focus lies on the (ε,δ)-*complexity* $n^{\mathrm{MC}}_{\mathrm{prob}}(\varepsilon,\delta,\mathcal{W})$, that is, the minimal number n of function values needed for randomized algorithms A_n in order to approximate the integral (1) such that

$$\mathbb{P}\{|A_n(f) - \mathrm{INT}\, f| > \varepsilon\} \leq \delta \qquad \text{for all } f \text{ with } \|f\|_{\mathcal{W}} \leq 1, \tag{2}$$

where $\|\cdot\|_{\mathcal{W}}$ is the (semi-)norm of $\mathcal{W}$. A method with this property of guaranteeing a small (absolute) *error* $\varepsilon > 0$ with *confidence* $1-\delta \in (0,1)$ (or *uncertainty* δ) for inputs from the unit ball $\mathcal{B}_{\mathcal{W}}$ is called (ε,δ)-*approximating in* $\mathcal{W}$, see also [16]. We also consider the *n-th minimal probabilistic Monte Carlo error at uncertainty* δ, defined by

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n,\delta,\mathcal{W}) := \inf\{\varepsilon > 0 \mid \exists\ (\varepsilon,\delta)\text{-approximating algorithm } A_n \text{ in } \mathcal{W}\}\ . \tag{3}$$

The probabilistic error criterion from above is less common in *information-based complexity* (IBC) where the standard notion of *Monte Carlo error* is some type of mean error. In general, for some $\ell \geq 1$ the *n-th minimal ℓ-mean Monte Carlo error* is given by

$$e^{\mathrm{MC}}_{\ell\text{-mean}}(n,\mathcal{W}) := \inf_{A_n} \sup_{\|f\|_{\mathcal{W}}\leq 1} \left(\mathbb{E}\,|A_n(f) - \mathrm{INT}\, f|^{\ell}\right)^{1/\ell}\ . \tag{4}$$

Here, the infimum is taken over all randomized algorithms which use at most n function values. Most frequently studied are the *root mean squared error* (RMSE), that is, $e^{\mathrm{MC}}_{2\text{-mean}}$, as well as the expected error $e^{\mathrm{MC}}_{1\text{-mean}}$. Accordingly, we define the ε-*complexity* $n^{\mathrm{MC}}_{\ell\text{-mean}}(\varepsilon,\mathcal{W})$ as the minimal number of function values needed by a randomized algorithm that guarantees a worst case ℓ-mean error smaller than ε. For more details on IBC we refer to the books [22, 23, 24, 26].

One might argue that the error criterion does not matter. However, this is not true. For example, using Markov's inequality it is always possible to construct (ε,δ)-approximating algorithms once we know methods which work for arbitrarily small mean errors. That way, however, the cost estimates are not optimal in terms of the δ-dependence, namely polynomial rather than logarithmic. In some situations this can be fixed by using more advanced inequalities such as Hoeffding bounds. In other situations commonly known algorithms may need to be modified which lead to more robust methods less prone to outliers. For this reason the probabilistic criterion is frequently used in statistics, see for example [8, 11, 12, 13]. Furthermore, there are numerical problems which can be solved with respect to the probabilistic (ε,δ)-criterion but the mean error is unbounded, see [16]. In other words, this criterion seems to be the right one for the concept of solvability.

In Section 2 we provide two generic lower bounds for the n-th minimal probabilistic Monte Carlo error based on bump functions. In Section 3 we discuss several approaches for deriving upper error bounds on Sobolev classes and discuss in which cases they lead to optimal rates. We mainly consider classical isotropic Sobolev spaces $W^r_p(G)$ on domains $G \subseteq \mathbb{R}^d$. For integer smoothness $r \in \mathbb{N}_0$ and integrability parameter $1 \leq p \leq \infty$, these spaces are given by

$$W_p^r(G) := \left\{ f \in L_p(G) \,\middle|\, \|f\|_{W_p^r(G)} := \Bigg(\sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}_0^d \\ |\boldsymbol{\alpha}|_1 \leq r}} \|D^{\boldsymbol{\alpha}} f\|_{L_p(G)}^p \Bigg)^{1/p} < \infty \right\},$$

with the usual modification for $p = \infty$ and the weak derivative $D^{\boldsymbol{\alpha}} f = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d} f$ for multi-index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$. Note that for $r = 0$ we obtain the Lebesgue spaces $L_p(G)$.

Our main result is for spaces $W_p^r(G)$ on bounded Lipschitz domains $G \subset \mathbb{R}^d$ (see [21] for a definition), and with sufficient smoothness, $rp > d$. In asymptotic notation (see definitions below) it states

$$e_{\mathrm{prob}}^{\mathrm{MC}}(n, \delta, W_p^r(G)) \asymp n^{-r/d} \min \left\{ 1, \left(\frac{\log \delta^{-1}}{n} \right)^{1-1/q} \right\} \tag{5}$$

with $q := \min\{p, 2\}$, or equivalently

$$n_{\mathrm{prob}}^{\mathrm{MC}}(\varepsilon, \delta, W_p^r(G)) \asymp \min \left\{ \varepsilon^{-d/r},\, \varepsilon^{-1/\left(\frac{r}{d} + \frac{q-1}{q}\right)} \left(\log \delta^{-1}\right)^{1/\left(\frac{q}{q-1} \cdot \frac{r}{d} + 1\right)} \right\}, \tag{6}$$

see Theorem 2.3 and Theorem 3.6. The condition $rp > d$ guarantees that the space $W_p^r(G)$ is compactly embedded in the space of continuous functions, see for instance [1]. Only then function evaluations are well defined and there exist deterministic integration methods, which in this case provide error bounds with rate $n^{-r/d}$. These worst case bounds come into play if we demand extremely high confidence $1 - \delta$ close to 1. It also turns out that for $p = 1$ the uncertainty δ does not play any role, which shows that deterministic methods are optimal in that case. In the power of n, we recover the well known gain of $1 - 1/p$ for $1 < p < 2$, and $1/2$ for $p \geq 2$, which Monte Carlo methods achieve compared to deterministic methods. The influence of the uncertainty δ grows with the gain in the error rate. In terms of the complexity (6) we observe that the higher the smoothness r the weaker the dependence on δ.

Asymptotic notation: For functions $e, f \colon \mathbb{N} \times (0,1) \to \mathbb{R}$ we use the notation $e(n, \delta) \preceq f(n, \delta)$, meaning that there is some $n_0 \in \mathbb{N}$ and $\delta_0 \in (0,1)$ such that $e(n, \delta) \leq c f(n, \delta)$ with some (possibly (d, r)-dependent) constant $c > 0$ for all $n \geq n_0$ and $\delta \in (0, \delta_0)$. Sometimes we add the restriction $n \succeq \log \delta^{-1}$, then $e(n, \delta) \leq c f(n, \delta)$ is only meant to hold for $\delta \in (0, \delta_0)$ and $n \geq n_0 \log \delta^{-1}$. Similarly we denote asymptotics for complexity functions $n(\varepsilon, \delta)$, describing a behaviour for small $\varepsilon, \delta > 0$. Asymptotic equivalence $e(n, \delta) \asymp f(n, \delta)$ is a shorthand for $e(n, \delta) \preceq f(n, \delta) \preceq e(n, \delta)$. The notion $e(n, \delta) \prec f(n, \delta)$ means "$e(n, \delta) \preceq f(n, \delta)$ but not $f(n, \delta) \preceq e(n, \delta)$".

2 Lower bounds

We start with the lower bounds as these are easily obtained for the whole parameter range of the function spaces we consider.

2.1 Auxiliary lemmas

As before, let $\mathcal{W}$ be a space of functions defined on a domain G, equipped with a (semi-)norm $\|\cdot\|_{\mathcal{W}}$. An abstract Monte Carlo algorithm defined for such functions is a family $A_n = (A_n^\omega)_{\omega\in\Omega}$ of mappings $A_n^\omega\colon \mathcal{W} \xrightarrow{N^\omega} \mathbb{R}^n \xrightarrow{\phi^\omega} \mathbb{R}$, indexed with elements ω from a probability space $(\Omega, \Sigma, \mathbb{P})$, such that the error functional $\omega \mapsto |A_n^\omega(f) - \mathrm{INT}\, f|$ is measurable. Here,

$$\mathbf{y} = N^\omega(f) = (f(\mathbf{x}_1^\omega), \ldots, f(\mathbf{x}_n^\omega))$$

is the information we collect about a problem instance f, from which the output $A_n^\omega(f) = \phi^\omega(\mathbf{y})$ is generated. One might consider adaptive strategies to acquire information, that is, $\mathbf{x}_i^\omega$ might depend on the previously obtained information $y_1, \ldots, y_{i-1}$. Our lower bounds do hold for this type of algorithms, but the upper bounds we present are based on non-adaptive methods. For simplicity, in this paper we restrict to methods with fixed cardinality n. In general, the number of function values an algorithm collects might be random and even depend on the input, see for instance [8, 12, 16]. Let us mention here that our auxiliary lemmas on lower bounds, Lemma 2.1 and 2.2, would then still hold with slightly worse constants.

In the spirit of Bakhvalov [4], for proving lower bounds we switch to an *average input setting* with a discrete probability measure μ supported within the input set—which in our case is the unit ball $\mathcal{B}_{\mathcal{W}}$ of the space $\mathcal{W}$—and make use of the relation

$$\begin{aligned}
\sup_{\|f\|_{\mathcal{W}}\leq 1} \mathbb{P}\{|A_n(f) - \mathrm{INT}\, f| > \varepsilon\} &\geq \int_{\mathcal{B}_{\mathcal{W}}} \int_{\Omega} \mathbb{1}_{\{|A_n^\omega(f) - \mathrm{INT}\, f| > \varepsilon\}} \,\mathrm{d}\mathbb{P}(\omega)\, \mathrm{d}\mu(f) \\
&= \int_{\Omega} \int_{\mathcal{B}_{\mathcal{W}}} \mathbb{1}_{\{|A_n^\omega(f) - \mathrm{INT}\, f| > \varepsilon\}} \,\mathrm{d}\mu(f)\, \mathrm{d}\mathbb{P}(\omega) \\
&\geq \inf_{Q_n} \mu\{f\colon |Q_n(f) - \mathrm{INT}\, f| > \varepsilon\}\,, \qquad (7)
\end{aligned}$$

where the infimum is taken over all *deterministic* integration methods Q_n that use n function values. (For fixed ω, the realisation A_n^ω of a given algorithm can be regarded as a deterministic algorithm.) In the proof of the lower bounds we use the implication

$$\sup_{\|f\|_{\mathcal{W}}\leq 1} \mathbb{P}\{|A_n(f) - \mathrm{INT}(f)| > \varepsilon\} > \delta \quad \Longrightarrow \quad e^{\mathrm{MC}}_{\mathrm{prob}}(A_n, \delta, \mathcal{W}) \geq \varepsilon\,, \qquad (8)$$

where $e^{\mathrm{MC}}_{\mathrm{prob}}(A_n, \delta, \mathcal{W})$ is the infimum of all $\varepsilon > 0$ such that the algorithm A_n is (ε, δ)-approximating in $\mathcal{W}$.

Depending on the integrability index p of the Sobolev classes we choose different probability measures μ in order to obtain appropriate lower bounds. Similarly to [20, Proposition 1 and 2 in Section 2.2.4] we have the following two generic lemmas, now for the probabilistic instead of the root mean squared error. The first one applies for integrability $2 \leq p \leq \infty$.

Lemma 2.1. *For $n \geq 17$ and a natural number $N \geq 5n + 6$, assume that there are functions $f_i \colon G \to \mathbb{R}$, with $i = 1, \ldots, N$, satisfying the following conditions:*

1. *for $i = 1, \ldots, N$, the sets $G_i := \{\mathbf{x} \in G \colon f_i(\mathbf{x}) \neq 0\}$ are pairwise disjoint, and $\mathrm{INT}(f_i) = \gamma$ for some $\gamma > 0$;*
2. *for signs $s_i \in \{\pm 1\}$, the function $f_{\mathbf{s}} := \sum_{i=1}^{N} s_i f_i$ is an element of the input set $\mathcal{B}_{\mathcal{W}}$, that is, $\|f_{\mathbf{s}}\|_{\mathcal{W}} \leq 1$.*

Then, for any uncertainty level $0 < \delta < 1/3$, we have

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, \mathcal{W}) \geq \gamma \min \left\{ n^{1/2} \sqrt{\log_4 \frac{1}{3\delta}},\, n \right\} .$$

Proof. Let μ be the uniform distribution on the finite set

$$\mathcal{F} := \left\{ f_{\mathbf{s}} = \sum_{i=1}^{N} s_i f_i \colon s_i \in \{\pm 1\} \right\} \subset \mathcal{B}_{\mathcal{W}} .$$

Let $Q_n \colon \mathcal{F} \to \mathbb{R}$ be a deterministic algorithm using n function values. Without loss of generality, we may assume that the algorithm computes function values $y_i = f(\mathbf{x}_i)$ with $\mathbf{x}_i \in G_i$ for $i = 1, \ldots, n$. Hence, from the i-th piece of information we learn whether $s_i = +1$ or -1 for $f = f_{\mathbf{s}}$.

Note that, given the information $\mathbf{y} = N(f)$, there are still $k := N - n$ unknown signs s_i. The conditional distribution of $\mathrm{INT}(f)$ given $\mathbf{y}$ can be represented as the distribution of

$$g_{\mathbf{y}} + \gamma X_k$$

where

$$g_{\mathbf{y}} := \gamma \sum_{i=1}^{n} s_i \,, \qquad \text{and} \qquad X_k := \sum_{i=1}^{k} Z_i \qquad \text{with } Z_i \overset{\text{iid}}{\sim} \text{Rademacher}.$$

Since this is the situation for all information $\mathbf{y} \mathrel{\widehat{=}} (s_1, \ldots, s_n)$, we obtain

$$\begin{aligned} \mu\{|Q_n(f) - \mathrm{INT}\, f| > \varepsilon\} \;&\geq\; \inf_{a \in \mathbb{R}} \mathbb{P}\{\gamma\, |X_k - a| > \varepsilon\} \\ &= \inf_{a \in \mathbb{R}} 2^{-k} \sum_{j=0}^{k} \binom{k}{j} \mathbb{1}\{\gamma\, |2j - k - a| > \varepsilon\} . \end{aligned}$$

At most $k' := \lfloor \varepsilon/\gamma \rfloor + 1$ terms are removed from the binomial sum, optimally the central ones, so we have

$$\mu\{|Q_n(f) - \mathrm{INT}\, f| > \varepsilon\} \;\geq\; 2^{-k} \left[\sum_{j=0}^{\lfloor \frac{k-k'}{2} \rfloor} \binom{k}{j} + \sum_{j=\lceil \frac{k+k'+1}{2} \rceil}^{k} \binom{k}{j} \right].$$

We employ Lemma A.1 twice, namely, for odd k with

$$t = \lceil (k'-1)/2 \rceil \quad \text{and} \quad t = \lceil k'/2 \rceil \leq \varepsilon/(2\gamma) + 1$$

as well as for even k with

$$t = \lceil k'/2 \rceil \quad \text{and} \quad t = \lceil (k'+1)/2 \rceil \leq \varepsilon/(2\gamma) + 3/2.$$

In order to match the conditions of Lemma A.1, we restrict to $\varepsilon/\gamma \leq (k-6)/4$. Hence under that assumption we have

$$\mu\{|Q_n(f) - \mathrm{INT}\, f| > \varepsilon\} \geq \frac{1}{1 + 2/\sqrt{\pi}} \exp\left(-\frac{4\,(\log 2)\,(\varepsilon/\gamma + 2)^2}{k} \right). \tag{9}$$

Note that $k = N - n \geq (5n+6) - n = 4n + 6$, so then the condition $\varepsilon/\gamma \leq n$ is sufficient for (9) to hold. The right-hand side of (9) can be further simplified via $(\varepsilon/\gamma + 2)^2 \leq 2\,(\varepsilon^2/\gamma^2 + 4)$, exploiting $k > 4n$ and also $n \geq 17$. For $0 < \varepsilon \leq \gamma n$ this leads to

$$\mu\{|Q_n(f) - \mathrm{INT}\, f| > \varepsilon\} > \frac{2^{-8/n}}{1 + 2/\sqrt{\pi}}\, 2^{-2\,\varepsilon^2/(n\,\gamma^2)} > \frac{1}{3}\, 4^{-\varepsilon^2/(n\,\gamma^2)}. \tag{10}$$

By Bakhvalov's trick (7) this is a lower bound for the worst case uncertainty $\sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A_n(f) - \mathrm{INT}(f)| > \varepsilon\}$, holding for any Monte Carlo algorithm A_n. Regarding the right-hand side of (10) as a given δ, the implication (8) finally provides the assertion. Pay attention that for too small δ, namely $0 < \delta < \frac{1}{3}\, 4^{-n}$, isolating ε in (10) is misleading to delusive error bounds exceeding γn which, however, violates the conditions on ε. In this case we can only conclude that $\varepsilon = \gamma n$ is a lower bound for $e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, \mathcal{W})$. $\square$

The following result will be useful for integrability $1 < p < 2$.

Lemma 2.2. *For $n \in \mathbb{N}$ and a natural number $N \geq 4n$, assume that there are functions $f_i \colon G \to \mathbb{R}$, with $i = 1, \ldots, N$, satisfying the following conditions:*

1. *for $i = 1, \ldots, N$, the sets $G_i := \{\mathbf{x} \in G \colon f_i(\mathbf{x}) \neq 0\}$ are pairwise disjoint, and $\mathrm{INT}(f_i) = \gamma$ for some $\gamma > 0$;*

2. *for $I \subset \{1, \ldots, N\}$ with $\#I = M$ for some given natural number $M \leq N$, and for signs $s_i \in \{\pm 1\}$, the function $f_{I,\mathbf{s}} := \sum_{i \in I} s_i\, f_i$ is an element of the input set $\mathcal{B}_{\mathcal{W}}$, that is, $\|f_{I,\mathbf{s}}\|_{\mathcal{W}} \leq 1$.*

Then, for any $0 < \delta < \frac{1}{2}\, 2^{-\lceil M/2 \rceil}$, we have

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, \mathcal{W}) \geq \tfrac{1}{2} \gamma M\,.$$

Proof. Let μ be the uniform distribution on the finite set

$$\mathcal{F} := \left\{ \sum_{i \in I} s_i f_i \colon I \subset \{1, \ldots, N\} \text{ with } \#I = M,\, s_i \in \{\pm 1\} \right\} \subset \mathcal{B}_{\mathcal{W}} .$$

Let $Q_n \colon \mathcal{F} \to \mathbb{R}$ be a deterministic algorithm using n function values. Without loss of generality, we may assume that the algorithm computes function values $y_i = f(\mathbf{x}_i)$ with $\mathbf{x}_i \in G_i$ for $i = 1, \ldots, n$. Hence, from the i-th piece of information we learn whether $i \in I$, and if so, whether $s_i = +1$ or -1 within the representation

$$f \;=\; f_{I,\mathbf{s}} \;:=\; \sum_{i \in I} s_i f_i \,.$$

Let $m(f) := \#(I \cap \{1, \ldots, n\})$ denote the number of detected subdomains G_i where the function f is non-zero. Under μ, the random variable $m(\cdot)$ is distributed according to a hypergeometric distribution with population of size N containing $M \leq N$ items of interest and admitting $n < N$ draws without replacement. The expected value is

$$\int_{\mathcal{F}} m(f) \, \mathrm{d}\mu(f) \;=\; \frac{n}{N} \, M \;\leq\; \frac{1}{4} \, M \,,$$

and using Markov's inequality we conclude

$$\mu \left\{ f \colon m(f) \leq \tfrac{1}{2} M \right\} \;\geq\; \frac{1}{2} \,. \tag{11}$$

Given the information $(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)) = \mathbf{y}$ with $m(f) = m$, there are still $k := M - m$ unknown s_i for subdomains G_i where the function does not vanish and the conditional distribution of $\mathrm{INT}(f)$ is given similarly to the proof of Lemma 2.1 (only n is substituted by m). Hence, the conditional uncertainty can be quantified via a binomial sum. For $0 < \varepsilon \leq \frac{1}{2} M \gamma$, up to $k' := \lfloor \varepsilon / \gamma \rfloor + 1 \leq \lceil \frac{1}{2} M \rceil$ terms are removed, and we obtain

$$\begin{aligned}
&\mu\bigl(|Q_n(f) - \mathrm{INT}(f)| > \varepsilon \;\big|\; (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)) = \mathbf{y},\, m(f) = m \bigr) \\
&\qquad \geq\; 2^{-k} \left[\sum_{j=0}^{\left\lfloor \frac{k-k'}{2} \right\rfloor} \binom{k}{j} + \sum_{j = \left\lceil \frac{k+k'+1}{2} \right\rceil}^{k} \binom{k}{j} \right] .
\end{aligned}$$

This bound is the same for all information outcomes $\mathbf{y}$ with the same number $m(f) = m$ of detected non-zero subdomains, and for $k \geq k'$, by Lemma A.2, we further estimate

$$\mu \left(|Q_n(f) - \mathrm{INT}(f)| > \varepsilon \mid m(f) = m \right) \geq 2^{-k'} . \tag{12}$$

Note that $m \leq \frac{1}{2}M$ implies $k = M - m \geq \lceil \frac{1}{2}M \rceil \geq k'$, so (12) can be used under the condition formulated in (11). Hence,

$$\begin{aligned}
\mu\{|Q_n(f) - &\mathrm{INT}(f)| > \varepsilon\} \\
&\geq \sum_{m=0}^{\lceil M/2 \rceil} \mu\left(|Q_n(f) - \mathrm{INT}(f)| > \varepsilon \mid m(f) = m\right) \cdot \mu\{f \colon m(f) = m\} \\
&\overset{(12)}{\geq} 2^{-k'} \cdot \mu\{f \colon m(f) \leq \tfrac{1}{2}M\} \;\geq\; \frac{1}{2}\, 2^{-\lceil M/2 \rceil}.
\end{aligned}$$

By Bakhvalov's trick (7) this is a lower bound for the worst case uncertainty $\sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A_n(f) - \mathrm{INT}(f)| > \varepsilon\}$, and for $0 < \delta < \frac{1}{2}\, 2^{-\lceil M/2 \rceil}$ the implication (8) proves the assertion. □

2.2 Lower bounds for Sobolev classes

The norms of classical Sobolev spaces $W_p^r(G)$ with *integrability index* p possess the property that for any decomposition of the support of a function $f \in W_p^r(G)$ into essentially disjoint sub-domains, say $G_1, \ldots, G_M$, we have

$$\|f\|_{W_p^r(G)} = \left(\sum_{i=1}^{M} \|f\|_{W_p^r(G_i)}^p \right)^{1/p} \leq M^{1/p} \max_{i=1,\ldots,M} \|f\|_{W_p^r(G_i)} \,. \tag{13}$$

The *smoothness* has an effect on scalings, namely for functions $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ and $\psi(\mathbf{x}) := \varphi(m\mathbf{x} - \mathbf{i})$, with $m > 0$ and $\mathbf{i} \in \mathbb{R}^d$, we have the following well known relation between the derivatives,

$$\|D^{\boldsymbol{\alpha}}\psi\|_{L_p(\mathbb{R}^d)} = m^{|\boldsymbol{\alpha}|_1 - d/p} \, \|D^{\boldsymbol{\alpha}}\varphi\|_{L_p(\mathbb{R}^d)} \,, \quad \text{for } \boldsymbol{\alpha} \in \mathbb{N}_0^d.$$

For $m \geq 1$ this leads to the scaling property

$$\|\psi\|_{W_p^r(\mathbb{R}^d)} \leq m^{r - d/p} \, \|\varphi\|_{W_p^r(\mathbb{R}^d)} \,. \tag{14}$$

If $\operatorname{supp}\varphi, \operatorname{supp}\psi \subseteq G$, then this relation holds also for the norms of the restricted space $W_p^r(G)$.

Theorem 2.3. *Let $G \subset \mathbb{R}^d$ be a domain with nonempty interior. Further, let $r \in \mathbb{N}_0$, $1 \leq p \leq \infty$ and define $q := \min\{p, 2\}$. Then we have the asymptotic lower bound*

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, W_p^r(G)) \succeq \min\left\{ n^{-r/d},\, n^{-(r/d + 1 - 1/q)} \left(\log \delta^{-1}\right)^{1 - 1/q} \right\} \,.$$

Proof. Since the interior of G is nonempty there exists a cubic subdomain. We restrict to functions which are supported within that rectangular subdomain, and by scaling, without loss of generality, we may assume $G = [0,1]^d$.

Let $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ be a sufficiently smooth function supported on $[0,1]^d$ with $\|\varphi\|_{W_p^r([0,1]^d)} \leq 1$ and $\gamma_0 := \mathrm{INT}\,\varphi > 0$. We call φ *bump function*. For $m \in \mathbb{N}$ we split $[0,1]^d$ into $N = m^d$ subcubes $G_{\mathbf{i}}$ with $\mathbf{i} \in [m]^d := \{0, \dots, m-1\}^d$ and equip each subcube with a scaled, shifted bump function $\psi_{\mathbf{i}}(\mathbf{x}) := \varphi(m\mathbf{x} - \mathbf{i})$.

If $2 \leq p \leq \infty$, we choose $m := \lceil (5n+6)^{1/d} \rceil$ and $f_{\mathbf{i}} := m^{-r}\, \psi_{\mathbf{i}}$. Hence, by (13) and (14) with $M = N$ we have

$$\left\| \sum_{\mathbf{i} \in [m]^d} s_{\mathbf{i}} f_{\mathbf{i}} \right\|_{W_p^r([0,1]^d)} \leq 1 \qquad \text{for arbitrary } s_{\mathbf{i}} \in \{\pm 1\}. \tag{15}$$

Then $\gamma = \mathrm{INT}\, f_{\mathbf{i}} = m^{-r-d} \gamma_0$. Restricting to $n \geq 17$ and $0 < \delta \leq 1/3$, we can apply Lemma 2.1 and obtain

$$\begin{aligned} e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, W_p^r([0,1]^d)) &\geq \gamma_0\, m^{-r-d} \min \left\{ n^{1/2} \sqrt{\log_4 \frac{1}{3\delta}},\, n \right\} \\ &\succeq \min\{ n^{-r/d-1/2} \sqrt{\log \delta^{-1}},\, n^{-r/d} \}\,. \end{aligned}$$

If $1 \leq p < 2$, we restrict to $0 \leq \delta < 1/4$ and choose $m := \lceil (4n)^{1/d} \rceil$. In the case $2^{-2n-1} \leq \delta$, we take $M = 2 \lceil \log_2 (4\delta)^{-1} \rceil \leq 2 \log_2 (2\delta)^{-1}$, and easily see that $M \leq 4n \leq N$ is fulfilled. Here, put $f_{\mathbf{i}} := m^{-r}\, (N/M)^{1/p}\, \psi_{\mathbf{i}}$, and note that by (13) and (14) we have

$$\left\| \sum_{\mathbf{i} \in I} s_{\mathbf{i}} f_{\mathbf{i}} \right\|_{W_p^r([0,1]^d)} \leq 1\,, \qquad \text{for arbitrary } s_{\mathbf{i}} \in \{\pm 1\} \text{ and } I \subseteq [m]^d \text{ with } \#I = M.$$

We have $\gamma = \mathrm{INT}\, f_{\mathbf{i}} = m^{-r-d+d/p}\, M^{-1/p}$, and Lemma 2.2 implies

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, W_p^r([0,1]^d)) \geq \tfrac{1}{2}\, \gamma_0\, m^{-r-d+d/p}\, M^{1-1/p} \succeq n^{-(r/d+1-1/p)} \left(\log \delta^{-1} \right)^{1-1/p}\,.$$

For small $\delta \in (0, 2^{-2n-1})$, however, we may just choose $M = 4n$, and similarly we obtain

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, W_p^r([0,1]^d)) \succeq n^{-r/d}\,,$$

which finishes the proof. □

Remark 2.4 (Lower bounds for non-integer smoothness)**.** There are several approaches to generalize Sobolev spaces for non-integer smoothness $r > 0$. For example the *Slobedeckii space* $W_p^r(G)$ is given as the set of functions with finite norm

$$\|f\|_{W_p^r(G)} := \left(\|f\|^p_{W_p^{\lfloor r \rfloor}(G)} + \sum_{|\boldsymbol{\alpha}|_1 = \lfloor r \rfloor} \int_G \int_G \frac{|D^{\boldsymbol{\alpha}} f(\mathbf{x}) - D^{\boldsymbol{\alpha}} f(\mathbf{z})|^p}{|\mathbf{x} - \mathbf{z}|^{d + (r - \lfloor r \rfloor)p}} \,\mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{z} \right)^{1/p},$$

where $1 \leq p < \infty$, see for instance the book of Triebel [27, p. 36]. For such spaces the inequality (13) does not hold anymore. However, we can still construct fooling functions composed of bumps on disjoint subcubes with random sign, but we need to introduce an additional constant in order to take the non-local nature of fractional derivatives into account.

Let us consider an easier example: Namely, classes of Hölder continuous functions with fractional smoothness $0 < \beta \leq 1$ (and integrability parameter $p = \infty$) given by

$$C^\beta([0,1]^d) := \left\{ f : [0,1]^d \to \mathbb{R} \,\middle|\, |f|_{C^\beta} := \sup_{\mathbf{x},\mathbf{z} \in [0,1]^d} \frac{|f(\mathbf{x}) - f(\mathbf{z})|}{|\mathbf{x} - \mathbf{z}|_\infty^\beta} < \infty \right\} . \tag{16}$$

With the choice $f_{\mathbf{i}} := \frac{1}{2}\, m^{-\beta}\, \psi_{\mathbf{i}}$, within the proof above, one can ensure (15). Thus, loosing just a factor $1/2$, we still have the same order as should be expected from generalizing the integer smoothness case,

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, C^\beta([0,1]^d)) \succeq n^{-\beta/d} \min\left\{1,\, \sqrt{\frac{\log \delta^{-1}}{n}}\right\} . \tag{17}$$

This fits very well to the upper bounds of Theorem 3.9.

3 Upper Bounds

3.1 Probability amplification

One of the most elementary methods of 'probability amplification' is the so-called 'median trick', see Alon et al. [2] and Jerrum et al. [14]. The following proposition is a minor modification of [19, Proposition 2.1, in particular (2.6)] from Niemiro and Pokarowski, now adapted to the language of algorithms and IBC.

As in Section 2.1, we consider a general function space $\mathcal{W}$ equipped with a (semi-)norm $\|\cdot\|_{\mathcal{W}}$ and take its unit ball $\mathcal{B}_{\mathcal{W}}$ as input set.

Proposition 3.1 (Median trick)**.** *For $\varepsilon > 0$ let A_m be an arbitrary Monte Carlo algorithm such that*

$$\sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A_m(f) - \mathrm{INT}\, f| > \varepsilon\} \leq \alpha \,,$$

where $0 < \alpha < 1/2$. For an odd natural number k, define

$$A_{k,m}(f) := \mathrm{med}\left\{A_m^{(1)}(f), \ldots, A_m^{(k)}(f)\right\}$$

as the median of k independent realisations of A_m. Then

$$\sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A_{m,k}(f) - \mathrm{INT}\, f| > \varepsilon\} \;\leq\; \frac{1}{2}(4\alpha(1-\alpha))^{k/2} \;<\; 2^{k-1}\, \alpha^{k/2} \,.$$

The previous proposition can be used to derive upper bounds for the probabilistic (ε, δ)-complexity $n^{\mathrm{MC}}_{\mathrm{prob}}(\varepsilon, \delta, \mathcal{W})$ in terms of the ℓ-mean error complexity $n^{\mathrm{MC}}_{\ell\text{-mean}}(\varepsilon, \mathcal{W})$, compare (4).

Theorem 3.2. *Let $\ell \geq 1$ and $0 < \delta \leq 1/2$. Then for the (ε, δ)-complexity holds*

$$n^{\mathrm{MC}}_{\mathrm{prob}}(\varepsilon, \delta, \mathcal{W}) \leq 2 \log_2 \delta^{-1} \cdot n^{\mathrm{MC}}_{\ell\text{-mean}}\left(8^{-1/\ell}\, \varepsilon,\, \mathcal{W}\right) .$$

In particular, if $e^{\mathrm{MC}}_{\ell\text{-mean}}(n, \mathcal{W}) \preceq n^{-\varrho}$ for some $\varrho > 0$, then we have

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, \mathcal{W}) \preceq \left(\frac{\log \delta^{-1}}{n}\right)^{\varrho} \qquad \text{for } n \succeq \log \delta^{-1}.$$

Proof. Without loss of generality, we assume that $n^{\mathrm{MC}}_{\ell\text{-mean}}(8^{-1/\ell}\varepsilon, \mathcal{W}) < \infty$, otherwise the claimed inequality is trivial. Hence, there is an $m \in \mathbb{N}$ such that $e^{\mathrm{MC}}_{\ell\text{-mean}}(m, \mathcal{W}) < 8^{-1/\ell}\varepsilon$. This implies that there is a Monte Carlo algorithm A_m such that

$$e^{\mathrm{MC}}_{\ell\text{-mean}}(A_m, \mathcal{W}) := \sup_{\|f\|_{\mathcal{W}} \leq 1} \left(\mathbb{E}\, \|A_m(f) - \mathrm{INT}\, f\|^{\ell}\right)^{1/\ell} \leq 8^{-1/\ell}\varepsilon.$$

Thus, for any $f \in \mathcal{B}_{\mathcal{W}}$ by Markov's inequality we have

$$\mathbb{P}\{|A_m(f) - \mathrm{INT}\, f| > \varepsilon\} \leq \left(\frac{e^{\mathrm{MC}}_{\ell\text{-mean}}(A_m, \mathcal{W})}{\varepsilon}\right)^{\ell} \leq \frac{1}{8}.$$

Now we aim to apply Proposition 3.1 with k chosen as the smallest odd natural number that satisfies $k \geq 2 \log_2 (2\delta)^{-1}$. (Note that $k \leq 2 \log_2 \delta^{-1}$ for $0 < \delta \leq 1/2$.) Then we obtain the desired complexity bound

$$n^{\mathrm{MC}}_{\mathrm{prob}}(\varepsilon, \delta, \mathcal{W}) \leq k \cdot n^{\mathrm{MC}}_{\ell\text{-mean}}\left(8^{-1/\ell}\, \varepsilon, \mathcal{W}\right) \leq 2 \log_2 \delta^{-1} \cdot\, n^{\mathrm{MC}}_{\ell\text{-mean}}\left(8^{-1/\ell}\, \varepsilon,\, \mathcal{W}\right) .$$

In terms of the error quantities, for fixed m and odd $k \geq 2 \log_2 (2\delta)^{-1}$ we can state

$$e^{\mathrm{MC}}_{\mathrm{prob}}(km, \delta, \mathcal{W}) \leq 8^{-1/\ell}\, e^{\mathrm{MC}}_{\ell\text{-mean}}(m, \mathcal{W}) .$$

Assuming $e^{\mathrm{MC}}_{\ell\text{-mean}}(m, \mathcal{W}) \leq C\, m^{-\varrho}$ for $m \geq m_0 \in \mathbb{N}$, and given an information budget $n \geq 2 m_0 \log_2 \delta^{-1}$, put $m := \lfloor n/(2 \log_2 \delta^{-1}) \rfloor \asymp n/\log \delta^{-1}$. Then the assertion follows by the assumption with hidden constant $\left(\frac{2}{\log 2}(1 + 1/m_0)\right)^{\varrho} \cdot C$. $\square$

Remark 3.3 (Integrating L_p-functions)**.** The lower bounds of Theorem 2.3 match the upper bounds from Theorem 3.2, iff the rate of convergence is related to the integrability index by $\varrho = 1 - 1/q$ where $q := \min\{p, 2\}$. This is only the case for smoothness $r = 0$, i.e., when L_p-balls are the considered input sets. In that case,

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, L_p) \asymp \left(\frac{\log \delta^{-1}}{n}\right)^{1-1/q}, \qquad \text{for } n \succeq \log \delta^{-1}. \tag{18}$$

Here we used estimates of the q-mean error for the standard i.i.d.-based Monte Carlo method applied to L_p-functions which, for example, can be found in [3, Theorem 2], [9, Proposition 5.4], [20, Sect. 2.2.8, Proposition 3], as well as [25, Proof of Theorem 1].

Remark 3.4 (Alternatives to the median trick)**.** Catoni [5] proposes an alternative scheme based on random samples of L_p-functions, which suppresses outliers and (in contrast to the median trick) is symmetric in the sense that permuting the sample data does not change the result. Compare also Huber [13] where this approach is combined with the median trick. Unfortunately, their methods are not homogeneous and shift invariant, that is, for the algorithm, say A, in general $A(af + b) = aA(f) + b$ does not hold.

3.2 Separation of the main part

Separation of the main part, also known as *control variates*, is a well established technique of variance reduction which uses the approximation of functions with respect to an L_q-norm in order to exploit the smoothness of the given input set.

Within this section we assume that $G \subset \mathbb{R}^d$ is a bounded Lipschitz domain, see [21] for details. For $q \geq 1$ let $L_q(G)$ be the Lebesgue space equipped with the norm $\|\cdot\|_{L_q(G)}$. Let $\mathcal{W} \subset L_q(G)$ be a normed linear space with corresponding unit ball $\mathcal{B}_{\mathcal{W}}$ and assume that function evaluations are continuous (well-defined) on $\mathcal{W}$. For the approximation step we only consider linear methods

$$A_n \colon \mathcal{W} \to L_q(G), \qquad f \mapsto g := \sum_{i=1}^{n} f(\mathbf{x}_i)\, g_i\,, \tag{19}$$

with nodes $\mathbf{x}_i \in G$ and functions $g_i \in L_q(G)$, where $\int_G g_i(\mathbf{x})\,\mathrm{d}\mathbf{x}$ is known for any $i = 1, \ldots, n$. The minimal L_q-approximation error of such methods is denoted by

$$e^{\mathrm{det}}(n, \mathcal{W} \hookrightarrow L_q) := \inf_{A_n} \sup_{\|f\|_{\mathcal{W}} \leq 1} \|A_n(f) - f\|_{L_q(G)}\,, \tag{20}$$

and the ε-complexity $n^{\mathrm{det}}(\varepsilon, \mathcal{W} \hookrightarrow L_q)$ is the minimal number of function evaluations needed in order to achieve an L_q-approximation error smaller than ε. The idea is to apply a Monte Carlo integration method $M_n \colon L_q \to \mathbb{R}$ to the difference $f - g$ between approximating and original function, while the integral of $g = A_n(f)$ is considered to be known. This approach leads to the following theorem.

Theorem 3.5 (Separation of the main part)**.** *For any $n \in \mathbb{N}$ we have*

$$e^{\mathrm{MC}}_{\mathrm{prob}}(2n, \delta, \mathcal{W}) \leq e^{\mathrm{det}}(n, \mathcal{W} \hookrightarrow L_q) \cdot e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, L_q)\,.$$

Proof. Let A_n be a linear approximation method, see (19), which guarantees for any $f \in \mathcal{B}_{\mathcal{W}}$ and some $\alpha > 0$ that

$$\|A_n(f) - f\|_{L_q(G)} \leq \alpha.$$

Further, let M_n^ω be a Monte Carlo method which approximates $\mathrm{INT}\, h$ for inputs $h \in L_q$. With this we define a new Monte Carlo method Q_{2n}^ω (a randomized quadrature rule) as follows:

1. Compute the approximation $g := A_n(f) \in L_q(G)$, using function values $f(\mathbf{x}_i)$ at nodes $\mathbf{x}_1, \dots, \mathbf{x}_n \in G$.

2. Return
$$Q_{2n}^\omega(f) := \operatorname{INT} g + \alpha M_n^\omega\left(\alpha^{-1}(f-g)\right)$$
where M_n evaluates $\alpha^{-1}(f-g)$ at random nodes $\mathbf{X}_{n+1}^\omega, \dots, \mathbf{X}_{2n}^\omega \in G$. (If M_n is homogeneous, that is, $M_n^\omega(\lambda f) = \lambda M_n^\omega(f)$ for any $\lambda \in \mathbb{R}$, then α cancels out.)

Note that, the information $\mathbf{y} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n), f(\mathbf{X}_{n+1}^\omega), \dots, f(\mathbf{X}_{2n}^\omega))$ suffices to execute the algorithm, namely,

$$\operatorname{INT} g = \sum_{i=1}^n f(\mathbf{x}_i) \operatorname{INT} g_i\,, \qquad \text{and}$$

$$[\alpha^{-1}(f-g)](\mathbf{X}_{n+j}^\omega) = \alpha^{-1}\left(f(\mathbf{X}_{n+j}^\omega) - \sum_{i=1}^n f(\mathbf{x}_i)\, g_i(\mathbf{X}_{n+j}^\omega)\right) \qquad \text{for } j = 1, \dots, n.$$

Indeed, $\operatorname{INT} g_i$ is assumed to be precomputed and computing function values of g_i is considered to belong to the combinatorial cost, rather than the information cost of the algorithm.

By writing $\varepsilon = \alpha\varepsilon'$, the uncertainty of the algorithm can be traced back to the uncertainty of M_n. Namely, if M_n is (ε', δ)-approximating in $L_q(G)$, then

$$\begin{aligned}\mathbb{P}\left\{|Q_{2n}(f) - \operatorname{INT} f| > \varepsilon\right\} &= \mathbb{P}\left\{|M_n(\alpha^{-1}(f-g)) - \operatorname{INT}(\alpha^{-1}(f-g))| > \varepsilon'\right\} \\ &\leq \delta\,.\end{aligned}$$

Optimal methods A_n and M_n^ω lead to $\alpha \to e^{\det}(n, \mathcal{W} \hookrightarrow L_q)$ and $\varepsilon' \to e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, L_q)$, while keeping the uncertainty bounded by δ, thus letting ε approach the stated error bound. ☐

As long as function evaluations are continuous, it suffices to work with deterministic approximation methods of the form (19). Note that for isotropic Sobolev spaces $W_p^r(G)$ on bounded Lipschitz domains $G \subset \mathbb{R}^d$, this is the case iff $rp > d$. In this setting it is well known that with $q := \min\{p, 2\}$,

$$e^{\det}\left(n, W_p^r(G) \hookrightarrow L_q(G)\right) \asymp n^{-r/d}\,, \qquad \text{if } rp > d. \tag{21}$$

For $G = [0,1]^d$, this result can be achieved with piecewise polynomial interpolation, see for instance Heinrich [9, Proposition 5.1], technical details of approximation methods are contained in Ciarlet [6]. For the general case of bounded Lipschitz domains $G \subset \mathbb{R}^d$, see Novak and Triebel [21, Theorem 23]. From this we conclude optimal upper bounds.

Theorem 3.6. *Let $G \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Further let $r \in \mathbb{N}$ and $1 \leq p \leq \infty$ with $rp > d$. Then we have the asymptotic rate*

$$e^{\mathrm{MC}}_{\mathrm{prob}}\left(n, \delta, W_p^r(G)\right) \asymp n^{-r/d} \min\left\{1, \left(\frac{\log \delta^{-1}}{n}\right)^{1-1/q}\right\},$$

where $q := \min\{2, p\}$.

Proof. The lower bounds follow from Theorem 2.3.

For $n \succeq \log \delta^{-1}$, we combine (18) with (21) via Theorem 3.5 and obtain

$$e^{\mathrm{MC}}_{\mathrm{prob}}\left(n, \delta, W^r_p(G)\right) \preceq n^{-r/d} \left(\frac{\log \delta^{-1}}{n}\right)^{1-1/q} .$$

For $n \prec \log \delta^{-1}$ we rely on deterministic quadrature. This problem is easier than approximation in the sense that if $g \in L_q(G)$ is an approximation of f, then $|\operatorname{INT} f - \operatorname{INT} g| \leq \operatorname{Vol}_d(G)^{1-1/q} \cdot \|f - g\|_{L_q(G)}$. Hence,

$$\begin{aligned} e^{\det}(n, W^r_p(G)) &:= \inf_{A_n} \sup_{\|f\|_{\mathcal{W}} \leq 1} |A_n(f) - \operatorname{INT} f| \\ &\preceq e^{\det}(n, W^r_p(G) \hookrightarrow L_q(G)) \asymp n^{-r/d} . \end{aligned}$$

See also Novak [20, 1.3.12] for a direct derivation on $G = [0, 1]^d$. □

Remark 3.7 (Lower smoothness)**.** The condition $rp > d$ is necessary to guarantee that the evaluation of functions on $W^r_p(G)$ for $1 < p \leq \infty$ is well-defined. (For $p = 1$, the condition $r = d$ is also sufficient, but then deterministic methods already provide the optimal error rates.) In the cases $1 < p < \infty$ with $rp \leq d$ one can still use separation of the main part, but with a randomized approximation scheme, see Heinrich [10] for the case $G = [0, 1]^d$. That way, for any $1 \leq p \leq \infty$ and general $r \in \mathbb{N}$ we have

$$e^{\mathrm{MC}}_{\text{1-mean}}(n, W^r_p([0, 1]^d)) \asymp n^{-(r/d+1-1/q)} ,$$

with $q := \min\{p, 2\}$. Probability amplification, see Theorem 3.2, yields

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, W^r_p([0, 1]^d)) \asymp \left(\frac{\log \delta^{-1}}{n}\right)^{r/d+1-1/q} , \qquad \text{for } n \succeq \log \delta^{-1}.$$

The power of $\log \delta^{-1}$ in this upper bound may exceed the power of the lower bound by r/d which can get close to 1 for $p \to 1$. We conjecture that the influence of δ is smaller at least if one is close to the regime where functions are continuous.

Instead of deterministic algorithms of the form (19) one might also consider general randomized methods for the approximation of functions in $\mathcal{W}$. For those let $e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, \mathcal{W} \hookrightarrow L_q)$ be the smallest $\varepsilon > 0$, such that there exists a general randomized approximation algorithm satisfying

$$\mathbb{P}\{\|A_n(f) - f\|_{L_q} > \varepsilon\} \leq \delta \qquad \text{for all } \|f\|_{\mathcal{W}} \leq 1.$$

Similarly to Theorem 3.5 one can show that

$$e^{\mathrm{MC}}_{\mathrm{prob}}(2n, \delta, \mathcal{W}) \leq e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta/2, \mathcal{W} \hookrightarrow L_q) \cdot e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta/2, L_q) . \tag{22}$$

Such an approach, however, seems to rely on complicated algorithms, since non-linearity might be inevitable in order to suppress outliers. There might be easier implementable Monte Carlo methods for integration which achieve a better order of convergence without relying on the approximation of functions. Such methods are needed in spaces of mixed smoothness, see the discussion in Section 4.

Anyway, studying Sobolev embeddings $W_p^r(G) \hookrightarrow L_q(G)$ in terms of approximation with high confidence within the regime $d(q-p) < rqp \leq dq$ for general integrability parameters $1 \leq p, q < \infty$, is an interesting problem on its own, compare Heinrich [10].

3.3 Stratified sampling

Let us introduce stratified sampling for the approximation of INT f for integrable functions defined on $G = [0,1]^d$. For $m \in \mathbb{N}$ we split the unit cube $[0,1]^d$ into m^d subcubes given by

$$G_{\mathbf{i}} = \prod_{j=1}^{d} \left[\frac{i_j}{m}, \frac{i_j+1}{m}\right) ,$$

with $\mathbf{i} \in [m]^d := \{0, \ldots, m-1\}^d$ and $\mathbf{i} = (i_1, \ldots, i_d)$. Let $(\mathbf{X_i})_{\mathbf{i}\in[m]^d}$ be a sequence of independent random variables with $\mathbf{X_i}$ uniformly distributed in $G_{\mathbf{i}}$. Then, stratified sampling is given by

$$S_m^d(f) := \frac{1}{m^d} \sum_{\mathbf{i}\in[m]^d} f(\mathbf{X_i}) , \tag{23}$$

which uses m^d function evaluations of f. Compared to the separation of the main part, stratified sampling is easier to implement. In some cases we show that it provides optimal results in terms of the (ε, δ)-complexity. Since the structure only depends on the information budget and not on δ (compare the median trick), we obtain a universal method in terms of the uncertainty.

We use Hoeffding's inequality which is, for the convenience of the reader, stated in the following proposition.

Proposition 3.8 (Hoeffding's inequality)**.** *Let $Y_1, \ldots, Y_n$ be independent random variables supported on intervals of length $b_i > 0$, that is,* $\operatorname{ess\,sup} Y_i - \operatorname{ess\,inf} Y_i \leq b_i$. *Then, for $S_n := \frac{1}{n}\sum_{i=1}^n Y_i$ and $\varepsilon > 0$ we have*

$$\mathbb{P}\left\{|S_n - \mathbb{E}\, S_n| > \varepsilon\right\} \leq 2 \exp\left(-\frac{2\, n^2\, \varepsilon^2}{\sum_{i=1}^n b_i^2}\right) .$$

First, we consider Hölder classes $C^\beta([0,1]^d)$ with smoothness $\beta \in (0,1]$, see (16). Compare also [4] for the result in terms of the root mean squared error.

Theorem 3.9. *For the classes of Hölder-continuous functions on $[0,1]^d$, stratified sampling achieves the optimal rate of convergence, namely*

$$e^{\mathrm{MC}}_{\mathrm{prob}}\left(n, \delta, C^\beta([0,1]^d)\right) \asymp n^{-\beta/d} \min\left\{1, \sqrt{\frac{\log \delta^{-1}}{n}}\right\} .$$

Proof. Concerning the lower bounds, see (17) in Remark 2.4.

For the upper bounds, we start with the case $n = m^d$ with $m \in \mathbb{N}$ and employ S_m^d. Obviously, this method is unbiased, i.e., $\mathbb{E}\, S_m^d(f) = \mathrm{INT}\, f$. By Hölder continuity, the random variables $Y_{\mathbf{i}} = f(\mathbf{X_i})$ are spread on intervals of length $b_{\mathbf{i}} \leq m^{-\beta}$ for $|f|_{C^\beta} \leq 1$. This implies $|S_m^d(f) - \mathrm{INT}\, f| \leq m^{-\beta} = n^{-\beta/d}$. Hoeffding's inequality, Proposition 3.8, leads to

$$\mathbb{P}\{|S_m^d(f) - \mathrm{INT}\, f| > \varepsilon\} \leq 2\exp\left(-\frac{2\, m^{2d}\, \varepsilon^2}{m^d \cdot m^{-2\beta}}\right) = 2\exp\left(-2\, m^{d+2\beta}\varepsilon^2\right) .$$

This is guaranteed to be at most δ for

$$\varepsilon = \frac{1}{\sqrt{2}}\, m^{-(\beta+d/2)} \sqrt{\log\frac{2}{\delta}} = \frac{1}{\sqrt{2}}\, n^{-(\beta/d+1/2)} \sqrt{\log\frac{2}{\delta}}\, .$$

Given an information budget $n \in \mathbb{N}$, we choose $m := \lfloor n^{1/d} \rfloor$. Employing the method S_m^d, see (23), we actually only use $m^d \leq n$ function values. For $n \geq 2^d$ we have $m \geq \frac{1}{2}\, n^{1/d}$, hence we obtain the stated asymptotics. $\square$

Now we consider the isotropic Sobolev classes $W_p^1([0,1]^d)$ of smoothness 1. For Hoeffding's inequality to be applicable, we need $W_p^1([0,1]^d) \hookrightarrow L_\infty([0,1]^d)$ which is the case for $p > d$.

Theorem 3.10. *Stratified sampling leads to*

$$e^{\mathrm{MC}}_{\mathrm{prob}}(n, \delta, W_p^1([0,1]^d)) \preceq n^{-1/d} \min\left\{1,\, n^{-(1-1/q)} \sqrt{\log \delta^{-1}}\right\}\,, \qquad \textit{for } p > d,$$

where $q := \min\{p, 2\}$. *For* $p \geq 2$, *this perfectly matches the lower bounds from Theorem 2.3.*

Proof. We start with the one-dimensional case considering the method S_n^1, see (23). Hence, the unit interval $[0,1]$ is split into intervals $G_0, \ldots, G_{n-1}$ of length n^{-1}. For $x_1 < x_2$ from $[0,1]$ we have

$$|f(x_2) - f(x_1)| = \left|\int_{[x_1,x_2]} f'(x)\, \mathrm{d}x\right| \leq \int_{[x_1,x_2]} |f'(x)|\, \mathrm{d}x\, .$$

Hence, on the ith interval G_i we have

$$b_i \leq \int_{G_i} |f'(x)|\, \mathrm{d}x \leq n^{-1}\left(n \int_{G_i} |f'(x)|^q\, \mathrm{d}x\right)^{1/q} = n^{-(1-1/q)}\, \|f'\|_{L_q(G_i)}\,,$$

where we used Jensen's inequality. Furthermore

$$\|f'\|_{L_q([0,1])} = \left(\sum_{i=1}^n \|f'\|_{L_q(G_i)}^q\right)^{1/q} \geq n^{1-1/q}\left(\sum_{i=1}^n b_i^q\right)^{1/q} \geq n^{1-1/q}\left(\sum_{i=1}^n b_i^2\right)^{1/2}\,,$$

exploiting $q \leq 2$ in the last inequality. Applying Hoeffding's inequality, Proposition 3.8, for $\|f\|_{W_p^1([0,1])} \leq 1$ we obtain

$$\mathbb{P}\{|S_n^1(f) - \mathrm{INT}\, f| > \varepsilon\} \leq 2 \exp\left(-2\, n^{4-2/q}\, \varepsilon^2\right) .$$

This is guaranteed to be at most δ for

$$\varepsilon = \frac{1}{\sqrt{2}}\, n^{-(2-1/q)} \sqrt{\log \frac{2}{\delta}} \,,$$

which shows the assertion for $d = 1$.

In higher dimension, $d \geq 2$, splitting $[0,1]^d$ into m^d subcubes $G_{\mathbf{i}}$ with $\mathbf{i} \in [m]^d$, we exploit the embedding $W_p^1([0,1]^d) \hookrightarrow L_\infty([0,1]^d)$. Namely, incorporating scaling we bound the spread of function values within $G_{\mathbf{i}}$ by

$$b_{\mathbf{i}} := \operatorname*{ess\,sup}_{G_{\mathbf{i}}} f - \operatorname*{ess\,inf}_{G_{\mathbf{i}}} f \leq C\, m^{d/p-1}\, \|f\|_{W_p^1(G_{\mathbf{i}})} \,,$$

with some constant $C > 0$ depending only on p and d. From this, with $p > d \geq 2$, we conclude

$$\left(\sum_{\mathbf{i}\in[m]^d} b_{\mathbf{i}}^2\right)^{1/2} \leq m^{d(1/2-1/p)} \left(\sum_{\mathbf{i}\in[m]^d} b_{\mathbf{i}}^p\right)^{1/p} \leq C\, m^{-(1-d/2)}\, \|f\|_{W_p^1([0,1]^d)} \,,$$

compare (13). Choosing $m := \lceil n^{1/d} \rceil$, we obtain the right order by applying Hoeffding's inequality similarly to the one-dimensional case. □

Remark 3.11. The one-dimensional problem contains cases of small integrability $1 < p < 2$ for which we do not obtain the optimal δ-dependence. It is not known to us whether this is a deficiency of the method or of the proof. In that case, we may use separation of the main part, which is equally simple as f may be approximated on $G_{\mathbf{i}}$ by just one function value.

In the case of discontinuous functions, $p < d$, it remains challenging to find methods which detect and discourage outliers within stratified sampling. One idea might be to take several function values out of each subcube. This could improve also on the above mentioned case $d = 1$ and $1 < p < 2$. Any result in that direction might offer reasonable alternatives to control variates, where the case of small smoothness is also open.

4 Challenges in mixed smoothness spaces

In the recent years spaces of dominating mixed smoothness gained a lot of interest in the study of high-dimensional problems. For a survey on this topic we refer to the paper of Dũng, Temlyakov, and Ullrich [7].

For integer smoothness $r \in \mathbb{N}$ and integrability $1 \leq p \leq \infty$, on domains $G \subset \mathbb{R}^d$, Sobolev spaces of dominating mixed smoothness can be defined by

$$\mathbf{W}_p^{\mathrm{mix},r}(G) := \left\{ f \in L_p(G) \;\middle|\; \|f\|_{\mathbf{W}_p^{\mathrm{mix},r}(G)} := \Bigg(\sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}_0^d \\ |\boldsymbol{\alpha}|_\infty \leq r}} \|D^{\boldsymbol{\alpha}} f\|_{L_p(G)}^p \Bigg)^{1/p} \leq \infty \right\}.$$

Lower bounds of the integration problem can be shown by scaling bump functions $\varphi\colon [0,1]^d \to \mathbb{R}$ in one coordinate, that is, for $m \in \mathbb{N}$ we define functions $\psi_i(\mathbf{x}) := \varphi(mx_1 - i, x_2, \ldots, x_d)$, where $i \in \{0, \ldots, m-1\} = [m]$. By using those, similarly to Theorem 2.3 one can obtain

$$e_{\mathrm{prob}}^{\mathrm{MC}}(n, \delta, \mathbf{W}_p^{\mathrm{mix},r}([0,1]^d)) \succeq \min \left\{ n^{-r},\, n^{-(r+1-1/q)} \,(\log \delta^{-1})^{1-1/q} \right\} . \tag{24}$$

When talking about upper bounds it is useful to note that the integration problem is as difficult for the non-periodic spaces $\mathbf{W}_p^{\mathrm{mix},r}(G)$ as for the zero-boundary space $\mathring{\mathbf{W}}_p^{\mathrm{mix},r}(G) := \{f \in \mathbf{W}_p^{\mathrm{mix},r}(\mathbb{R}^d) \mid \operatorname{supp} f \subseteq G\}$. Namely, the integral of any function $f \in \mathbf{W}_p^{\mathrm{mix},r}([0,1]^d)$, via a change of variables, can be traced back to the integral of a function $h := |\det \Phi'| \cdot (f \circ \Phi) \in \mathring{\mathbf{W}}_p^{\mathrm{mix},r}([0,1]^d)$ with zero boundary condition, where $\Phi\colon [0,1]^d \to [0,1]^d$ is a smooth bijection. That way we only loose a constant, see Nguyen, Ullrich, and Ullrich [18]. Let us mention that our lower bounds are based on bump functions with zero boundary, so the lower bounds hold with the same constants.

The optimal order of convergence in terms of the root mean squared error is determined by Ullrich [28], namely

$$e_{2\text{-mean}}^{\mathrm{MC}}(n, \mathbf{W}_p^{\mathrm{mix},r}([0,1]^d)) \asymp n^{-(r+1-1/q)}\,, \qquad \text{for } r \geq \max\{1/p - 1/2, 0\},$$

where $q = \min\{p, 2\}$. The result is based on a randomly shifted and dilated Frolov rule, developed by Krieg and Novak [15], given by

$$Q_{B,\mathbf{v}}(f) := \frac{1}{|\det B|} \sum_{\mathbf{m} \in \mathbb{Z}^d} f(B^{-\top}(\mathbf{m} + \mathbf{v}))\,,$$

where $f \in \mathring{\mathbf{W}}_p^{\mathrm{mix},r}([0,1]^d)$, which is of course only evaluated inside $[0,1]^d$. Here, $B = \operatorname{diag}(\mathbf{u}) B_n$ with dilation random variable $\mathbf{u}$ and independent shift random variable $\mathbf{v}$ distributed according to the uniform distribution in $[1/2, 3/2]^d$ and $[0,1]^d$, respectively, as well as a suitable generator matrix $B_n = n^{1/d} B_1$. 'Suitable' means $\det B_n = n$ and $\prod_{j=1}^d |(B_1 \mathbf{m})_j| \geq c > 0$ for all $\mathbf{m} \in \mathbb{Z}^d \setminus \{0\}$. In particular, the expected number of function evaluations is n. (As mentioned before, the lower bounds from Section 2 can be extended to methods with varying cardinality which will only affect constants.) Via Theorem 3.2 one can build a method by independent repetition which provides

$$e_{\mathrm{prob}}^{\mathrm{MC}}(n, \delta, \mathbf{W}_p^{\mathrm{mix},r}([0,1]^d)) \preceq \left(\frac{\log \delta^{-1}}{n} \right)^{r+1-1/q} \qquad \text{for } n \succeq \log \delta^{-1}, \tag{25}$$

with $q := \min\{p, 2\}$. Unfortunately, we do not achieve the optimal dependence on δ. The original algorithm alone does not possess desirable confidence guarantees, as the following one-dimensional counter example shows. This is not surprising as the number of random parameters is fixed by the dimension and thus we do not expect to observe concentration phenomena, which is in contrast to stratified sampling.

Example 4.1. We consider the integration problem in a one dimensional setting on $\mathring{\mathbf{W}}_2^{\mathrm{mix},r}([0,1])$. The random Frolov rule that uses n function values on average is determined by

$$Q_n(f) := \frac{1}{un} \sum_{m \in \mathbb{Z}} f\left(\frac{m+v}{un}\right),$$

with independent random variables u and v uniformly distributed in $[1/2, 3/2]$ and $[0,1]$, respectively. Let $\varphi \in \mathring{\mathbf{W}}_2^{\mathrm{mix},r}([0,1])$ be a bump function with integral $\gamma_0 := \int_0^1 \varphi \,\mathrm{d}x$ and norm $\|\varphi\|_{\mathbf{W}_2^{\mathrm{mix},r}} \leq 1$. For

$$f_n(x) := (2n)^{-r} \sum_{k=0}^{n-1} \varphi(2nx - 2k)$$

observe that $\|f_n\|_{\mathbf{W}_2^{\mathrm{mix},r}} \leq 1$ and $\int_0^1 f_n \,\mathrm{d}x = \gamma_0/2^{r+1} \cdot n^{-r}$. Furthermore, the algorithm returns 0 if all the function values are computed inside $\bigcup_{k=0}^{n-1} \left[\frac{2k+1}{2n}, \frac{k+1}{n}\right]$, which is where f_n vanishes. This happens if $\frac{v}{un} \in [\frac{1}{2n}, \frac{1}{n}]$ and $\frac{n-1+v}{un} \in \left[\frac{2n-1}{2n}, 1\right]$, in particular for shifts $v \in [\frac{1}{2}, \frac{3}{4}]$ and dilations $u \in [1 - \frac{1}{4n}, 1]$. This means, with probability exceeding $\delta_n := \frac{1}{16n}$ the error is $\gamma_0/2^{r+1} \cdot n^{-r}$, hence,

$$e_{\mathrm{prob}}^{\mathrm{MC}}(Q_n, \delta_n, \mathring{\mathbf{W}}_2^{\mathrm{mix},r}([0,1])) \succeq n^{-r} \succ n^{-(r+1/2)} \sqrt{\log \delta_n^{-1}} \asymp n^{-(r+1/2)} \sqrt{\log n}\,.$$

This reveals a significant gap to the general lower bound (24). $\square$

Separation of the main part does not provide the optimal rate in n, but the dependence on δ can be reduced. Since we may restrict to the integration problem for functions with zero boundary condition, $\mathring{\mathbf{W}}_p^{\mathrm{mix},r}([0,1])$, we may apply results for the approximation of periodic functions, denoted by $\widetilde{\mathbf{W}}_p^{\mathrm{mix},r}([0,1]^d)$. Namely,

$$e^{\det}(n, \widetilde{\mathbf{W}}_p^{\mathrm{mix},r}([0,1]^d) \hookrightarrow L_p) \preceq n^{-r} (\log n)^{(r+1/2)(d-1)}, \qquad \text{for } 1 < p < \infty,$$

which can be found in [7, (5.11)]. Applying Theorem 3.5, for $n \succeq \log \delta^{-1}$ we conclude that

$$e_{\mathrm{prob}}^{\mathrm{MC}}(n, \delta, \mathbf{W}_p^{\mathrm{mix},r}([0,1]^d)) \preceq n^{-(r+1-1/q)} (\log n)^{(r+1/2)(d-1)} (\log \delta^{-1})^{1-1/q}, \tag{26}$$

where $q := \min\{p, 2\}$. Here the δ-dependence is optimal, but the rate in n is affected by logarithmic terms.

Finally, deterministic quadrature is known to achieve

$$e^{\det}(n, \mathbf{W}_p^{\mathrm{mix},r}([0,1]^d)) \asymp n^{-r} (\log n)^{(d-1)/2}, \qquad \text{for } 1 < p < \infty, \tag{27}$$

see [7, Theorem 8.14]. This catches the case $n \prec \log \delta^{-1}$.

It remains a challenging open problem to find randomized integration methods which have the right dependence on the uncertainty while fully exploiting the smoothness.

A Technical Proofs

In Section 2 we need the following two inequalities about binomial sums. The first lemma is a minor extension of [17, Proposition 7.3.2], holding also for odd k, and with slightly improved constants.

Lemma A.1. *For $k \in \mathbb{N}$ and $t \in \mathbb{N}_0$ we have*

$$\begin{aligned} 2^{-k} \sum_{j=0}^{\lfloor k/2 \rfloor - t} \binom{k}{j} &= 2^{-k} \sum_{j=\lceil k/2 \rceil + t}^{k} \binom{k}{j} \\ &\geq \frac{1}{2 + 4/\sqrt{\pi}} \begin{cases} \exp\left(-\dfrac{16\,(\log 2)\,t^2}{k}\right) & \text{for odd } k \text{ and } t \in [0, \frac{k+3}{8}], \\ \exp\left(-\dfrac{16\,(\log 2)\,(t-1/2)^2}{k}\right) & \text{for even } k \text{ and } t \in [0, \frac{k+6}{8}]. \end{cases} \end{aligned}$$

Proof. First, recall that $\binom{k}{\lfloor k/2 \rfloor} < 2^k / \sqrt{\pi \lceil k/2 \rceil}$, which for even k follows from Stirling's formula and for odd k can be derived from $k+1$ via Pascal's rule. Hence,

$$2^{-k} \sum_{j=0}^{\lfloor k/2 \rfloor - t} \binom{k}{j} \geq \frac{1}{2} - 2^{-k}\, t \binom{k}{\lfloor k/2 \rfloor} > \frac{1}{2} - \frac{t}{\sqrt{\pi \lceil k/2 \rceil}}\,.$$

For $0 \leq t \leq \sqrt{\lceil k/2 \rceil}/(1 + 2/\sqrt{\pi})$, this gives the absolute lower bound $\frac{1}{2+4/\sqrt{\pi}}$.

For larger t we follow the approach of [17, Proposition 7.3.2]. Basic estimates yield

$$\begin{aligned} 2^{-k} \sum_{j=0}^{\lfloor k/2 \rfloor - t} \binom{k}{j} &\geq 2^{-k} \sum_{j=\lfloor k/2 \rfloor - 2t + 1}^{\lfloor k/2 \rfloor - t} \binom{k}{j} \\ &\geq 2^{-k}\, t \binom{k}{\lfloor k/2 \rfloor - 2t + 1} \\ &= 2^{-k}\, t \binom{k}{\lfloor k/2 \rfloor} \prod_{i=1}^{2t-1} \frac{\lfloor k/2 \rfloor - 2t + 1 + i}{\lceil k/2 \rceil + i} \\ &\geq 2^{-k}\, t \binom{k}{\lfloor k/2 \rfloor} \left(\frac{\lfloor k/2 \rfloor - 2t + 2}{\lceil k/2 \rceil + 1}\right)^{2t-1}. \end{aligned}$$

Next, we use $1 - x \geq \exp(-2\,(\log 2)\,x)$ for $0 \leq x \leq 1/2$. For odd k we set $x = 2t/(\lceil k/2 \rceil + 1)$ and for even k we set $x = (2t-1)/(k/2+1)$. This is where $t \leq$

$(k+6)/8$ for even k, and $t \le (k+3)/8$ for odd k, comes into play. Finally, we use $\binom{k}{\lceil k/2 \rceil} \ge 2^k/(2\sqrt{\lceil k/2 \rceil})$, and obtain

$$2^{-k} \sum_{j=0}^{\lfloor k/2 \rfloor - t} \binom{k}{j} \ge \frac{t}{2\sqrt{\lceil k/2 \rceil}} \begin{cases} \exp\Big[-\dfrac{8\,(\log 2)\, t\,(t-1/2)}{\lceil k/2 \rceil + 1}\Big] & \text{for odd } k, \\ \exp\Big[-\dfrac{8\,(\log 2)\,(t-1/2)^2}{k/2+1}\Big] & \text{for even } k. \end{cases}$$

For $t \ge \sqrt{\lceil k/2 \rceil}/(1+2/\sqrt{\pi})$, the prefactor simplifies as stated in the claimed inequality. $\square$

Lemma A.2. *For $k, k' \in \mathbb{N}_0$ we have for all $k \ge k'$ that*

$$2^{-k} \left[\sum_{j=0}^{\lfloor \frac{k-k'}{2} \rfloor} \binom{k}{j} + \sum_{j=\lceil \frac{k+k'+1}{2} \rceil}^{k} \binom{k}{j} \right] \ge 2^{-k'} .$$

Proof. The proof follows by induction over $k \ge k'$. A speciality here is that in the induction step we assume the statement for k and prove it for $k+2$, which is sufficient when the base case is verified for $k = k'$ and $k = k'+1$.

For $k = k'$ and $k = k'+1$ we have $2^{-k'}\binom{k}{0}$ and $2^{-(k'+1)}[\binom{k'}{0} + \binom{k'+1}{k'+1}]$ which proves the inequality. (We even have equality.)

For the induction step from k to $k+2$ where $k \ge k'$, via Pascal's rule, as well as using $\binom{k}{\lfloor \frac{k+2-k'}{2} \rfloor} \ge \binom{k}{\lfloor \frac{k-k'}{2} \rfloor}$, we obtain

$$\sum_{j=0}^{\lfloor \frac{k+2-k'}{2} \rfloor} \binom{k+2}{j} = 4 \sum_{j=0}^{\lfloor \frac{k-k'}{2} \rfloor - 1} \binom{k}{j} + 3 \binom{k}{\lfloor \frac{k-k'}{2} \rfloor} + \binom{k}{\lfloor \frac{k+2-k'}{2} \rfloor} \ge 4 \sum_{j=0}^{\lfloor \frac{k-k'}{2} \rfloor} \binom{k}{j}.$$

Similarly, with $\binom{k}{\lfloor \frac{k+k'+1}{2} \rfloor} \ge \binom{k}{\lfloor \frac{k+k'+3}{2} \rfloor}$, one can show

$$\sum_{j=\lceil \frac{k+k'+3}{2} \rceil}^{k+2} \binom{k+2}{j} \ge 4 \sum_{j=\lceil \frac{k+k'+1}{2} \rceil}^{k} \binom{k}{j}.$$

Now, by the induction hypothesis the assertion is proven. $\square$

Acknowledgements

The authors wish to express their gratitude to Erich Novak for many detailed hints and discussions during the work on this paper. We also wish to thank Glenn Byrenheidt, Stefan Heinrich, Lutz Kämmerer, David Krieg, and Mario Ullrich for their advice. Daniel Rudolf gratefully acknowledges support of the Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences (Volkswagen Foundation) and the Campus laboratory AIMS.

References

[1] R.A. Adams, J.J.F. Fournier. *Sobolev Spaces*, second edition. Academic Press, 2002.

[2] N. Alon, Y. Matias, M. Szegedy. The space complexity of approximating the frequency moments. *STOC '96 Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, 20–29, 1996.

[3] B. von Bahr, C. Esseen. Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics*, 36:299–303, 1965.

[4] N. S. Bakhvalov. On the approximate calculation of multiple integrals. *Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem.* 4:3–18, 1959, in Russian. English translation: *Journal of Complexity*, 31(4):502–516, 2015.

[5] O. Catoni. Challenging the empirical mean and empirical variance: A deviation study. *Annales de lInstitut Henri Poincaré – Probabilités et Statistiques*, 48(4):1148–1185, 2012.

[6] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems.* SIAM, 1978.

[7] D. Dũng, V.N. Temlyakov, T. Ullrich. *Hyperbolic Cross Approximation.* Advanced Courses in Mathematics. CRM Barcelona. Birkhuser/Springer, to appear.

[8] L. Gajek, W. Niemiro, P. Pokarowski. Optimal Monte Carlo integration with fixed relative precision. *Journal of Complexity*, 29:4–26, 2013.

[9] S. Heinrich. Random approximation in numerical analysis. *Proceedings of the Conference "Functional Analysis" Essen*, pp. 123-171, 1994.

[10] S. Heinrich. Randomized approximation of Sobolev embeddings. *Proceedings of the MCQMC 2006*, 445–459, 2008.

[11] F.J. Hickernell, L. Jiang, Y.W. Liu, A. Owen. Guaranteed conservative fixed width confidence intervals via Monte Carlo sampling. Proceedings of the MCQMC 2012, Springer, 2013.

[12] M. Huber. A Bernoulli mean estimate with known relative error distribution. *Random Structures and Algorithms*, 50:173–182, 2017.

[13] M. Huber. An optimal (ϵ, δ)-approximation scheme for the mean of random variables with bounded relative variance. Available on arXiv:1706.01478v1 [stat.CO], 2017.

[14] M. Jerrum, L. Valiant, V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Sciences*, 43:169–188, 1986.

[15] D. Krieg, E. Novak. A universal algorithm for univariate integration. *Foundations of Computational Mathematics*, 17:895-916, 2017.

[16] R.J. Kunsch, E. Novak, D. Rudolf. Solvable integration problems and optimal sample size selection. Preprint available on arXiv:1805.08637 [math.NA], 2018.

[17] J. Matoušek, J. Vondrak. *The Probabilistic Method*, manuscript.

[18] V.K. Nguyen, M. Ullrich, T. Ullrich. Change of variable in spaces of mixed smoothness and numerical integration of multivariate functions on the unit cube. *Constructive Approximation*, 46:69–108, 2017.

[19] W. Niemiro, P. Pokarowski. Fixed precision MCMC estimation by median of product of averages. *Journal of Applied Probability*, 46:309–329, 2009.

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

[21] E. Novak, H. Triebel. Function spaces in Lipschitz domains and optimal rates of convergence for sampling. *Constructive Approximation*, 23:325–350, 2006.

[22] E. Novak, H. Woźniakowski. *Tractability of Multivariate Problems*, Volume I, Linear Information. European Mathematical Society, 2008.

[23] E. Novak, H. Woźniakowski. *Tractability of Multivariate Problems*, Volume II, Standard Information for Functionals. European Mathematical Society, 2010.

[24] E. Novak, H. Woźniakowski. *Tractability of Multivariate Problems*, Volume III, Standard Information for Operators. European Mathematical Society, 2012.

[25] D. Rudolf, N. Schweizer Error bounds of MCMC for functions with unbounded stationary variance. *Statistics & Probability Letters*, 99:6–12, 2015.

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

[27] H. Triebel. *Theory of Function Spaces*. Birkhäuser, 1983.

[28] M. Ullrich. A Monte Carlo method for integration of multivariate smooth functions. *SIAM Journal on Numerical Analysis*, 55(3):1188-1200, 2017.