



Optimal Monte Carlo Methods for L^2 -Approximation

David Krieg¹

Received: 12 May 2017 / Revised: 23 February 2018 / Accepted: 13 March 2018 /
Published online: 6 April 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract We construct Monte Carlo methods for the L^2 -approximation in Hilbert spaces of multivariate functions sampling not more than n function values of the target function. Their errors catch up with the rate of convergence and the preasymptotic behavior of the error of any algorithm sampling n pieces of arbitrary linear information, including function values.

Keywords Approximation of multivariate functions · Monte Carlo methods · Optimal order of convergence · Preasymptotic estimates · Multivariate integration

Mathematics Subject Classification 41A25 · 41A63 · 65C05 · 65D15 · 65D30 · 68Q25 · 65Y20

1 Introduction

Assume we want to approximate an unknown real or complex-valued function on a set D based on a finite number n of function values that may be evaluated at randomly and adaptively chosen points. In general, these function values do not determine the function uniquely, and so we cannot expect our approximation to be correct. We make an approximation error, which we measure in the space $L^2(D, \mathcal{A}, \mu)$ of quadratically integrable functions on D with respect to an arbitrary measure μ . In order to make any meaningful statement regarding this error, we need to have additional a priori

Communicated by Wolfgang Dahmen.

✉ David Krieg
david.krieg@uni-jena.de

¹ Mathematisches Institut, Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany

knowledge of the unknown function. Here, we assume structural knowledge of the form that it is contained in the unit ball F_\circ of a Hilbert space F that is compactly embedded in $L^2(D, \mathcal{A}, \mu)$. For instance, it may be bounded with respect to some Sobolev norm on a compact manifold D . The error of the randomized algorithm or Monte Carlo method A_n is the quantity

$$e^{\text{ran}}(A_n) = \sup_{f \in F_\circ} \left(\mathbb{E} \int_D |f - A_n(f)|^2 d\mu \right)^{1/2}.$$

The error of an optimal randomized algorithm that asks for at most n function values is denoted by

$$e(n) = \inf_{A_n} e^{\text{ran}}(A_n).$$

While it seems impossible to provide such algorithms, the optimal deterministic algorithm evaluating n arbitrary linear functionals is well known. It is given by the orthogonal projection P_n onto the span of the first n functions in the singular value decomposition of the embedding $T : F \hookrightarrow L^2$. Its worst case error is the $(n + 1)$ -st largest singular value or approximation number $\sigma(n + 1)$ of that embedding, the square root of the $(n + 1)$ -st largest eigenvalue of the operator $W = T^*T$.

The algorithm P_n asks for the first n coefficients of f with respect to the singular value decomposition of the embedding T . In most applications, however, it is not possible to sample these coefficients, and we may only make use of function values. This leads to the following questions:

- How does the error $e(n)$ of optimal randomized algorithms using n function values compare to the error $\sigma(n + 1)$ of the orthogonal projection P_n ?
- If possible, find a randomized algorithm A_n whose error is close to $\sigma(n + 1)$.

These are not new questions in the fields of Monte Carlo methods and information-based complexity. There are several results for particular spaces F where $e(n)$ behaves similarly to the error of P_n . See, for instance, Traub, Wasilkowski and Woźniakowski [20], Mathé [13] and Heinrich [5]. Results by Cohen, Davenport and Leviatan [2] and Cohen and Migliorati [3] contain a similar message, see Remark 3. In 1992, Novak [16] proved that

$$e(n) \geq \frac{\sigma(2n)}{\sqrt{2}}$$

holds for arbitrary spaces F . This means that optimal randomized algorithms using n function values are never much better than the orthogonal projection P_n . On the other hand, Wasilkowski and Woźniakowski [23] proved in 2006 that

$$\sigma(n) \preccurlyeq n^{-p}(\ln n)^q \quad \Rightarrow \quad e(n) \preccurlyeq n^{-p}(\ln n)^q(\ln \ln n)^{p+1/2}$$

for all $p > 0$ and $q \geq 0$. Here, we write $x_n \preccurlyeq y_n$ if there is some $C > 0$ and $n_0 \in \mathbb{N}$ such that $x_n \leq C y_n$ for all $n \geq n_0$. If $x_n \preccurlyeq y_n$ and $y_n \preccurlyeq x_n$, we write $x_n \asymp y_n$. This

means that optimal randomized algorithms using function values are always almost as good as the orthogonal projection P_n . The proof of this result is constructive. It raises the question whether the additional power of the double logarithm is necessary or not. In fact, Novak and Woźniakowski showed in 2012 that this is not the case for $q = 0$; that is,

$$\sigma(n) \preceq n^{-p} \Rightarrow e(n) \preceq n^{-p}$$

for all $p > 0$. The proof of this result, however, is not constructive. Both proofs can be found in their monograph [18, Chapter 22]. In the present paper, we prove the corresponding statement for $q > 0$. More generally, we consider upper bounds with the following property. We say that the sequence $L : \mathbb{N} \rightarrow (0, \infty)$ is *regularly decreasing* if there is some $r \geq 0$ such that

$$L(m) \geq 2^{-r} L(n) \text{ whenever } n \leq m \leq 2n. \tag{1}$$

If there is some $n_0 \in \mathbb{N}$ such that $L(n)$ is nonincreasing for $n \geq n_0$, this is equivalent to $L(2n) \asymp L(n)$. Property (1) is satisfied if $L(n)n^r$ is nondecreasing. The sequence

$$L(n) = n^{-p} (1 + \log_2 n)^q$$

is regularly decreasing for any $p > 0$ and $q \geq 0$. It satisfies (1) for $r = p$. Another example is

$$L(n) = (1 + \log_2 n)^{-q}$$

for any $q > 0$, which satisfies (1) for $r = q$. The sequence is not regularly decreasing if it decays exponentially or has huge jumps. We obtain the following result.

Theorem 1 *If $L : \mathbb{N} \rightarrow (0, \infty)$ is regularly decreasing, then*

$$\sigma(n) \preceq L(n) \Rightarrow e(n) \preceq L(n).$$

This solves Open Problem 99 as posed by Novak and Woźniakowski in [18]. One problem with this result is that it does not provide any algorithm; it only states the existence of good algorithms. Another problem is that the error bound is only asymptotic. The preasymptotic behavior of $e(n)$ may, however, be very different from its asymptotic behavior. This is typically the case if the set D is a domain in high-dimensional Euclidean space.

These problems are tackled by Theorem 2. In Sect. 3, we provide a randomized algorithm A_n^r for any $n \in \mathbb{N}$ and $r \geq 0$. This algorithm is a refinement of the algorithm proposed by Wasilkowski and Woźniakowski [23]. It asks for at most n function values and satisfies the following error bound.

Theorem 2 Assume that $L : \mathbb{N} \rightarrow (0, \infty)$ satisfies (1), and let $c_r = 2^{r[2r+3]+1}$.

$$\begin{array}{lll} \text{If} & \sigma(n) \leq L(n) & \text{for all } n \in \mathbb{N}, \\ \text{then} & e^{\text{ran}}(A_n^r) \leq c_r L(n) & \text{for all } n \in \mathbb{N}. \end{array}$$

The constant c_r only depends on the order r . If D is a domain in d -dimensional Euclidean space, this order is often independent of d or even strictly decreasing with d . See Sect. 3 for the definition of this algorithm and several examples.

We find that the error of randomized algorithms using n function values of the target function can get very close to the error of the orthogonal projection P_n and that this is achieved by the algorithm A_n^r .

In Sect. 4, we use these algorithms for the integration of functions f in F with respect to probability measures μ . We simply exploit the relation

$$\int_D f \, d\mu = \int_D A_n^r f \, d\mu + \int_D (f - A_n^r f) \, d\mu.$$

We compute the integral of $A_n^r f$ and use a direct simulation to approximate the integral of $(f - A_n^r f)$, which has a small variance. This technique is called variance reduction and widely used for Monte Carlo integration. See Heinrich [5, Theorem 5.3] for another example. Even if D is a high-dimensional domain, the resulting method can significantly improve on the error of a sole direct simulation for a relatively small number of samples.

These results are based on the a priori knowledge that our target function is contained in the unit ball of the space F . In Sect. 5, we discuss how this assumption can be weakened.

2 The Setting

Let (D, \mathcal{A}, μ) be a measure space and $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. The space $L^2 = L^2(D, \mathcal{A}, \mu)$ is the space of quadratically integrable \mathbb{K} -valued functions on (D, \mathcal{A}, μ) , equipped with the scalar product

$$\langle f, g \rangle_2 = \int_D f \cdot \bar{g} \, d\mu.$$

Let F be a second Hilbert space and F_\circ be its unit ball. We assume that F is a subset of L^2 and that

$$T : F \rightarrow L^2, \quad Tf = f,$$

is compact. With the embedding T , we associate a positive semi-definite and compact operator $W = T^*T$ on the space F . By the spectral theorem, there is a (possibly finite) orthogonal basis $\mathcal{B} = \{b_1, b_2, \dots\}$ of F , consisting of eigenvectors corresponding to a nonincreasing zero sequence $(\lambda_n)_{n \in \mathbb{N}}$ of eigenvalues of W . Let N be the cardinality of

\mathcal{B} . One can easily check that \mathcal{B} is orthogonal in L^2 , as well. We take the eigenvectors b_n to be normalized in L^2 . We call this basis the singular value decomposition of T .¹ The number $\sigma(n) = \sqrt{\lambda_n}$ is called its n -th singular value or approximation number.

The worst case error of a deterministic algorithm $A : F \rightarrow L^2$ is the quantity

$$e^{\det}(A) = \sup_{f \in F_\circ} \|f - A(f)\|_2.$$

The worst case error of a measurable randomized algorithm

$$A : F \times \Omega \rightarrow L^2, \quad (f, \omega) \rightarrow A^\omega(f),$$

where Ω is the sample space of some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is the quantity

$$e^{\text{ran}}(A) = \sup_{f \in F_\circ} \left(\mathbb{E}_\omega \|f - A^\omega(f)\|_2^2 \right)^{1/2}.$$

We usually skip the ω in the notation. See Novak and Woźniakowski [17, Chapter 4] for a precise definition of such algorithms. We furthermore define the following minimal worst case errors within certain classes of algorithms.

The quantity

$$e^{\det}(n, T, \Lambda^{\text{all}}) = \inf_{A \in \mathcal{A}_n^{\det, \text{all}}} e^{\det}(A)$$

is the minimal worst case error within the class $\mathcal{A}_n^{\det, \text{all}}$ of all deterministic algorithms evaluating at most n linear functionals of the input function.

The quantity

$$e^{\text{ran}}(n, T, \Lambda^{\text{all}}) = \inf_{A \in \mathcal{A}_n^{\text{ran}, \text{all}}} e^{\text{ran}}(A)$$

is the minimal worst case error within the class $\mathcal{A}_n^{\text{ran}, \text{all}}$ of all measurable randomized algorithms evaluating at most n linear functionals.

The quantity

$$e^{\det}(n, T, \Lambda^{\text{std}}) = \inf_{A \in \mathcal{A}_n^{\det, \text{std}}} e^{\det}(A)$$

is the minimal worst case error within the class $\mathcal{A}_n^{\det, \text{std}}$ of all deterministic algorithms evaluating at most n function values of the input function.

The quantity

$$e(n) = e^{\text{ran}}(n, T, \Lambda^{\text{std}}) = \inf_{A \in \mathcal{A}_n^{\text{ran}, \text{std}}} e^{\text{ran}}(A)$$

¹ This term is more commonly used to refer to the representation $Tf = \sum_{b \in \mathcal{B}} \langle f, b \rangle Tb$ of the compact operator. Here, the altered terminology shall ease the notation.

finally is the minimal worst case error within the class $\mathcal{A}_n^{\text{ran, std}}$ of all measurable randomized algorithms evaluating at most n function values. This is the error to be analyzed. It was proved by Novak [16] that

$$e^{\text{ran}}(n, T, \Lambda^{\text{std}}) \geq e^{\text{ran}}(n, T, \Lambda^{\text{all}}) \geq \frac{1}{\sqrt{2}} e^{\text{det}}(2n - 1, T, \Lambda^{\text{all}}). \tag{2}$$

The error $e^{\text{det}}(n, T, \Lambda^{\text{all}})$ is known to coincide with $\sigma(n + 1)$. We refer to Novak and Woźniakowski [17, Section 4.2.3]. The infimum is attained for the nonadaptive linear algorithm

$$P_n : F \rightarrow L^2, \quad P_n(f) = \sum_{k=1}^{n \wedge N} \langle f, b_k \rangle_2 b_k.$$

Here, $\log_2 x$ denotes the logarithm of $x > 0$ in base 2, whereas $\ln x$ denotes its natural logarithm. The minimum of $a \in \mathbb{R}$ and $b \in \mathbb{R}$ is denoted by $a \wedge b$. Recall that we write $x_n \preccurlyeq y_n$ if there is a positive constant C and some $n_0 \in \mathbb{N}$ such that $x_n \leq C y_n$ for all $n \geq n_0$. We write $x_n \asymp y_n$ if $x_n \preccurlyeq y_n$ and $y_n \preccurlyeq x_n$.

3 A Method for Multivariate Approximation

Let us keep the notation of the previous section. For any $m \in \mathbb{N}$ with $m \leq N$, we define

$$u_m = \frac{1}{m} \sum_{j=1}^m |b_j|^2.$$

This is a probability density with respect to μ . We consider the probability measure

$$\mu_m : \mathcal{A} \rightarrow [0, 1], \quad \mu_m(E) = \int_E u_m \, d\mu$$

on (D, \mathcal{A}) . In view of optimal algorithms in $\mathcal{A}_n^{\text{det, all}}$, we introduce the following family of algorithms in $\mathcal{A}_n^{\text{ran, std}}$.

Algorithm Let $\mathbf{n} = (n_1, n_2, \dots)$ and $\mathbf{m} = (m_1, m_2, \dots)$ be sequences of nonnegative integers such that \mathbf{m} is nondecreasing and bounded above by $N = |\mathcal{B}|$. We define the algorithms $M_{\mathbf{n}, \mathbf{m}}^{(k)} : L^2 \rightarrow L^2$ for $k \in \mathbb{N}_0$ as follows.

- Set $M_{\mathbf{n}, \mathbf{m}}^{(0)} = 0$.
- For $k \geq 1$ and $f \in L^2$, let $X_1^{(k)}, \dots, X_{n_k}^{(k)}$ be random variables with distribution μ_{m_k} that are each independent of all the other random variables, and set

$$M_{\mathbf{n}, \mathbf{m}}^{(k)} f = M_{\mathbf{n}, \mathbf{m}}^{(k-1)} f + \sum_{j=1}^{m_k} \left[\frac{1}{n_k} \sum_{i=1}^{n_k} \frac{(f - M_{\mathbf{n}, \mathbf{m}}^{(k-1)} f) \bar{b}_j}{u_{m_k}} (X_i^{(k)}) \right] b_j.$$

Note that the expectation of each term in the inner sum is $\langle f - M_{n,m}^{(k-1)} f, b_j \rangle_2$. The algorithm $M_{n,m}^{(k)}$ hence approximates f in k steps. In the first step, n_1 function values of f are used for standard Monte Carlo type approximations of its m_1 leading coefficients with respect to the orthonormal system \mathcal{B} . In the second step, n_2 values of the residue are used for standard Monte Carlo type approximations of its m_2 leading coefficients and so on. In total, $M_{n,m}^{(k)}$ uses $\sum_{j=1}^k n_j$ function values of f . The total number of approximated coefficients is m_k .

Algorithms of this type have already been studied by Wasilkowski and Woźniakowski in [23]. The simple but crucial difference with the above algorithms is the variable number n_j of nodes in each approximation step. Note that this stepwise approximation is similar to several multilevel Monte Carlo methods as introduced by Heinrich in 1998, see [4].

The benefit from the k th step is controlled by m_k and n_k as follows.

Lemma 1 *For all nondecreasing sequences \mathbf{n} and \mathbf{m} of nonnegative integers and all $k \in \mathbb{N}$, we have*

$$\sigma(m_k + 1)^2 \leq e^{\text{ran}} \left(M_{n,m}^{(k)} \right)^2 \leq \frac{m_k}{n_k} e^{\text{ran}} \left(M_{n,m}^{(k-1)} \right)^2 + \sigma(m_k + 1)^2.$$

Lemma 1 corresponds to Theorem 22.14 by Novak and Woźniakowski [18]. The setting of the present paper is slightly more general, but the proof is the same. Since Lemma 1 is essential for the following investigation, I present the proof.

Proof The lower bound holds true, since $M_{n,m}^{(k)}(b_{m_k+1})$ is perpendicular to b_{m_k+1} . To prove the upper bound, let $f \in F_\circ$. By \mathbb{E}_I we denote the expectation with respect to the random variables $X_i^{(j)}$ for $j \in I$ and $i = 1 \dots n_j$. We need to estimate

$$\mathbb{E}_{\{1\dots k\}} \left\| f - M_{n,m}^{(k)} f \right\|_2^2 = \sum_{j=1}^N \mathbb{E}_{\{1\dots k\}} \left| \langle f - M_{n,m}^{(k)} f, b_j \rangle_2 \right|^2.$$

On the one hand, we have

$$\begin{aligned} \sum_{j=m_k+1}^N \mathbb{E}_{\{1\dots k\}} \left| \langle f - M_{n,m}^{(k)} f, b_j \rangle_2 \right|^2 &= \sum_{j=m_k+1}^N \left| \langle f, b_j \rangle_2 \right|^2 = \sum_{j=m_k+1}^N \left| \langle f, W b_j \rangle_F \right|^2 \\ &= \sum_{j=m_k+1}^N \left| \langle f, \sigma(j) b_j \rangle_F \right|^2 \sigma(j)^2 \leq \sigma(m_k + 1)^2 \|f\|_F^2 \leq \sigma(m_k + 1)^2. \end{aligned}$$

We use the abbreviation

$$g_j = \frac{\left(f - M_{n,m}^{(k-1)} f \right) \overline{b_j}}{u_{m_k}}$$

for each $j \leq m_k$. Note that $u_{m_k} = 0$ implies $b_j = 0$, and we set $g_j = 0$ in this case. We then obtain, on the other hand, for each $j \leq m_k$, that

$$\begin{aligned} \mathbb{E}_{\{k\}} \left| \left\langle f - M_{n,m}^{(k)} f, b_j \right\rangle_2 \right|^2 &= \mathbb{E}_{\{k\}} \left| \left\langle f - M_{n,m}^{(k-1)} f, b_j \right\rangle_2 - \frac{1}{n_k} \sum_{i=1}^{n_k} g_j \left(X_i^{(k)} \right) \right|^2 \\ &= \mathbb{E}_{\{k\}} \left| \int_D g_j(x) \, d\mu_{m_k}(x) - \frac{1}{n_k} \sum_{i=1}^{n_k} g_j \left(X_i^{(k)} \right) \right|^2 \\ &\leq \frac{1}{n_k} \int_D |g_j(x)|^2 \, d\mu_{m_k}(x) = \frac{1}{n_k} \int_D |g_j(x)|^2 u_{m_k}(x) \, d\mu(x), \end{aligned}$$

and hence

$$\begin{aligned} \sum_{j=1}^{m_k} \mathbb{E}_{\{k\}} \left| \left\langle f - M_{n,m}^{(k)} f, b_j \right\rangle_2 \right|^2 &\leq \frac{1}{n_k} \int_D \sum_{j=1}^{m_k} |g_j(x)|^2 u_{m_k}(x) \, d\mu(x) \\ &= \frac{m_k}{n_k} \int_D \left| \left(f - M_{n,m}^{(k-1)} f \right) (x) \right|^2 \, d\mu(x) = \frac{m_k}{n_k} \left\| f - M_{n,m}^{(k-1)} f \right\|_2^2. \end{aligned}$$

With Fubini’s theorem, this yields that

$$\mathbb{E}_{\{1\dots k\}} \left\| f - M_{n,m}^{(k)} f \right\|_2^2 \leq \frac{m_k}{n_k} \mathbb{E}_{\{1\dots k-1\}} \left\| f - M_{n,m}^{(k-1)} f \right\|_2^2 + \sigma(m_k + 1)^2,$$

and the upper bound is proved. □

We now define the algorithm of Theorem 2. We consider such algorithms $M_{n,m}^{(k)}$, where the number of nodes n_j is doubled in each step and the ratio $\frac{m_j}{n_j}$ of approximated coefficients and computed function values is constant, say $2^{-\ell}$. This way, the total number m_k of approximated coefficients is linear in the total number n of computed function values. This is necessary to achieve an error of the same order as with optimal algorithms using arbitrary linear information, which precisely compute the first n coefficients. The algorithms by Wasilkowski and Woźniakowski [23] do not have this property. If the ratio is small enough, Lemma 1 ensures that $M_{n,m}^{(k)}$ inherits optimal error bounds from $M_{n,m}^{(k-1)}$.

Algorithm Given $r \geq 0$, we set $\ell_r = \lceil 2r + 1 \rceil$ and define the sequences n and m by

$$n_j = \begin{cases} 0, & \text{for } j \leq \ell_r, \\ 2^{j-1}, & \text{for } j > \ell_r, \end{cases} \quad m_j = \begin{cases} 0, & \text{for } j \leq \ell_r, \\ 2^{j-1-\ell_r} \wedge N, & \text{for } j > \ell_r. \end{cases}$$

For $n \in \mathbb{N}$, we choose $k \in \mathbb{N}_0$ such that $2^k \leq n < 2^{k+1}$ and set

$$A_n^r = M_{n,m}^{(k)}.$$

The algorithm A_n^r obviously performs less than n function evaluations.

Proof of Theorem 2 Let n and m be defined as above and $k \in \mathbb{N}_0$. We first show that

$$e^{\text{ran}} \left(M_{n,m}^{(k)} \right) \leq \bar{c}_r L(2^k), \tag{3}$$

where $\bar{c}_r = 2^{r(\ell_r+1)+1}$. We use induction on k . If $k \leq \ell_r$, we have $M_{n,m}^{(k)} = 0$ and

$$e^{\text{ran}} \left(M_{n,m}^{(k)} \right) = \sigma(1) \leq L(1) \leq 2^{rk} L(2^k) \leq \bar{c}_r L(2^k).$$

For $k > \ell_r$, we inductively obtain with Lemma 1 that

$$\begin{aligned} e^{\text{ran}} \left(M_{n,m}^{(k)} \right)^2 &\leq 2^{-\ell_r} e^{\text{ran}} \left(M_{n,m}^{(k-1)} \right)^2 + \sigma(m_k + 1)^2 \\ &\leq 2^{-\ell_r} \bar{c}_r^2 L \left(2^{k-1} \right)^2 + L \left(2^{k-\ell_r-1} \right)^2 \\ &\leq 2^{-\ell_r} \bar{c}_r^2 2^{2r} L(2^k)^2 + 2^{2r(\ell_r+1)} L(2^k)^2 \\ &= \left(2^{2r-\ell_r} + 2^{-2} \right) \bar{c}_r^2 L(2^k)^2, \end{aligned}$$

where the term in brackets is smaller than 1. This shows (3). For $n \in \mathbb{N}$, we choose $k \in \mathbb{N}_0$ with $2^k \leq n < 2^{k+1}$ and obtain

$$e^{\text{ran}} \left(A_n^r \right) = e^{\text{ran}} \left(M_{n,m}^{(k)} \right) \leq \bar{c}_r L(2^k) \leq 2^r \bar{c}_r L(n) = c_r L(n),$$

as was to be proved. □

Note that Theorem 1 is a direct consequence of Theorem 2. Of course, the best possible upper bound for $\sigma(n)$ is $\sigma(n)$ itself. If we combine Theorem 1 for $L(n) = \sigma(n)$ with Novak’s lower bound (2), we obtain the following statement on the order of convergence.

Corollary 1 *Assume that $\sigma(2n) \asymp \sigma(n)$. Then*

$$e^{\text{ran}} \left(n, F \hookrightarrow L^2, \Lambda^{\text{std}} \right) \asymp e^{\text{ran}} \left(n, F \hookrightarrow L^2, \Lambda^{\text{all}} \right) \asymp e^{\text{det}} \left(n, F \hookrightarrow L^2, \Lambda^{\text{all}} \right).$$

Note that the error $e^{\text{det}} \left(n, F \hookrightarrow L^2, \Lambda^{\text{std}} \right)$ of optimal deterministic algorithms based on function values may perform much worse, as shown by Hinrichs, Novak and Vybíral [7], see also Novak and Woźniakowski [18, Section 26.6.1]. It is a very interesting question whether the condition on the decay of the singular values can be relaxed. Note that we use this condition both to prove the upper and the lower bound of Corollary 1. On the other hand, if we combine Theorem 2 for $L(n) = \sigma(n)$ and the lower bound (2), we obtain the following optimality result.

Corollary 2 *Assume that there is some $r \geq 0$ such that $\sigma(2n) \geq 2^{-r} \sigma(n)$ holds for all $n \in \mathbb{N}$. We set $\tilde{c}_r = 2^{r \lceil 2r+4 \rceil + 3/2}$. Then we have*

$$e^{\text{ran}} \left(A_n^r \right) \leq \tilde{c}_r e^{\text{ran}} \left(n, T, \Lambda^{\text{std}} \right) \quad \text{for all } n \in \mathbb{N}.$$

Let us now consider some examples. In each example, we first discuss the order of convergence of $e^{\text{ran}}(n, F \hookrightarrow L^2, \Lambda^{\text{std}})$. We then talk about explicit upper bounds.

Example 1 (Approximation of mixed order Sobolev functions on the torus). Let D be the d -dimensional torus \mathbb{T}^d , represented by the unit cube $[0, 1]^d$, where opposite faces are identified. Let \mathcal{A} be the Borel σ -algebra on \mathbb{T}^d and μ the Lebesgue measure. Let F be the Sobolev space of complex-valued functions on D with dominating mixed smoothness $r \in \mathbb{N}$, equipped with the scalar product

$$\langle f, g \rangle_F = \sum_{\|\alpha\|_\infty \leq r} \langle D^\alpha f, D^\alpha g \rangle_2. \tag{4}$$

We know that

$$e^{\text{det}}(n, F \hookrightarrow L^2, \Lambda^{\text{all}}) \asymp n^{-r} \ln^{r(d-1)} n.$$

This classical result goes back to Babenko [1] and Mityagin [14]. Corollary 1 yields

$$e^{\text{ran}}(n, F \hookrightarrow L^2, \Lambda^{\text{std}}) \asymp n^{-r} \ln^{r(d-1)} n.$$

This is a new result. The optimal order is achieved by the algorithm A'_n and the author does not know of any other algorithm with this property. It is still an open problem whether the same rate can be achieved with deterministic algorithms based on function values. So far, it is only known that

$$n^{-r} \ln^{r(d-1)} n \preceq e^{\text{det}}(n, F \hookrightarrow L^2, \Lambda^{\text{std}}) \preceq n^{-r} \ln^{(r+1/2)(d-1)} n.$$

The upper bound is achieved by Smolyak’s algorithm, see Sickel and Ullrich [19].

We now turn to explicit estimates. We know that there is some $C_{r,d} > 0$ such that

$$e^{\text{ran}}(n, F \hookrightarrow L^2, \Lambda^{\text{std}}) \leq C_{r,d} n^{-r} \ln^{r(d-1)} n \quad \text{for all } n \geq 2. \tag{5}$$

This upper bound is optimal as n tends to infinity. However, it is not useful to describe the error numbers for small values of n . Simple calculus shows that the right-hand side in (5) is increasing for $n \leq e^{d-1}$. The error numbers, on the other hand, are decreasing. Moreover, the right-hand side attains its minimum for $n = 2$ if restricted to $n \leq (d - 1)^{d-1}$ and is hence larger than $e^{\text{ran}}(2, F \hookrightarrow L^2, \Lambda^{\text{std}})$. This means that the trivial upper bound

$$e^{\text{ran}}(n, F \hookrightarrow L^2, \Lambda^{\text{std}}) \leq e^{\text{ran}}(2, F \hookrightarrow L^2, \Lambda^{\text{std}}) \quad \text{for all } n \geq 2$$

is better than (5) for all $n \leq (d - 1)^{d-1}$ and regardless of the value of $C_{r,d}$. For these reasons, it is important to consider different error bounds, if the dimension d is large.

See also the paper of Kühn, Sickel, and Ullrich [12]. Based on this paper, it is shown by the author [9] that

$$\sigma(n) \leq (2/n)^p \quad \text{for all } n \in \mathbb{N}, \quad \text{if } p = \frac{r}{2 + \ln d}.$$

We obtain with Theorem 2 that

$$e^{\text{ran}}(A_n^p) \leq 2 \cdot \left(2^{\lceil 2p+4 \rceil} / n\right)^p \quad \text{for } n \in \mathbb{N}. \tag{6}$$

Example 2 (Approximation of mixed order Sobolev functions on the cube). Now, let D be the d -dimensional unit cube $[0, 1]^d$ with the induced topology, and let \mathcal{A} be the Borel σ -algebra and μ the Lebesgue measure. Let F be the Sobolev space of complex-valued functions on $[0, 1]^d$ with dominating mixed smoothness $r \in \mathbb{N}$, equipped with the scalar product (4). Just like on the torus, we have

$$e^{\text{ran}}\left(n, F \hookrightarrow L^2, \Lambda^{\text{std}}\right) \asymp e^{\det}\left(n, F \hookrightarrow L^2, \Lambda^{\text{all}}\right) \asymp n^{-r} \ln^{r(d-1)} n,$$

where the optimal rate is achieved by A_n^r . Like in Example 1, the corresponding upper bounds are bad for $n \leq (d - 1)^{d-1}$. In this range, we need different estimates for the approximation numbers. It is known that

$$\sigma(n) \leq (2/n)^p \quad \text{for } n \in \mathbb{N}, \quad \text{if } p = \frac{1.1929}{2 + \ln d}.$$

This estimate cannot be improved significantly for $n \leq 2^d$, even if $r = \infty$. See the author’s paper [9] for more details. With Theorem 2, we obtain the upper bound

$$e^{\text{ran}}(A_n^p) \leq 2 \cdot (2^6/n)^p \quad \text{for } n \in \mathbb{N}.$$

Example 3 (Approximation in tensor product spaces). This example is more general than the previous ones. By $H_1 \otimes H_2$ we denote the tensor product of two Hilbert spaces H_1 and H_2 . For $j = 1 \dots d$, let $(D_j, \mathcal{A}_j, \nu_j)$ be a σ -finite measure space and F_j be a Hilbert space of \mathbb{K} -valued functions that is compactly embedded in $L^2(D_j, \mathcal{A}_j, \nu_j)$. The σ -finity of the measure spaces ensures that

$$L^2(D_1, \mathcal{A}_1, \nu_1) \otimes \dots \otimes L^2(D_d, \mathcal{A}_d, \nu_d) = L^2(D, \mathcal{A}, \mu),$$

where D is the Cartesian product of the sets D_j and μ is the unique product measure of the measures ν_j on the tensor product \mathcal{A} of the σ -algebras \mathcal{A}_j . The tensor product space

$$F = F_1 \otimes \dots \otimes F_d$$

is compactly embedded in $L^2(D, \mathcal{A}, \mu)$. Assuming that the approximation numbers of the univariate embeddings $F_j \hookrightarrow L^2(D_j, \mathcal{A}_j, \nu_j)$ are of polynomial decay, that is,

$$e^{\det} \left(n, F_j \hookrightarrow L^2(D_j, \mathcal{A}_j, \nu_j), \Lambda^{\text{all}} \right) \asymp n^{-r_j}$$

for some $r_j > 0$, it can be derived from Mityagin [14] and Nikol’skaya [15] that

$$e^{\det} \left(n, F \hookrightarrow L^2(D, \mathcal{A}, \mu), \Lambda^{\text{all}} \right) \asymp n^{-r} \ln^{r(d_0-1)} n,$$

where r is the minimum among all numbers r_j and d_0 is its multiplicity. Corollary 1 implies

$$e^{\text{ran}} \left(n, F \hookrightarrow L^2(D, \mathcal{A}, \mu), \Lambda^{\text{std}} \right) \asymp n^{-r} \ln^{r(d_0-1)} n,$$

where the optimal order is achieved by A_n^r . We do not discuss explicit estimates in this abstract setting.

Example 4 (Approximation of isotropic Sobolev functions on the torus). Let D again be the d -torus, this time represented by $[0, 2\pi]^d$. Let F be the Sobolev space of complex-valued functions on D with isotropic smoothness $r \in \mathbb{N}$, equipped with the scalar product

$$\langle f, g \rangle_F = \sum_{\|\alpha\|_1 \leq r} \langle D^\alpha f, D^\alpha g \rangle_2.$$

This example is not a tensor product problem. For this classical problem, it is known that

$$\begin{aligned} e^{\det} \left(n, F \hookrightarrow L^2, \Lambda^{\text{std}} \right) &\asymp e^{\text{ran}} \left(n, F \hookrightarrow L^2, \Lambda^{\text{std}} \right) \\ &\asymp e^{\det} \left(n, F \hookrightarrow L^2, \Lambda^{\text{all}} \right) \asymp e^{\text{ran}} \left(n, F \hookrightarrow L^2, \Lambda^{\text{all}} \right) \asymp n^{-r/d} \end{aligned}$$

for $r > d/2$. In the case $r \leq d/2$, where function values are only defined almost everywhere, the last three relations stay valid. See Jerome [8], Triebel [21], Mathé [13], and Heinrich [6]. For $n \leq 2^d$, however, the function $n^{-r/d}$ is not suited to describe the behavior of $\sigma(n)$. It has been proved by Kühn, Mayer, and Ullrich [11] that there are positive constants b_r and B_r that do not depend on d such that

$$b_r \left(\frac{\log_2(1 + d/\log_2 n)}{\log_2 n} \right)^{r/2} \leq \sigma(n) \leq B_r \left(\frac{\log_2(1 + d/\log_2 n)}{\log_2 n} \right)^{r/2} \tag{7}$$

for all $d > 1$ and $n \in \mathbb{N}$ with $d \leq n \leq 2^d$. If we apply Relation (2) and Theorem 2,² we obtain the existence of d -independent positive constants \tilde{b}_r and \tilde{B}_r such that

$$\tilde{b}_r \left(\frac{\log_2(1 + d/\log_2 n)}{\log_2 n} \right)^{r/2} \leq e(n) \leq \tilde{B}_r \left(\frac{\log_2(1 + d/\log_2 n)}{\log_2 n} \right)^{r/2}$$

for all $d > 1$ and $n \in \mathbb{N}$ with $d \leq n \leq 2^{d-1}$. This optimal behavior is achieved by the algorithm A_n^r .

Remark 1 (Implementation of these algorithms). The construction of the algorithms A_n^r is completely explicit. We are able to implement these algorithms if we know the singular value decomposition \mathcal{B} of the embedding $F \hookrightarrow L^2$ and if we are able to sample from the probability distributions μ_m . This task may be very hard. In Example 1 and 4, however, it is not. Here, \mathcal{B} is the Fourier basis of L^2 , and all the random variables are independent and uniformly distributed on the unit cube. Also the case of general tensor product spaces F and L^2 can be handled if the singular value decompositions \mathcal{B}_j of the univariate embeddings $F_j \hookrightarrow L^2(D_j, \mathcal{A}_j, \nu_j)$ are known. Then, the singular value decomposition of the embedding $F \hookrightarrow L^2$ is given by

$$\mathcal{B} = \left\{ b^{(1)} \otimes \dots \otimes b^{(d)} \mid b^{(j)} \in \mathcal{B}_j \text{ for } j = 1 \dots d \right\},$$

and the probability measure μ_m is the average of m product densities; that is,

$$\mu_m = \frac{1}{m} \sum_{i=1}^m \bigotimes_{j=1}^d \eta_{i,j},$$

where $d\eta_{i,j} = |b_{i,j}|^2 d\nu_j$ with some $b_{i,j} \in \mathcal{B}_j$. A random sample x from this distribution can be obtained as follows:

- (1) Get i from the uniform distribution on $\{1, \dots, m\}$.
- (2) Get x_1, \dots, x_d independently from the probability distributions $\eta_{i,1}, \dots, \eta_{i,d}$.

The second step can, for example, be done by rejection sampling, if the measures $\eta_{i,j}$ have a bounded Lebesgue density. This way, the total sampling costs are linear in d . Another method of sampling from μ_m is proposed by Cohen and Migliorati in [3, Section 5].

² We take $L(n)$ as the right-hand side in (7) for $d \leq n \leq 2^d$, $L(n) = L(2^d)$ for $n > 2^d$, and $L(n) = \max\{1, L(d)\}$ for $n < d$. Then $\sigma(n) \leq L(n)$ for $n \in \mathbb{N}$, and $L(n)n^r$ is nondecreasing.

4 A Method for Multivariate Integration

In this section, we require the measure μ to be finite. This ensures that the integral operator

$$I : F \rightarrow \mathbb{K}, \quad I(f) = \int_D f d\mu$$

is well defined and continuous on F . Let us assume that μ is a probability measure. We want to approximate $I(f)$ for an unknown function $f \in F_\circ$ by a randomized algorithm Q_n that evaluates at most n function values of f . The worst case error of Q_n is the quantity

$$e^{\text{ran}}(Q_n) = \sup_{f \in F_\circ} \left(\mathbb{E} |I(f) - Q_n(f)|^2 \right)^{1/2}.$$

The minimal worst case error among such algorithms is denoted by

$$e^{\text{ran}}(n, I, \Lambda^{\text{std}}) = \inf_{Q_n} e^{\text{ran}}(Q_n).$$

Like any method for L^2 -approximation, the algorithm A_n^r from Sect. 3 can also be used for numerical integration.

Algorithm For all $r > 0$, any $n \in \mathbb{N}$, and $f \in L^2$, let

$$Q_{2n}^r(f) = I(A_n^r f) + \frac{1}{n} \sum_{j=1}^n (f - A_n^r f)(X_j),$$

where X_1, \dots, X_n are random variables with distribution μ that are independent of each other and the random variables in A_n^r .

It is easy to verify that Q_{2n}^r is unbiased, evaluates at most $2n$ function values of f , and satisfies

$$\mathbb{E} |I(f) - Q_{2n}^r(f)|^2 \leq \frac{1}{n} \mathbb{E} \|f - A_n^r f\|_2^2$$

for each f in L^2 . We thus obtain the following corollary.

Corollary 3 Assume that $L : \mathbb{N} \rightarrow (0, \infty)$ satisfies (1), and let $c_r = 2^{r \lceil 2r+3 \rceil + 1}$.

$$\begin{array}{lll} \text{If} & \sigma(n) \leq L(n) & \text{for all } n \in \mathbb{N}, \\ \text{then} & e^{\text{ran}}(Q_{2n}^r) \leq c_r n^{-1/2} L(n) & \text{for all } n \in \mathbb{N}. \end{array}$$

In particular:

$$\begin{aligned}
 e^{\det} \left(n, F \hookrightarrow L^2, \Lambda^{\text{all}} \right) &\asymp n^{-p} \ln^q n \\
 \Rightarrow e^{\text{ran}} \left(n, I, \Lambda^{\text{std}} \right) &\asymp n^{-p-1/2} \ln^q n.
 \end{aligned}$$

The result on the order of convergence is quite general but not always optimal. An example is given by integration with respect to the Lebesgue measure μ on the Sobolev space F with dominating mixed smoothness r on the d -dimensional unit cube, as treated by Novak and the author [10] and Ullrich [22]. In this case, we have

$$\begin{aligned}
 e^{\det} \left(n, F \hookrightarrow L^2, \Lambda^{\text{all}} \right) &\asymp n^{-r} \ln^{r(d-1)} n, \\
 e^{\text{ran}} \left(n, I, \Lambda^{\text{std}} \right) &\asymp n^{-r-1/2}.
 \end{aligned}$$

The main strength of Corollary 3 is that it provides us with unbiased methods for high-dimensional integration achieving a small error with a modest number of function values.

Example 5 (Integration of mixed order Sobolev functions on the torus). As in Example 1, let F be the Sobolev space of dominating mixed smoothness r on the d -torus, and let μ be the Lebesgue measure. Among all randomized algorithms for multivariate integration in F , the randomized Frolov algorithm Q_n^* is known to have the optimal error rate. It is shown by Ullrich [22] that there is some constant $c > 2^d$ such that

$$e^{\text{ran}} \left(Q_n^* \right) \leq c n^{-r-1/2} \quad \text{for } n \in \mathbb{N}. \tag{8}$$

However, this estimate is trivial if n is not exponentially large in d . For smaller values of n , an error less than one is guaranteed by the direct simulation

$$S_n(f) = \frac{1}{n} \sum_{j=1}^n f(X_j),$$

with independent and uniformly distributed random variables X_j . It satisfies

$$e^{\text{ran}} \left(S_n \right) \leq n^{-1/2} \quad \text{for } n \in \mathbb{N}. \tag{9}$$

However, this error bound converges only slowly, as n tends to infinity. It does not reflect the smoothness of the integrands at all. The above method also guarantees nontrivial error bounds for smaller values of n but converges faster than S_n . Relation (6) immediately yields that

$$e^{\text{ran}} \left(Q_{2n}^p \right) \leq C n^{-p-1/2} \quad \text{for } n \in \mathbb{N}, \tag{10}$$

with $p = \frac{r}{2+\ln d}$ and $C = 2^{p\lceil 2p+4 \rceil + 1}$. For example, let $d = 500$ and $r = 8$. For one million function values, the estimate (8) for the Frolov algorithm is larger than one,

the estimate (9) for the direct simulation gives the error 10^{-3} , and the estimate (10) for our new algorithm gives an error smaller than $5 \cdot 10^{-7}$.

Remark 2 (Implementation of these algorithms). We are able to implement the algorithms Q_{2n}^r under the following assumptions:

- We are able to implement A_n^r . This issue is discussed in Remark 1.
- We know the integrals $I(b_j)$ of the eigenfunctions $b_j \in \mathcal{B}$ for all $j \leq 2^{-\ell_r} n$.
- We can sample from the probability distribution μ .

In the above example, the implementation is particularly easy, since \mathcal{B} is the Fourier basis and all the random variables are independent and uniformly distributed on the unit cube.

5 A Weaker Type of A Priori Knowledge

In the previous sections, we assumed that the target function f is contained in the unit ball of a Hilbert space F that is compactly embedded into L^2 , that is,

$$\|f\|_F \leq 1. \tag{11}$$

As we have seen in Sect. 2, the space F induces a nonincreasing sequence σ , the singular numbers

$$\sigma(1) \geq \sigma(2) \geq \dots > 0$$

of the embedding $F \hookrightarrow L^2$. This sequence is either finite or tends to zero. It also induces a nested sequence V of subspaces

$$V_0 \subset V_1 \subset V_2 \subset \dots \subset L^2, \quad \dim(V_m) = m,$$

where V_m is spanned by the first m elements of the singular value decomposition.

In turn, any such pair (σ, V) induces a Hilbert space F that is compactly embedded into L^2 . We choose b_m as an element of the orthogonal complement of V_{m-1} in V_m with $\|b_m\|_2 = 1$ and define F by its orthonormal basis $\{\sigma(1)b_1, \sigma(2)b_2, \dots\}$. It has the scalar product

$$\langle f, g \rangle_F = \sum \sigma(k)^{-2} \langle f, b_k \rangle_2 \overline{\langle g, b_k \rangle_2},$$

where we take the sum over the whole sequence σ . It is not hard to see that the correspondence between F and the pair (σ, V) is bijective up to the choice of the spaces V_m for which we have $\sigma(m + 1) = \sigma(m)$.

Let P_m denote the orthogonal projection onto V_m in L^2 . It is readily verified that our assumption (11) on the target function f implies that

$$\|f - P_m f\|_2^2 \leq \sigma(m + 1)^2 \quad \text{for all } m \in \mathbb{N}_0. \tag{12}$$

In general, however, (12) is strictly weaker than (11). For example, if $\sigma(k) = 1/k$ for $k \in \mathbb{N}$, the function

$$f = \sum (\sigma(k)^2 - \sigma(k + 1)^2)^{1/2} b_k$$

satisfies (12) but is not even contained in the space F . In Sect. 3, we constructed a randomized algorithm $A_n^r : L^2 \rightarrow V_m$ and proved upper bounds on the mean square error $\mathbb{E} \|f - A_n^r(f)\|_2^2$ for any f from (11). In fact, the same error bounds hold for any f from (12). We state this as Theorem 3.

Theorem 3 *Let (D, \mathcal{A}, μ) be a measure space and $L^2 = L^2(D, \mathcal{A}, \mu)$. For any $m \in \mathbb{N}_0$, let V_m be an m -dimensional subspace of L^2 such that $V_m \subset V_{m+1}$, and let $P_m : L^2 \rightarrow V_m$ be the orthogonal projection onto V_m . Assume that $f \in L^2$ satisfies*

$$\|f - P_m f\|_2^2 \leq \varepsilon(m) \quad \text{for all } m \in \mathbb{N}_0, \tag{13}$$

with some $\varepsilon : \mathbb{N}_0 \rightarrow (0, \infty)$. Then the randomized algorithm $Q_m : L^2 \rightarrow V_m$ as defined below satisfies

$$\mathbb{E} \|f - Q_m f\|_2^2 \leq 2\varepsilon(m)$$

for any $m = 2^k$ and $k \in \mathbb{N}_0$. The number of requested function values is at most

$$n(Q_m) \leq 4m \cdot \max_{0 \leq j \leq k} \left\lceil \frac{\varepsilon(\lfloor 2^{j-1} \rfloor)}{\varepsilon(2^j)} \right\rceil. \tag{14}$$

To define the algorithm Q_m , we choose b_n in the orthogonal complement of V_{n-1} in V_n with $\|b_n\|_2 = 1$ for all $n \in \mathbb{N}$. For $j \in \mathbb{N}$, we set

$$m_j = 2^{j-1} \quad \text{and} \quad n_j = 2^j \left\lceil \frac{\varepsilon(\lfloor 2^{j-2} \rfloor)}{\varepsilon(2^{j-1})} \right\rceil.$$

Then the method $M_{n,m}^{(k)} : L^2 \rightarrow V_{m_k}$ for $k \in \mathbb{N}_0$ can be defined as in Sect. 3. Given $m = 2^k$ for some $k \in \mathbb{N}_0$, we define $Q_m = M_{n,m}^{(k+1)} : L^2 \rightarrow V_m$.

Proof We only sketch the proof since it is very similar to the proof of Theorem 2. Just like in Lemma 1, we can show for any $k \in \mathbb{N}_0$ that

$$\mathbb{E} \left\| f - M_{n,m}^{(k+1)} f \right\|_2^2 \leq \frac{m_{k+1}}{n_{k+1}} \mathbb{E} \left\| f - M_{n,m}^{(k)} f \right\|_2^2 + \varepsilon(m_{k+1}).$$

The statement follows by induction on $k \in \mathbb{N}_0$. □

Note that we did not impose any condition on the upper bound $\varepsilon : \mathbb{N}_0 \rightarrow (0, \infty)$. If ε is regularly decreasing, the maximum in (14) is bounded by a constant that does not depend on m . Roughly speaking, the algorithm Q_m admits a mean square error of order $\varepsilon(m)$ with a sample size of order m for any f from (13).

Remark 3 (Optimal approximation within a subspace). Let D be a Borel subset of \mathbb{R}^d with positive Lebesgue measure, \mathcal{A} be the Borel sigma algebra on D , and μ be a probability measure on (D, \mathcal{A}) . The best approximation of $f \in L^2(D, \mathcal{A}, \mu)$ within the subspace V_m is given by $P_m f$. Its error is given by the number

$$e_m(f) = \inf_{v \in V_m} \|f - v\|_2 = \|f - P_m f\|_2.$$

In general, we cannot find $P_m f$ by sampling only a finite number of function values of f . What we can provide is a random approximation v_m within V_m whose root mean square error

$$\left(\mathbb{E} \|f - v_m\|_2^2\right)^{1/2}$$

is close to $e_m(f)$. If we know the numbers $e_m(f)$ for all $m \in \mathbb{N}$ (or some good upper bound) and if they are regularly decreasing, we can choose v_m as the output of the method Q_m from Theorem 3, which uses a sample size of order m . But even if we do not know anything about $f \in L^2$, we can still find an approximation v_m like above. We only need the mild assumption that V_m consists of functions defined everywhere on D and that for each $x \in D$, there is some $v \in V_m$ with $v(x) \neq 0$. We can then choose v_m as the output of a weighted least squares method, see Cohen and Migliorati [3, Theorem 2.1 (iv)]. The sample size of this method, however, is at least of order $m \ln m$. In both cases, the involved proportionality constants are independent of the dimension of the domain D .

Acknowledgements I wish to thank Erich Novak, Robert Kunsch, Winfried Sickel, and two anonymous referees, whose comments and questions led to the present generality of the theorems.

References

1. Babenko, K.I.: About the approximation of periodic functions of many variable trigonometric polynomials. Dokl. Akad. Nauk SSR **32**, 247–250 (1960)
2. Cohen, A., Davenport, M.A., Leviatan, D.: On the stability and accuracy of least squares approximations. Found. Comput. Math. **13**, 819–834 (2013)
3. Cohen, A., Migliorati, G.: Optimal weighted least-squares methods. SMAI J. Comput. Math. **3**, 181–203 (2017)
4. Heinrich, S.: Multilevel Monte Carlo methods. In: Proceedings of the Third International Conference on Large-Scale Scientific Computing, Sozopol (Bulgaria), pp. 58–67. Springer, Berlin (2001)
5. Heinrich, S.: Random approximation in numerical analysis. In: Proceedings of the Conference Functional Analysis, Essen (Germany), pp. 123–171. Marcel Dekker (1994)
6. Heinrich, S.: Randomized approximation of Sobolev embeddings. In: Keller, A., Heinrich, S., Niederreiter, H. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2006, pp. 445–459. Springer, Berlin (2008)
7. Hinrichs, A., Novak, E., Vyřbřal, J.: Linear information versus function evaluations for L_2 -approximation. J. Approx. Theory **153**, 97–107 (2008)
8. Jerome, J.W.: On the L_2 n -width of certain classes of functions of several variables. J. Math. Anal. Appl. **20**, 110–123 (1967)
9. Krieg, D.: Tensor power sequences and the approximation of tensor product operators. J. Complex. **44**, 30–51 (2018)

10. Krieg, D., Novak, E.: A universal algorithm for multivariate integration. *Found. Comput. Math.* **17**(4), 895–916 (2017)
11. Kühn, T., Mayer, S., Ullrich, T.: Counting via entropy: new preasymptotics for the approximation numbers of Sobolev embeddings. *SIAM J. Numer. Anal.* **54**(6), 3625–3647 (2016)
12. Kühn, T., Sickel, W., Ullrich, T.: Approximation of mixed order Sobolev functions on the d -torus— asymptotics, preasymptotics and d -dependence. *Constr. Approx.* **42**, 353–398 (2015)
13. Mathé, P.: Random approximation of Sobolev embeddings. *J. Complex.* **7**, 261–281 (1991)
14. Mityagin, B.S.: Approximation of functions in L^p and C on the torus. *Math. Notes* **58**, 397–414 (1962)
15. Nikol'skaya, N.S.: Approximation of differentiable functions of several variables by Fourier sums in the L_p -metric. *Sibirsk. Mat. Zh.* **15**, 395–412 (1974); English transl. in *Siberian Math. J.* **15** (1974)
16. Novak, E.: Optimal linear randomized methods for linear operators in Hilbert spaces. *J. Complex.* **8**, 22–36 (1992)
17. Novak, E., Woźniakowski, H.: *Tractability of Multivariate Problems. Volume I: Linear Information.* EMS, Zürich (2008)
18. Novak, E., Woźniakowski, H.: *Tractability of Multivariate Problems. Volume III: Standard Information for Operators.* EMS, Zürich (2012)
19. Sickel, W., Ullrich, T.: Spline interpolation on sparse grids. *Appl. Anal.* **90**, 337–383 (2010)
20. Traub, J.F., Wasilkowski, G.W., Woźniakowski, H.: *Information-Based Complexity.* Academic Press, Cambridge (1988)
21. Triebel, H.: Sampling numbers and embedding constants. *Proc. Steklov Inst. Math.* **248**, 268–277 (2005)
22. Ullrich, M.: A Monte Carlo method for integration of multivariate smooth functions. *SIAM J. Numer. Anal.* **55**(3), 1188–1200 (2017)
23. Wasilkowski, G.W., Woźniakowski, H.: The power of standard information for multivariate approximation in the randomized setting. *Math. Comput.* **76**, 965–988 (2006)