ON THE POWER OF RANDOM INFORMATION

AICKE HINRICHS, DAVID KRIEG, ERICH NOVAK, JOSCHA PROCHNO, AND MARIO ULLRICH

ABSTRACT. We study approximation and integration problems and compare the quality of optimal information with the quality of random information. For some problems random information is almost optimal and for some other problems random information is much worse than optimal information. We prove new results and give a short survey of known results.

1. INTRODUCTION

Optimal algorithms use optimal information about the problem instances. Quite often, we do not have access to optimal information. One reason may be that we do not know how to choose the most significant measurements. In this paper, we assume that information comes in randomly, which is a standard assumption in learning theory and uncertainty quantification. We ask the following question:

> What is the typical quality of random information?

Of course, random information cannot be better than optimal information. In many cases, however, it will turn out that random information is practically as good as optimal information while in other situations random information is a little worse. Sometimes, random information is completely useless.

To phrase our question precisely, we need to clarify how we measure the quality of information and what we mean by random. The first is done with the so-called radius of information, which is the worst case error of the best algorithm that uses nothing but the given information and the a priori knowledge about the problem instance. Random information, on the other hand, shall be obtained from a certain number of independent measurements that all follow the same law. We study the question for two different natural choices of uniform distributions.

Let us go a little more into detail. A linear problem is given by a linear solution operator S that maps from a convex and symmetric subset F of a normed space to a normed space G and a class Λ of continuous linear functionals on F, the class of admissible measurements. We may think of an integration problem, where $S(f)$ is the integral of a function f, or a recovery problem, where S is an embedding. One wants to approximate the solution $S(f)$ for unknown $f \in F$ based on n of these measurements such that we can guarantee a small error with respect to the norm in G. We refer the reader to [29, 36] for a detailed exposition.

An information mapping has the form

$$(1) \qquad N_n : F \to \mathbb{R}^n, \quad N_n(f) = \big(L_1(f), \dots, L_n(f)\big),$$

where $L_1, \dots, L_n \in \Lambda$. The power or quality of the information mapping is measured by the radius of information. This is the worst case error of the best algorithm $A_n = \varphi \circ N_n$ based on this information, i.e.,

$$(2) \qquad r(N_n) = \inf_{\varphi : \mathbb{R}^n \to G} \sup_{f \in F} \| S(f) - \varphi(N_n(f)) \|_G .$$

We say that information is optimal, and write N_n^* instead of N_n, if

$$(3) \qquad r(N_n^*) = \inf_{N_n} r(N_n).$$

Here we study random information of the form (1), where the random functionals $L_1, \dots, L_n \in \Lambda$ are independent and identically distributed. The goal is to compare

$$r(N_n^*) \qquad \text{vs.} \qquad \mathbb{E}[r(N_n)],$$

the radius of optimal information and the expected radius of random information.

If the infimum and the expected value are comparable, this means that there are many good algorithms based on many different information mappings. In this case, optimal information and therefore optimal algorithms are not very special. On the other hand, if the infimum is significantly smaller than the expected value, this means that optimal information is indeed special. It seems to be an interesting characteristic of a problem whether optimal information is special or not. Of course, the results of our comparisons heavily depend on the distribution underlying the measurements. While the question may be interesting for many distributions, we feel that there often is a natural choice. If the distribution only depends on the class Λ of admissible measurements, collecting random information might even be a good idea if optimal information is available. It may happen that we do not loose much in terms of the radius but gain the following useful properties:

- Since the distribution is independent of n, it is easy to increase the number of measurements if our current approximation is not yet satisfactory.
- The information can be used for many different input classes F, solution operators S, and target spaces G. In that sense it is *universal*.

The second property does not mean that the corresponding algorithm $A_n = \varphi \circ N_n$ is universal. The optimal choice of φ may depend on F, S, and G. It is well known, see again [36], that any interpolatory algorithm is optimal up to a factor two.

We note that examples of the sort '*random information is good*' can be deduced from work using the probabilistic method to prove upper bounds of $r(N_n^*)$. Here one uses upper bounds of $\mathbb{E}[r(N_n)]$ and the trivial inequality $r(N_n^*) \leq \mathbb{E}[r(N_n)]$, see [17, 29, 31, 32, 33] for many such results. The idea is to introduce a random family of algorithms and to show that the expected worst case error is small. This is used to obtain the existence of good algorithms. However, it actually implies that *most* of the algorithms in that family are good.

Therefore, the expected radius of the random information underlying these algorithms must be small as well. We note that for this approach only upper bounds of $\mathbb{E}[r(N_n)]$ are needed and lower bounds are usually not studied.

2. UNIFORMLY DISTRIBUTED FUNCTION EVALUATIONS

If the input class F consists of functions, one natural restriction of the class of information functionals is so called *standard information* consisting of function evaluations. Moreover, if the domain of definition of functions in F carries a probability measure, a natural source of random information are function values $f(x)$ with x distributed according to the probability measure. In this section, we analyze this setting for different input classes F of smooth functions defined on the unit cube $[0,1]^d$. Then the random information consists of function values $f(x)$ with x uniformly distributed in $[0,1]^d$.

2.1. **Integration and Approximation in univariate Sobolev spaces of smoothness 1.** We consider the L_q-approximation problem

$$\text{APP}\colon W_p^1([0,1]) \to L_q([0,1])$$

given by $\text{APP}(f) = f$ using function values as information. The Sobolev spaces $W_p^1([0,1])$ are usually normed by

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

It is well known that the optimal information is given by

$$N_n^*(f) = \big(f(1/n), f(3/n), \ldots, f((2n-1)/n) \big),$$

i.e., point evaluations at equidistant points. The minimal radius of information then satisfies

$$r(N_n^*) \asymp n^{-1+(1/p-1/q)_+},$$

see (6).

Here and later we use asymptotic notation as follows. For sequences (a_n) and (b_n) of positive real numbers, we write $a_n \preccurlyeq b_n$ and $b_n \succcurlyeq a_n$ to indicate that there exists a constant $C \in (0,\infty)$ such that $a_n \leq C\, b_n$ for all n. We shall write $a_n \asymp b_n$ if $a_n \preccurlyeq b_n$ and $b_n \preccurlyeq a_n$. Sometimes we use similar notation for double sequences.

We will show that information given by n independent and uniformly distributed points in $[0,1]$ is as good as optimal information provided that $p > q$, whereas in the case $p \leq q$ there is a loss of a logarithmic factor.

Theorem 1. *Consider L_q-approximation of functions from $W_p^s([0,1])$ with $1 \leq p,q \leq \infty$, using function values at uniformly distributed points in $[0,1]$.*

$$\mathbb{E}[r(N_n)] \asymp \begin{cases} r(N_n^*) \asymp n^{-1} & \text{if} \quad p > q \\[2ex] r(N_{n/\log n}^*) \asymp \left(\frac{n}{\log n} \right)^{-1+1/p-1/q} & \text{if} \quad p \leq q. \end{cases}$$

The proof of Theorem 1 is based on the following formula for the radius of information $r(N_n)$ for information $N_n(f) = \big(f(x_1), f(x_2), \ldots, f(x_n)\big)$, where we assume $0 = x_0 \leq x_1 \leq \cdots \leq x_n \leq x_{n+1} = 1$.

Lemma 2.

$$r(N_n) \asymp \begin{cases} \left(\sum_{i=0}^n (x_{i+1} - x_i)^{(pq+p-q)/(p-q)} \right)^{1/q-1/p} & \text{if} \quad p > q \\ \max_{0 \leq i \leq n} (x_{i+1} - x_i)^{1-1/p+1/q} & \text{if} \quad p \leq q. \end{cases}$$

Proof. Note that in this case the radius of information satisfies

$$r(N_n) \asymp \sup_{\substack{f \in W_p^1([0,1]) \\ N_n(f)=0}} \frac{\|f\|_q}{\|f\|_{W_p^1}},$$

with an equivalence constant of at most 2, see [29, Section 4.2]. We start with the lower bound in the case $p \leq q$. Let $i_0 \in \{0, 1, \ldots, n\}$ be such that

$$\ell = x_{i_0+1} - x_{i_0} = \max_{0 \leq i \leq n} (x_{i+1} - x_i)$$

and consider a continuous piecewise linear function $f : [0,1] \to \mathbb{R}$ that is zero outside of the interval (x_{i_0}, x_{i_0+1}), satisfies $f'(x) = 1$ in the left half of this interval and $f'(x) = -1$ in the right half. Then

$$N_n(f) = (0, \ldots, 0), \quad \|f\|_q \asymp \ell^{1+1/q}, \quad \text{and} \quad \|f\|_{W_p^1} \asymp \ell^{1/p}$$

show that $r(N_n) \succcurlyeq \ell^{1-1/p+1/q}$.

We turn to the lower bound in the case $p > q$. In this case, let $f : [0,1] \to \mathbb{R}$ be a continuous piecewise linear function satisfying $f(x_i) = 0$ for $i = 0, 1, \ldots, n+1$ and $f'(x) = \ell_i^\alpha$ in the left half of each interval $[x_i, x_{i+1}]$ and $f'(x) = -\ell_i^\alpha$ in the right half, where $\ell_i = x_{i+1} - x_i$ and $\alpha = q/(p-q)$. Then $N_n(f) = (0, \ldots, 0)$ and

$$\|f\|_q \asymp \left(\sum_{i=0}^n \ell_i^{(pq+p-q)/(p-q)} \right)^{1/q} \quad \text{and} \quad \|f\|_{W_p^1} \asymp \left(\sum_{i=0}^n \ell_i^{(pq+p-q)/(p-q)} \right)^{1/p}$$

show that $r(N_n) \succcurlyeq \left(\sum_{i=0}^n \ell_i^{(pq+p-q)/(p-q)} \right)^{1/q-1/p}$.

Finally, we prove the upper bound. Observe that $\|f\|_\infty \leq \|f'\|_1$ holds for any function $f \in W_1^1([0,1])$ with $f(0) = 0$. Consequently, $\|f\|_q \leq \|f'\|_p$ holds for any function $f \in W_p^1([0,1])$ with $f(0) = 0$. By scaling, it follows that $\|f\|_q \leq (b-a)^{1-1/p+1/q}\|f'\|_p$ holds for any function $f \in W_p^1([a,b])$ satisfying either $f(a) = 0$ or $f(b) = 0$. Let now $f \in W_p^1([0,1])$ be such that $N_n(f) = \big(f(x_1), f(x_2), \ldots, f(x_n)\big) = (0, \ldots, 0)$ and $\|f'\|_p = 1$. Let f_i be the restriction of f to the interval $[x_i, x_{i+1}]$. Then

$$\|f\|_q = \left(\sum_{i=0}^n \|f_i\|_q^q \right)^{1/q} \leq \left(\sum_{i=0}^n (x_{i+1} - x_i)^{q-q/p+1} \|f_i'\|_p^q \right)^{1/q}.$$

In the case $p \leq q$, let again ℓ be the length of the largest of the intervals $[x_i, x_{i+1}]$. It follows that

$$\|f\|_q \leq \ell^{1-1/p+1/q} \left(\sum_{i=0}^{n} \|f_i'\|_p^q \right)^{1/q} \leq \ell^{1-1/p+1/q} \left(\sum_{i=0}^{n} \|f_i'\|_p^p \right)^{1/p}$$

$$= \ell^{1-1/p+1/q} \|f'\|_p = \ell^{1-1/p+1/q}.$$

This implies $r(N_n) \leq \ell^{1-1/p+1/q}$ as claimed. In the case $p > q$, let again $\ell_i = x_{i+1} - x_i$ and $s > 0$ be given by $\frac{1}{s} = \frac{1}{q} - \frac{1}{p}$. Using Hölder's inequality, we arrive at

$$\|f\|_q \leq \left(\sum_{i=0}^{n} \ell_i^{s+1} \right)^{1/s} \left(\sum_{i=0}^{n} \|f_i'\|_p^p \right)^{1/p} = \left(\sum_{i=0}^{n} \ell_i^{s+1} \right)^{1/s} \|f'\|_p$$

$$= \left(\sum_{i=0}^{n} \ell_i^{s+1} \right)^{1/s}.$$

This implies the claimed upper bound also when $p > q$. $\qquad\square$

Proof of Theorem 1. In the case $p \leq q$ the proof follows from Lemma 2 and the fact that, in expectation, the largest gap among n independent and uniformly distributed points in $[0, 1]$ is of order $\log(n)/n$. This is a simple consequence of Lemma 7. In the case $p > q$ the proof follows from

$$\mathbb{E}\left[\left(\sum_{i=0}^{n} \ell_i^{s+1} \right)^{1/s} \right] \asymp \frac{1}{n}$$

for any $s > 0$, where the ℓ_i are again the spacings of independent and uniformly distributed points in $[0, 1]$. In fact, the lower bound follows trivially since the expected radius of random information is larger than the radius of optimal information. To prove the upper bound observe first that the function $s \mapsto \left(\sum_{i=0}^{n} \ell_i^{s+1} \right)^{1/s}$ is non-decreasing, since $\sum_{i=0}^{n} \ell_i = 1$, and that Jensen's inequality implies

$$\mathbb{E}\left[\left(\sum_{i=0}^{n} \ell_i^{s+1} \right)^{1/s} \right] \leq \left(\mathbb{E}\left[\sum_{i=0}^{n} \ell_i^{s+1} \right] \right)^{1/s} \qquad \text{whenever } s \geq 1.$$

Hence, it is enough to show that

$$\mathbb{E}\left[\sum_{i=0}^{n} \ell_i^{s+1} \right] \asymp n^{-s}$$

for positive integers s. It was proved in [9] that

$$\mathbb{E}\left[\sum_{i=0}^{n} h(\ell_i) \right] = n(n+1) \int_0^1 (1-r)^{n-1} h(r) \mathrm{d}r$$

for any integrable function $h : [0, 1] \to \mathbb{R}$. With $h(r) = r^{s+1}$, we obtain

$$\mathbb{E}\left[\sum_{i=0}^{n} \ell_i^{s+1} \right] = n(n+1)B(s+2, n) = \frac{(n+1)!(s+1)!}{(n+s+1)!} \asymp n^{-s},$$

where $B(x, y)$ is the Beta function. This completes the proof. $\qquad\square$

Remark 3. For the classes studied, i.e., when the smoothness is one, the radius of information $r(N_n)$ for the integration problem with $S(f) = \int_0^1 f(x)\,\mathrm{d}x$ is the same as for the L_1-approximation problem. Hence, for the integration problem for functions in $W_p^1([0,1])$, we have

$$\mathbb{E}[r(N_n)] \asymp \begin{cases} r(N_n^*) \asymp n^{-1} & \text{if} \quad p > 1 \\ \\ r(N_{n/\log n}^*) \asymp \left(\frac{n}{\log n}\right)^{-1} & \text{if} \quad p = 1. \end{cases}$$

Some cases of Theorem 1 are known, sometimes in a slightly different language since authors had in mind discrepancy (with optimal weights) instead of the integration problem. The star discrepany for $d = 1$, i.e., $p = q = 1$, was studied in [38]. Here we have $r(N_n^*) = \frac{1}{2n}$. From Corollary 3.3 and Theorem 3.5 of [38] we know that

$$\frac{c \log n}{n} \leq \mathbb{E}[r(N_n)] \leq \frac{8 \log n}{n}$$

for some $c \in (0, \infty)$. Therefore, random information is good, but not optimal. Also the case $p = 2$ and $q = 1$ is covered by [38, Theorems 4.2 and 4.3].

2.2. Approximation of multivariate Lipschitz functions.

We study the problem of L_q-approximation of functions from the class

$$(4) \qquad F_d = \left\{ f : [0,1]^d \to \mathbb{R} \, : \, |f(x) - f(y)| \leq \mathrm{dist}(x,y) \ \forall x, y \in [0,1]^d \right\}.$$

Here, dist denotes the maximum metric on the d-torus, i.e.,

$$\mathrm{dist}(x, y) = \min_{k \in \mathbb{Z}^d} \|x + k - y\|_\infty \quad \text{for} \quad x, y \in [0,1]^d.$$

We consider standard information of the form

$$N_n(f) = (f(x_1), \ldots, f(x_n)), \qquad x_1, \ldots, x_n \in [0,1]^d.$$

An optimal algorithm based on N_n works as follows:

Algorithm. For N_n as above and $f \in F_d$, we define $A_n(f) = (f^+ + f^-)/2$, where

$$f^+(x) = \min_{i \leq n} \left(f(x_i) + \mathrm{dist}(x, x_i) \right) \quad \text{and} \quad f^-(x) = \max_{i \leq n} \left(f(x_i) - \mathrm{dist}(x, x_i) \right).$$

Note that f^+ and f^- are the maximal and the minimal function in F_d that interpolate f at the nodes $x_1, \ldots, x_n$. We have the following relations that we prove for the sake of completeness, see e.g. [29, 36].

Lemma 4.

$$r(N_n) = \sup_{f \in F_d} \|f - A_n(f)\|_q = \sup_{\substack{f \in F_d \\ N_n(f) = 0}} \|f\|_q.$$

Proof. Clearly A_n is of the form $\varphi \circ N_n$ for some mapping $\varphi : \mathbb{R}^n \to L_q$. The definition of the radius yields

$$\sup_{f \in F_d} \|f - A_n(f)\|_q \geq r(N_n) \geq \sup_{\substack{f \in F_d \\ N_n(f)=0}} \|f\|_q .$$

On the other hand, any $f \in F_d$ satisfies the pointwise estimate

(5) $$|f - A_n(f)| \leq \frac{f^+ - f^-}{2} .$$

This implies

$$\sup_{\substack{f \in F_d \\ N_n(f)=0}} \|f\|_q \geq \sup_{f \in F_d} \|f - A_n(f)\|_q$$

since the right-hand side of (5) is an element of F_d that vanishes on $\{x_1, \ldots, x_n\}$. Altogether, we obtain the claimed identities. $\qquad\square$

It is also well known that optimal information is given by function values on a regular grid if $n = m^d$ for some $m \in \mathbb{N}$. The radius of optimal information N_n^* satisfies

$$r(N_n^*) \asymp n^{-1/d}$$

for all $1 \leq q \leq \infty$. This follows from the upper bound on the complexity of uniform approximation as studied in [34] and the lower bound on the complexity of numerical integration as studied in [35]. Using the technique of proof of the latter, we even obtain a precise formula for the radius of optimal information.

Proposition 5. *Let $n = m^d$ for some $m \in \mathbb{N}$. Then*

$$r(N_n^*) = \begin{cases} \dfrac{1}{2} \left(\dfrac{d}{d+q} \right)^{1/q} n^{-1/d} & if \quad 1 \leq q < \infty, \\[2ex] \dfrac{1}{2} \, n^{-1/d} & if \quad q = \infty. \end{cases}$$

Proof. Let N_n be as above and let $P_n = \{x_1, \ldots, x_n\}$. Note that the function $\mathrm{dist}(\cdot, P_n)$ is contained in F_d and vanishes on P_n. On the other hand, every other function $f \in F_d$ that vanishes on P_n must satisfy

$$|f(x)| \leq \mathrm{dist}(x, P_n) \qquad \text{for all } x \in [0, 1]^d.$$

This yields

$$r(N_n) = \sup_{\substack{f \in F_d \\ N_n(f)=0}} \|f\|_q = \|\mathrm{dist}(\cdot, P_n)\|_q .$$

Let us first consider the case $q = \infty$. Since the volume of the union of the balls $B_r^\infty(x)$ over $x \in P_n$ is smaller than 1 for all $r < 1/(2m)$, there must be some $x \in [0, 1]^d$ with $\mathrm{dist}(x, P_n) \geq r$. Therefore,

$$\|\mathrm{dist}(\cdot, P_n)\|_\infty \geq \frac{1}{2m},$$

where equality holds when $P_n = \{i/m \,:\, 0 \le i < m\}^d$. Let us turn to the case $q < \infty$. Let λ^d denote Lebesgue measure on $[0,1]^d$. Then

$$r(N_n)^q = \int_{[0,1]^d} \mathrm{dist}(x, P_n)^q \, \mathrm{d}x = \int_0^\infty \lambda^d \left(\mathrm{dist}(x, P_n)^q \ge t \right) \mathrm{d}t.$$

Note that

$$\lambda^d \left(\mathrm{dist}(x, P_n)^q \ge t \right) = 1 - \lambda^d \left(\mathrm{dist}(x, P_n)^q < t \right) \ge 1 - n \cdot 2^d t^{d/q},$$

where equality holds if the sets $B^\infty_{t^{1/q}}(y)$ for $y \in P_n$ are pairwise disjoint. Hence,

$$r(N_n)^q \ge \int_0^{(1/2m)^q} \lambda^d \left(\mathrm{dist}(x, P_n)^q \ge t \right) \mathrm{d}t$$

$$\ge \frac{1}{(2m)^q} - 2^d n \int_0^{(1/2m)^q} t^{d/q} \mathrm{d}t = \frac{1}{(2m)^q} \frac{d}{d+q}$$

with equality for $P_n = \{i/m \,:\, 0 \le i < m\}^d$. This proves the statement. $\qquad\square$

In the following, we study the expected radius of random information

$$N_n(f) = \big(f(x_1), \ldots, f(x_n) \big),$$

where the points $x_1, \ldots, x_n$ are independent and uniformly distributed in $[0,1]^d$. If q is finite, the qth moment of the radius at zero can be computed precisely.

Proposition 6. *Consider L_q-approximation of functions from F_d defined in (4) with $1 \le q < \infty$, using function values at uniformly distributed points in $[0,1]^d$. Then*

$$\mathbb{E}\left[r(N_n)^q\right] = \frac{1}{2^q} \frac{n!}{(q/d+1)\cdots(q/d+n)}.$$

In particular, we have

$$\lim_{n\to\infty} n^{1/d} \sqrt[q]{\mathbb{E}\left[r(N_n)^q\right]} = \frac{1}{2} \sqrt[q]{\Gamma(q/d+1)}.$$

Proof. Let $P_n = \{x_1, \ldots, x_n\}$. Recall that

$$r(N_n)^q = \int_{[0,1]^d} \mathrm{dist}(x, P_n)^q \, \mathrm{d}x.$$

Using Tonelli's theorem, we obtain

$$\mathbb{E}\left[r(N_n)^q\right] = \int_{[0,1]^d} \mathbb{E}\left[\mathrm{dist}(x, P_n)^q\right] \mathrm{d}x.$$

We will show that the integrand of the latter integral is constant. Let us fix $x \in [0,1]^d$ and note that $\mathrm{dist}(x, P_n) \in [0, 1/2]$. For any $t \in [0, 1/2]$, we have

$$\mathrm{dist}(x, P_n) \ge t \quad \Longleftrightarrow \quad \forall i \in \{1, \ldots, n\} : x_i \notin B^\infty_t(x).$$

Thus

$$\mathbb{P}[\mathrm{dist}(x, P_n) \ge t] = \left(1 - (2t)^d \right)^n.$$

The substitution $s = 1 - (2t^{1/q})^d$ and integration by parts yield

$$\mathbb{E}\big[\operatorname{dist}(x, P_n)^q\big] = \int_0^{2^{-q}} \mathbb{P}[\operatorname{dist}(x, P_n)^q \geq t]\,\mathrm{d}t = \int_0^{2^{-q}} \left(1 - (2t^{1/q})^d\right)^n \mathrm{d}t$$

$$= \frac{q/d}{2^q} \int_0^1 s^n (1-s)^{q/d-1}\,\mathrm{d}s = \frac{1}{2^q} \frac{n!}{(q/d+1)\cdots(q/d+n)},$$

which implies the statement of the theorem. $\square$

We now turn to the case of uniform approximation, i.e., $q = \infty$. In this case, the expected radius of information is closely related to the so called coupon collector's problem. This is the question for the random number τ_ℓ of coupons that a coupon collector has to collect to obtain a complete set of ℓ distinct coupons. The following facts on the distribution of τ_ℓ are well known, see [27]. Here $H_\ell = \sum_{k=1}^\ell 1/k$ is the ℓth harmonic number. Note that $H_\ell \sim \log \ell$ as $\ell \to \infty$.

Lemma 7. *Let $(Y_i)_{i=1}^\infty$ be a sequence of random variables that are uniformly distributed in the set $\{1, \ldots, \ell\}$ and let*

$$\tau_\ell = \min\left\{n \in \mathbb{N} : \{Y_1, \ldots, Y_n\} = \{1, \ldots, \ell\}\right\}.$$

Then

$$\mathbb{E}[\tau_\ell] = \ell H_\ell \quad and \quad \operatorname{Var}[\tau_\ell] \leq \ell^2 \sum_{k=1}^\ell 1/k^2$$

and for any $c \in (0, \infty)$,

$$\mathbb{P}\big[\tau_\ell > \lceil c\,\ell \log \ell \rceil\big] \leq \ell^{-c+1}.$$

This leads to the following estimates of the expected radius for $q = \infty$.

Proposition 8. *Consider L_∞-approximation of functions from F_d defined in (4). For $n \in \mathbb{N}$, let*

$$m_1 = \min\left\{m \in \mathbb{N} : m^d(H_{m^d} - 2) \geq n\right\},$$
$$m_2 = \max\left\{m \in \mathbb{N} : 2m^d \log(m^d) \leq n\right\}.$$

Then

$$\frac{1}{4m_1} \leq \mathbb{E}[r(N_n)] \leq \frac{2}{m_2}.$$

Proof. Let $m \in \mathbb{N}$ and decompose $[0, 1]^d$ into $\ell = m^d$ boxes

$$\prod_{i=1}^d \left[\frac{k_i - 1}{m}, \frac{k_i}{m}\right), \quad k_1, \ldots, k_d \in \{1, 2, \ldots, m\}$$

of equal volume. Again, let $P_n = \{x_1, \ldots, x_n\}$. Recall that the radius of the information $N_n(f) = \big(f(x_1), \ldots, f(x_n)\big)$ is given by

$$r(N_n) = \max_{x \in [0,1]^d} \operatorname{dist}(x, P_n).$$

Therefore, $r(N_n)$ is bounded above by $1/m$ if every box contains a point of P_n, and bounded below by $1/(2m)$ if one of the boxes does not contain a point of P_n. Let A be the event that every box contains a point. Note that the number of random points x_i that it takes to hit all the boxes follows the distribution of the coupon collector's variable τ_ℓ as defined in Lemma 7. For the upper bound, we choose $m = m_2$. Lemma 7 yields

$$\mathbb{P}[A^c] = \mathbb{P}[\tau_\ell > n] \leq 1/\ell$$

and hence

$$\mathbb{E}[r(N_n)] \leq \mathbb{P}[A] \cdot \frac{1}{m} + \mathbb{P}[A^c] \cdot 1 \leq \frac{2}{m}.$$

For the lower bound, we choose $m = m_1$. Chebyshev's inequality yields

$$\mathbb{P}[A] = \mathbb{P}[\tau_\ell \leq n] \leq \mathbb{P}[\tau_\ell \leq \ell H_\ell - 2\ell] \leq \frac{\mathrm{Var}[\tau_\ell]}{4\ell^2} \leq \frac{1}{2}.$$

We obtain

$$\mathbb{E}[r(N_n)] \geq \mathbb{P}[A^c] \frac{1}{2m} \geq \frac{1}{4m},$$

as it was to be proven. $\qquad\square$

We arrive at the following result on the power of random information proved in [25]. Note that the case $q = \infty$ is already known from [2], where the authors study the uniform approximation of functions on $[0, 1]^d$ with bounded r^{th} derivative.

Theorem 9 (Krieg)**.** *Consider L_q-approximation of functions from F_d defined in (4) with $1 \leq q \leq \infty$, using function values at uniformly distributed points in $[0, 1]^d$. Then*

$$\mathbb{E}[r(N_n)] \asymp \begin{cases} r(N_n^*) \asymp n^{-1/d} & \text{if} \quad q < \infty \\[2ex] r\left(N_{n/\log n}^*\right) \asymp \left(\frac{n}{\log n}\right)^{-1/d} & \text{if} \quad q = \infty. \end{cases}$$

Proof. The statements on the rate of the optimal radius of information follow from Proposition 5. The upper bound on the expected radius in the first case follows from Proposition 6, since the expected radius is bounded above by its qth moment. The lower bound is trivial. The upper and lower bound on the expected radius in the second case follow from Proposition 8, since both m_1^d and m_2^d are of order $n/\log n$. $\qquad\square$

Thus, in the sense of order of convergence, random information is as good as optimal information for the problem of L_q-approximation on F_d if $q < \infty$. If $q = \infty$, we loose a logarithmic factor.

Remark 10 (Modifications of F_d)**.** The rates of convergence of the expected and the optimal radius do not change if we replace the maximum metric on the torus by some equivalent metric. The same holds true if we change the Lipschitz constant or if we switch to the non-periodic setting.

2.3. **Classical Sobolev spaces.** We study the general L_q-approximation problem for Sobolev spaces,

$$\text{APP} \colon W_p^s([0,1]^d) \to L_q([0,1]^d)$$

given by $\text{APP}(f) = f$ with standard information. A typical choice for the norm in this space is

$$\|f\|_{W_p^s([0,1]^d)} = \left(\sum_{\beta \in \mathbb{N}_0^d \colon \|\beta\|_1 \leq s} \|D^\beta f\|_p^p \right)^{1/p}.$$

It is known that the radius of optimal information N_n^* satisfies

$$(6) \qquad\qquad r(N_n^*) \asymp n^{-s/d + (1/p - 1/q)_+},$$

where we assume that $s > d/p$ holds, to ensure that the functions are continuous. This can be found in [28] for general Lipschitz domains and was known earlier for special cases, such as the cube, see [28, Remark 24]. Here we ask for the expected radius of random information N_n given by n independent and uniformly distributed points in $[0,1]^d$. We conjecture that random information is as good as optimal information provided that $p > q$, whereas in the case $p \leq q$ there is a loss of a logarithmic factor. This conjecture is true in the following special cases:

- Section 2.1 covers univariate Sobolev spaces of smoothness 1, i.e., $s = d = 1$.
- Section 2.2 covers Lipschitz functions, i.e., $s = 1$ and $p = \infty$.
- The paper [2] covers uniform approximation on Hölder classes, i.e., $p = q = \infty$.

In the general case, we are only able to prove the second part of our conjecture.

Theorem 11. *Consider L_q-approximation of functions from $W_p^s([0,1]^d)$ with $1 \leq p \leq q \leq \infty$, using function values at uniformly distributed points in $[0,1]^d$. Then*

$$\mathbb{E}[r(N_n)] \asymp r\left(N_{n/\log n}^* \right) \asymp \left(\frac{n}{\log n} \right)^{-s/d + 1/p - 1/q}.$$

We use the following criterion for optimal point sets, which can be found in [28].

Lemma 12. *Consider L_q-approximation of functions from $W_p^s([0,1]^d)$ with $1 \leq p, q \leq \infty$. For all $n \in \mathbb{N}$, let $P_n \subset [0,1]^d$ be a point set of cardinality n,*

$$d_n = \min_{x,y \in P_n} \|x - y\|_\infty, \quad \text{and} \quad r_n = \max_{y \in [0,1]^d} \min_{x \in P_n} \|x - y\|_\infty.$$

If $r_n \preccurlyeq d_n$, then the information N_n with nodes P_n satisfies $r(N_n) \asymp r(N_n^*)$.

Proof of Theorem 11. Let $\alpha = s/d - 1/p + 1/q$. We start with the lower bound. Let $g \colon \mathbb{R}^d \to \mathbb{R}$ be a smooth function with support in the ℓ_∞-unit ball that is not identically zero. Using Lemma 7, we obtain with probability $1/2$ that $[0,1]^d$ contains an ℓ_∞-ball of radius

$$\tau_n \asymp \left(\frac{\log n}{n} \right)^{1/d}.$$

not containing any of the points in P_n. If x_0 denotes the center of this ball, then the function $f \in W_p^s([0,1]^d)$ given by $f(x) = g(\tau_n^{-1}(x - x_0))$ has support in this ball and satisfies

$$\|f\|_{W_p^s([0,1]^d)} \asymp \tau_n^{-s+d/p} \quad \text{and} \quad \|f\|_q \asymp \tau_n^{d/q}.$$

This yields

$$\mathbb{E}[r(N_n)] \geq \frac{1}{2} \frac{\|f\|_q}{\|f\|_{W_p^s([0,1]^d)}} \asymp \tau_n^{s-d/p+d/q} \asymp \left(\frac{n}{\log n}\right)^{-\alpha}.$$

We turn to the upper bound. We choose $m \in \mathbb{N}_0$ maximal such that $\ell = (3m)^d$ satisfies

$$n \geq (\alpha + 1)\ell \log \ell.$$

We split the unit cube into $k = m^d$ subcubes of equal volume. Moreover, we split each of these into 3^d subcubes of equal volume. This gives us $\ell = (3m)^d$ small cubes. Let A be the event that we hit every small cube. Lemma 7 yields

(7) $$\mathbb{P}[A^c] = \mathbb{P}[\tau_\ell > n] \leq \ell^{-\alpha} \asymp \left(\frac{n}{\log n}\right)^{-\alpha}.$$

Assume now that A takes place. We choose exactly one point out of the small cube in the center of every large cube to obtain a subset P_k of P_n with cardinality k. Then

$$r_k \leq \frac{2}{3m} \leq d_k.$$

Let N_k' be the information map with nodes P_k. Lemma 12 yields

(8) $$r(N_n) \leq r(N_k') \preccurlyeq k^{-\alpha} \asymp \left(\frac{n}{\log n}\right)^{-\alpha}.$$

Together, (7) and (8) imply the statement. $\qquad\square$

We note that the authors of [12] study the integration problem for very general Sobolev spaces on closed manifolds. One can obtain the optimal rate $r(N_n^*) \asymp n^{-s/d}$. The authors prove the upper bound

$$\mathbb{E}[r(N_n)] \preccurlyeq n^{-s/d} \log(n)^{s/d},$$

where the random points are chosen with respect to the probability measure that defines the integral to be approximated. The given proof works with the covering radius of random points which is an important characteristic for $q = \infty$. For the integration problem or $q = 1$, we expect, however, that the additional log factor is needed only for $p = 1$.

Remark 13. Another important case are tensor products of univariate Sobolev spaces. These spaces are usually called Sobolev spaces with (dominating) mixed smoothness or in the periodic case sometimes Korobov spaces. One of the main features is that the optimal (main) orders of convergence are independent of the dimension, only the logarithmic terms depend on the dimension. We refer to [11] for a survey. There exist upper bounds for the numbers $\mathbb{E}[r(N_n)]$, but these bounds are weaker than for isotropic Sobolev spaces. One reason is that Lemma 12 does not hold for these spaces and the geometrical structure of

good sampling points is not known. See [1, 17, 20, 22, 29, 31, 32] for upper bounds on $\mathbb{E}[r(N_n)]$ in various cases.

Remark 14. Recovery of functions of many variables usually suffers from the curse of dimensionality, even for smooth C^∞ functions, see [30]. To avoid the curse one can impose structural properties, such as the sparsity with respect to the Fourier coefficients or another orthonormal system. Then one can find an approximation of the function with relatively few random function values, see [6, 8]. Another way to study structural properties is to define weighted norms as was first done in [33], see [29, 31, 32] for a survey.

3. Uniformly distributed linear information

In this section we study *linear information* consisting of arbitrary continuous linear functionals. We are going to present two recovery problems for which we compare random information with optimal information. The first is a classical result due to Kashin, Garnaev, and Gluskin [24, 14] on the recovery of ℓ_1^m-vectors in the Euclidean distance, while the second are novel results of the authors on the recovery of vectors in an m-dimensional ellipsoid in the Euclidean distance [18].

3.1. **Recovery of ℓ_1^m-vectors in the Euclidean norm.** We study the problem of recovery of vectors in the unit ball of ℓ_1^m with general linear information and error measured in the Euclidean norm. This is the approximation problem with $F = \ell_1^m$ and $G = \ell_2^m$, i.e., S is the identity map $\ell_1^m \hookrightarrow \ell_2^m$. The minimal worst case error, which is the minimal radius of information achievable with general linear information, is closely connected to the notion of Gelfand numbers or Gelfand widths. Gelfand numbers and their dual counterpart Kolmogorov numbers are fundamental concepts in classical and modern approximation and complexity theory.

For $n \in \mathbb{N}$, the nth Gelfand number of a linear bounded operator $S : F \to G$ between normed spaces F and G can be defined as

$$c_n(S) = \inf_{\substack{F_n \subseteq F \\ \operatorname{codim}(F_n) \leq n}} \sup_{\|f\|_F \leq 1, f \in F_n} \|S(f)\|_G.$$

If the subspace F_n of F is the kernel of a linear information map $N_n : S \to \mathbb{R}^n$, then

$$r(N_n, 0) = \sup_{\|f\|_F \leq 1, f \in F_n} \|S(f)\|_G$$

is the local radius of zero information. It is not hard to see that

$$r(N_n, 0) \leq r(N_n) \leq 2r(N_n, 0).$$

Since every subspace F_n of codimension at most n is the kernel of a linear information map $N_n : S \to \mathbb{R}^n$, we have the well-known relation

$$c_n(S) \leq r(N_n^*) \leq 2c_n(S)$$

between the radius of the optimal information N_n^* for the linear problem $S : F \to G$ and the Gelfand numbers $c_n(S)$.

The systematic study of the Gelfand numbers of identity maps $\ell_p^m \hookrightarrow \ell_q^m$ has a long tradition and is the basis for the study of Gelfand numbers of embeddings of Sobolev spaces and, therefore, also for the complexity of L_q-approximation of functions from Sobolev spaces. We want to discuss here the case $p = 1$ and $q = 2$ which was one of the difficult cases. Its solution is an example for the power of random information. While lower bounds had been obtained before, the breakthrough regarding the upper bounds was made using random approximation, a groundbreaking method having its origin in the work of Kashin [24]. The problem was finally settled in the series of papers by Kashin [23, 24], Gluskin [15, 16], and Garnaev and Gluskin [14].

Let us consider now the particular case of the identity $\ell_1^m \hookrightarrow \ell_2^m$, which is of importance in information based complexity since it also demonstrates that non-linear recovery algorithms can be much better than linear ones. Although the results in the literature are usually formulated in the language of Gelfand numbers, we discuss the results directly for the radius of information.

Recall that we now allow arbitrary linear information $N_n : \ell_1^m \to \mathbb{R}^n$. The random information we consider is Gaussian information, where $N_n = N_{n,m}$ is given by an $n \times m$-matrix with independent standard normal entries. The kernels of these maps are distributed according to the Haar measure on the Grassmannian manifold of n-codimensional subspaces in $\mathbb{R}^m$. In this sense, we can speak about uniformly distributed linear information.

In [24, Theorem 1], Kashin introduced his fundamentally new approach via random subspaces and obtained that

$$\mathbb{E}\big[r(N_{n,m})\big] \preccurlyeq \frac{1}{\sqrt{n}}\left(1 + \log \frac{m}{n}\right)^{3/2}.$$

Through a refinement of Kashin's arguments, Garnaev and Gluskin [14] were able to decrease the power $3/2$ to a square root, while at the same time providing the matching lower bounds. Altogether, we have the following result.

Theorem 15 (Kashin, Garnaev, Gluskin). *Consider the recovery problem of vectors from* ℓ_1^m *in the Euclidean norm and Gaussian information. Then*

$$\mathbb{E}\big[r(N_{n,m})\big] \asymp r(N_{n,m}^*) \asymp \min\left\{1, \sqrt{\frac{\log(1 + \frac{m}{n})}{n}}\right\}.$$

We want to sketch the proof of the upper bound

$$\mathbb{E}\big[r(N_{n,m})\big] \preccurlyeq \min\left\{1, \sqrt{\frac{\log(1 + \frac{m}{n})}{n}}\right\}$$

of Theorem 15 and demonstrate the power of the probabilistic method of Kashin, Garnaev, and Gluskin.

Step 1. As already explained, one considers a Gaussian random matrix $G = (g_{ij})_{i,j=1}^{n,m}$ with independent and identically distributed standard normal entries and defines the random information mapping

$$N_n : \mathbb{R}^m \to \mathbb{R}^n, \quad N_n(x) = Gx.$$

Step 2. One defines the mapping φ to be

$$\varphi : \mathbb{R}^n \to \mathbb{R}^m, \quad y \mapsto \mathrm{argmin}_{x \in \mathbb{R}^m, \, N_n(x)=y} \|x\|_1$$

and considers the corresponding random algorithm $A_n = \varphi \circ N_n$.

Step 3. One can now prove that with high probability the error of the random algorithm is small for any $x \in \mathbb{B}_1^m$. More precisely, if $\delta \in (0, \infty)$, then taking

$$n \asymp \frac{\log m}{\varepsilon^2} + \log \frac{2}{\delta}$$

yields a worst case error

$$\sup_{x \in \mathbb{B}_1^m} \|x - A_n(x)\|_2 \le \varepsilon$$

for the random algorithm A_n with probability at least $1 - \delta$.

As a matter of fact, one actually shows that most linear information mappings $N_n : \mathbb{R}^m \to \mathbb{R}^n$ lead to a small error and so random information is essentially as good as optimal information. We want to emphasize that no explicit construction for such a mapping is known. So the optimal information is not accessible in this case and one has to settle for random information. We also want to mention that linear algorithms corresponding to linear widths or approximation numbers are much worse than nonlinear algorithms for this recovery problem.

Let us mention that Garnaev and Gluskin actually obtained sharp bounds in the more general setting of recovery of vectors from ℓ_1^m in the ℓ_q^m-norm or Gelfand numbers of the identity $\ell_1^m \hookrightarrow \ell_q^m$. If the random information $N_{n,m}$ is again Gaussian information and $1 < q \le 2$, then

$$\mathbb{E}\big[r(N_{n,m})\big] \asymp r(N_{n,m}^*) \asymp \min\left\{1, \left(\frac{\log(1 + \frac{m}{n})}{n}\right)^{1-1/q}\right\}.$$

Later, Gelfand widths have also attracted quite some attention in the area of compressive sensing over the last two decades, see, e.g., [3, 4, 7, 10]. The goal of compressive sensing is to recover compressible vectors $x \in \mathbb{R}^m$ (those which are close to sparse vectors with only few non-zero coordinates) from n pieces of incomplete linear information. In fact, vectors in the unit ball of ℓ_p^m with $0 < p \le 1$ serve as a good model for sparse vectors. In this context, also the recovery of vectors in ℓ_p^m with $0 < p \le 1$ or the Gelfand numbers of the identity $\ell_p^m \hookrightarrow \ell_q^m$ were studied. Donoho [10] and Foucart, Pajor, Rauhut, and T. Ullrich [13, Theorem 1.1] extended the Garnaev-Gluskin result also to the case $0 < p \le 1$ and

$p < q \leq 2$ and proved

$$\mathbb{E}\big[r(N_{n,m})\big] \asymp r(N_{n,m}^*) \asymp \min\left\{1, \left(\frac{\log(1 + \frac{m}{n})}{n}\right)^{1/p - 1/q}\right\}.$$

Again, it is not known how two explicitly construct measurement maps performing almost as good as $N_{n,m}^*$.

A further problem in compressed sensing is the matrix recovery problem. The analogon to the recovery problem of vectors from ℓ_p^m in the ℓ_q^m-norm is the recovery problem of matrices from Schatten p-classes in the Schatten q-norm. The particular case $q = 2$ is is the Frobenius norm. This problem was studied in [5, 21].

3.2. **Recovery of vectors from an ellipsoid.** We continue with another recovery problem and recent results of the authors concerning the quality of random information for ℓ_2-approximation, see [18].

Let F be a centered ellipsoid in $\mathbb{R}^m$ with semi-axes $\sigma_1 \geq \cdots \geq \sigma_m$. Clearly, F can be seen as the unit ball of some Hilbert space. We want to recover $x \in F$ from the data $N_n(x) \in \mathbb{R}^n$ with linear information mapping $N_n \in \mathbb{R}^{n \times m}$ and measure the error in the Euclidean distance. While the power of the information mapping is

$$r(N_n) = \inf_{\varphi: \mathbb{R}^n \to \mathbb{R}^m} \sup_{x \in F} \|\varphi(N_n(x)) - x\|_2,$$

it is known that, for linear problems in Hilbert spaces, the worst data is the zero data and so

$$r(N_n) = \sup_{x \in F \cap E_n} \|x\|_2,$$

where E_n is the kernel of N_n, see [29, 36]. In this case, the power or radius of optimal information is given by $r(N_n^*) = \sigma_{n+1}$. The question is now how good random information N_n is for the ℓ_2-recovery problem, when random information is provided by a Gaussian random matrix $G = (g_{ij})_{i,j=1}^{n,m}$? Is it comparable to optimal information?

As is explained in [18], rephrased in geometric terms, one is interested in the circumradius of the intersection of a centered ellipsoid F in $\mathbb{R}^m$ with a random subspace E_n of codimension n. While the maximal radius is the length of the largest semi-axis σ_1, the minimal radius is the length of the $(n+1)^{\text{st}}$ largest semi-axis σ_{n+1}. But how large is the radius of a typical intersection? Is it comparable to the minimal or the maximal radius or does it behave completely different?

Using various probabilistic tools, such as exponential estimates for sums of chi-squared random variables, Gordon's min-max theorem for Gaussian processes, or estimates for the extreme singular values of (structured) Gaussian matrices, the authors obtained upper and lower bounds on the radius of Gaussian information for the ℓ_2-approximation problem which hold with overwhelming probability. In many cases these bounds are sharp up to absolute constants and show that random information is comparable to optimal information.

Let us only present one case which is of particular importance, where the semi-axes behave like the singular values of Sobolev embeddings [18, Corollary].

Theorem 16 (Hinrichs, Krieg, Novak, Prochno, Ullrich). *For $k \in \mathbb{N}$ let*

$$\sigma_k \asymp k^{-\alpha} \ln^{-\beta}(k+1),$$

where $\alpha \in (0, \infty)$ and $\beta \in \mathbb{R}$. Then

$$\mathbb{E}[r(N_n)] \asymp \begin{cases} \sigma_1 & \text{for } n < c_m, \quad \text{if } \alpha < 1/2 \ \text{ or } \ \beta \leq \alpha = 1/2, \\ \sigma_{n+1}\sqrt{\ln(n+1)} & \text{for } n < \sqrt{m}, \ \text{if } \beta > \alpha = 1/2, \\ \sigma_{n+1} & \text{for } n < m, \quad \text{if } \alpha > 1/2, \end{cases}$$

with

$$c_m = \begin{cases} c\,m^{1-2\alpha}\ln^{-\max\{2\beta,0\}} m & \text{for } \alpha < 1/2, \\ c\ln^{1-\max\{2\beta,0\}} m & \text{for } \beta < \alpha = 1/2, \\ c\ln\ln m & \text{for } \beta = \alpha = 1/2, \end{cases}$$

where $c \in (0, \infty)$ is an absolute constant.

This means that random information is just as good as optimal information if the singular values decay with a polynomial rate greater than $1/2$. Then, in geometric terms, the size of a typical intersection ellipsoid is comparable to the size of the smallest intersection. On the other hand, if the singular values decay too slowly, random information is rather useless, which means that a typical intersection ellipsoid is almost as large as the largest.

The ellipsoid recovery problem can be considered in an infinite dimensional setting as well, even though our geometric intuition might fail. In this case, where $m = \infty$, we consider the ellipsoid

$$F = \left\{ x \in \ell_2 \colon \sum_{j \in \mathbb{N}} \frac{x_j^2}{\sigma_j^2} \leq 1 \right\}$$

and the matrix $(g_{ij})_{1 \leq i \leq n, j \in \mathbb{N}}$ with independent standard Gaussian entries as random information mapping N_n.

As a consequence of the probabilistic estimates derived in [18, Theorems 3 and 4], one obtains an ℓ_2-dichotomy showing that random information is useful if and only if $\sigma \in \ell_2$ [18, Corollary 5].

Theorem 17 (Hinrichs, Krieg, Novak, Prochno, Ullrich). *If $\sigma \notin \ell_2$, then $r(N_n) = \sigma_1$ holds almost surely for all $n \in \mathbb{N}$. If $\sigma \in \ell_2$, then*

$$\lim_{n \to \infty} \sqrt{n}\,\mathbb{E}[r(N_n)] = 0.$$

Remark 18. Such an ℓ_2-dichotomy is known from a related problem. There F is the unit ball of a reproducing kernel Hilbert space H, i.e., $H \subset L_2(D)$ consists of functions on a common domain D and the evaluation functional $f \mapsto f(x)$ is a bounded operator on H for every $x \in D$. The optimal linear information N_n for the L_2-approximation problem is

given by the singular value decomposition and has radius σ_{n+1}. This information might be difficult to implement and hence one might allow only information N_n of the form

$$N_n(f) = \big(f(x_1), \ldots, f(x_n)\big), \quad x_1, \ldots, x_n \in D.$$

The goal is to relate the power of function evaluations to the power of all continuous linear functionals and one would like to prove that their power is roughly the same. Unfortunately, this is *not* true in general. When $\sigma \notin \ell_2$ the convergence of optimal algorithms that may only use function values can be arbitrarily slow [19]. The situation is much better if we assume that $\sigma \in \ell_2$. It was shown in [26] and [37] that function values are almost as good as general linear information. We refer to [32, Chapter 26] for a presentation of these results.

Acknowledgement

References

[1] F. Bach. On the equivalence between kernel quadrature rules and random feature expansions. *J. Mach. Learn. Res.*, 18: Paper No. 21, 38, 2017.

[2] B. Bauer, L. Devroye, M. Kohler, A. Krzyżak, and H. Walk. Nonparametric estimation of a function from noiseless observations at random points. *J. Multivariate Anal.*, 160: 93–104, 2017.

[3] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2): 489–509, 2006.

[4] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. *Interactions between compressed sensing random matrices and high dimensional geometry*, volume 37 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2012.

[5] J. A. Chávez-Domínguez and D. Kutzarova. Stability of low-rank matrix recovery and its connections to Banach space geometry. *J. Math. Anal. Appl.*, 427(1): 320–335, 2015.

[6] A. Chkifa, N. Dexter, H. Tran, and Webster C. G. Polynomial approximation via compressed sensing of high-dimensional functions on lower sets. *Math. Comp.*, 87(311): 1415–1450, 2018.

[7] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22(1): 211–231, 2009.

[8] A. Cohen, R. DeVore, S. Foucart, and H. Rauhut. Recovery of functions of many variables via compressive sensing. In *Proc. SampTA 2011*, Singapore, 2010.

[9] D. A. Darling. On a class of problems related to the random division of an interval. *Ann. Math. Statistics*, 24: 239–253, 1953.

[10] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4): 1289–1306, 2006.

[11] D. Dung, V.T. Temlyakov, and T. Ullrich. *Hyperbolic Cross Approximation*. Advanced Courses in Mathematics - CRM Barcelona. Springer International Publishing, 2018.

[12] M. Ehler, M. Graef, and C. J. Oates. Optimal Monte Carlo integration on closed manifolds. *arXiv e-prints*, page arXiv: 1707.04723, Jul 2017. Statistics and Computing, to appear.

[13] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich. The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$. *J. Complexity*, 26(6): 629–640, 2010.

[14] A. Yu. Garnaev and E. D. Gluskin. The widths of a Euclidean ball. *Dokl. Akad. Nauk SSSR*, 277(5): 1048–1052, 1984. English translation: Soviet Math. Dokl. 30 (1984), no. 1, 200–204.

[15] E. D. Gluskin. On some finite-dimensional problems of width theory. *Physis—Riv. Internaz. Storia Sci.*, 23(2): 5–10, 124, 1981.

[16] E. D. Gluskin. Norms of random matrices and diameters of finite-dimensional sets. *Mat. Sb. (N.S.)*, 120(162)(2): 180–189, 286, 1983.

[17] S. Heinrich, E. Novak, G. W. Wasilkowski, and H. Woźniakowski. The inverse of the star-discrepancy depends linearly on the dimension. *Acta Arith.*, 96(3): 279–302, 2001.

[18] A. Hinrichs, D. Krieg, E. Novak, J. Prochno, and M. Ullrich. Random sections of ellipsoids and the power of random information. *arXiv e-prints*, page arXiv: 1901.06639, Jan 2019.

[19] A. Hinrichs, E. Novak, and J. Vybíral. Linear information versus function evaluations for L_2-approximation. *J. Approx. Theory*, 153(1): 97–107, 2008.

[20] A. Hinrichs, J. Oettershagen, and M. Ullrich. Numerical integration of smooth functions using random point sets. In preparation.

[21] A. Hinrichs, J. Prochno, and J. Vybíral. Entropy numbers of embeddings of Schatten classes. *J. Funct. Anal.*, 273(10): 3241–3261, 2017.

[22] M. Kanagawa, B. K. Sriperumbudur, and K. Fukumizu. Convergence guarantees for kernel-based quadrature rules in misspecified settings. *arXiv e-prints*, page arXiv: 1605.07254, May 2016. 29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

[23] B. S. Kašin. Kolmogorov diameters of octahedra. *Dokl. Akad. Nauk SSSR*, 214: 1024–1026, 1974.

[24] B. S. Kašin. The widths of certain finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat.*, 41(2): 334–351, 478, 1977. English translation: Math. USSR-Izv. 11 (1977), no. 2, 317–333 (1978).

[25] D. Krieg. *Algorithms and Complexity of some Multivariate Problems*. PhD thesis, Friedrich Schiller University Jena, 2019.

[26] F. Y. Kuo, G. W. Wasilkowski, and H. Woźniakowski. On the power of standard information for multivariate approximation in the worst case setting. *J. Approx. Theory*, 158(1): 97–125, 2009.

[27] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[28] E. Novak and H. Triebel. Function spaces in Lipschitz domains and optimal rates of convergence for sampling. *Constr. Approx.*, 23(3): 325–350, 2006.

[29] E. Novak and H. Woźniakowski. *Tractability of multivariate problems. Vol. 1: Linear information*, volume 6 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2008.

[30] E. Novak and H. Woźniakowski. Approximation of infinitely differentiable multivariate functions is intractable. *J. Complexity*, 25(4): 398–404, 2009.

[31] E. Novak and H. Woźniakowski. *Tractability of multivariate problems. Volume II: Standard information for functionals*, volume 12 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2010.

[32] E. Novak and H. Woźniakowski. *Tractability of multivariate problems. Volume III: Standard information for operators*, volume 18 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2012.

[33] I. H. Sloan and H. Woźniakowski. When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals? *J. Complexity*, 14(1): 1–33, 1998.

[34] A. G. Sukharev. An optimal method for the construction of the best uniform approximations for a certain class of functions. *Ž. Vyčisl. Mat. i Mat. Fiz.*, 18(2): 302–313, 523, 1978. English translation: U.S.S.R. Comput. Math. Math. Phys. 18 (1978), no. 2, 21–31.

[35] A. G. Sukharev. Optimal formulas of numerical integration for some classes of functions of several variables. *Dokl. Akad. Nauk SSSR*, 246(2): 282–285, 1979. English translation: Soviet Math. Dokl. 20 (1979), no. 3, 472–475.

[36] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information-based complexity*. Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA, 1988. With contributions by A. G. Werschulz and T. Boult.

[37] G. W. Wasilkowski and H. Woźniakowski. On the power of standard information for weighted approximation. *Found. Comput. Math.*, 1(4): 417–434, 2001.

[38] H. Weyhausen. *Expected Discrepancies*. PhD thesis, Friedrich Schiller University Jena, 2015.

(A. Hinrichs, M. Ullrich) Institut für Analysis, Johannes Kepler Universität Linz, Altenbergerstrasse 69, 4040 Linz, Austria

(D. Krieg, E. Novak) Mathematisches Institut, Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany

(J. Prochno) Institut für Mathematik & Wissenschaftliches Rechnen, Karl-Franzens-Universität Graz, Heinrichstrasse 36, 8010 Graz, Austria