# On the Power of Function Values for the Approximation Problem in Various Settings

Erich Novak and Henryk Woźniakowski

June 2, 2011

**Abstract**

This is an expository paper on approximating functions from general Hilbert or Banach spaces in the worst case, average case and randomized settings with error measured in the $L_p$ sense. We define the *power function* as the ratio between the best rate of convergence of algorithms that use function values over the best rate of convergence of algorithms that use arbitrary linear functionals for a worst possible Hilbert or Banach space for which the problem of approximating functions is well defined. Obviously, the power function takes values at most one. If these values are one or close to one than the power of function values is the same or almost the same as the power of arbitrary linear functionals. We summarize and supply a few new estimates on the power function. We also indicate eight open problems related to the power function since this function has not yet been studied in many cases. We believe that the open problems will be of interest to a general audience of mathematicians.

MSC: 41A25, 41A46, 65Y20

Keywords: degree of approximation, widths and sampling numbers, complexity of numerical algorithms



Surveys in Approximation Theory
Volume 6, 2011. pp. 1–23.

ISSN 1555-578X

## 1 Introduction

This is an expository paper on the problem of approximating functions from general Hilbert or Banach spaces, which has been thoroughly studied in many books and papers. This problem has many variants depending on how we measure the error of such approximations (algorithms). A popular choice is to take the norm of an $L_p$ space and all values of $p \in [1, \infty]$ have been considered. Furthermore, the error of algorithms can be defined in the worst case, average case or randomized setting. For the worst and average case settings, we consider deterministic algorithms. The worst case error is defined as the maximal error over the unit ball of a given space whereas the average case error is defined as the average error over the whole space with respect to a given measure. The usual choice is a zero mean Gaussian measure. For the randomized setting we consider randomized algorithms and the error is defined as the maximal expected error over the unit ball of a given space. Here, the expected error is given with respect to a probability distribution of randomized elements.

We approximate functions $f$ by algorithms that use information about $f$ given by finitely many functionals of $f$. Information is called *linear* if we can choose arbitrary linear functionals, and it is called *standard* if only function values may be used. Clearly, linear information is at least as powerful as standard information. For many applications, only standard information is available. But even in this case, it is a good idea to study linear information and learn how difficult is the function approximation problem. For example, if we can prove that even for linear information the problem is too difficult then, obviously, the same also holds for standard information. On the other hand, all positive results for linear information do not have to hold for standard information.

The main question addressed in this expository paper is the study of the power of standard information or equivalently the power of function values. We want to know how much we lose if function values are used instead of linear information. Or more optimistically, we ask when the power of standard information is the same or nearly the same as the power of linear information. Such questions have been addressed in a number of papers and we will refer to them in the course of this paper. It has been usually done for specific spaces and only a few papers addressed these questions for some classes of spaces.

Our approach is a little more general and we want to verify the power of function values/standard information for all Hilbert or Banach spaces for which the problem of function approximation is well defined. More precisely, we define the *power* function[1]

$$\ell^{\mathrm{sett-x}} \,:\, (0,\infty) \times [1,\infty] \to [0,1].$$

Here sett $\in \{\mathrm{wor}, \mathrm{ran}, \mathrm{avg}\}$ denotes the setting we use for the error definition. Hence, wor stands for the worst case setting, ran for the randomized setting, and avg for the average case setting. The second superscript $\mathrm{x} \in \{H, B\}$ tells us if we consider only Hilbert spaces ($\mathrm{x} = H$) or if we allow all Banach spaces ($\mathrm{x} = B$).

We now explain the meaning of the value

$$\ell^{\mathrm{sett-x}}(r,p).$$

The first argument $r$ means that the $n$th minimal error (formally defined in Definition 1) behaves like $n^{-r}$ if we use linear information. Since $r > 0$, we consider Hilbert or Banach spaces which

[1] We needed to find a good one-letter name for the power function. Since in English and in Polish this would indicate the letter "$p$" which is already used as the parameter of the $L_p$ space, we turn to German and use the word "Leistung". That is why the letter $\ell$ denotes the power function.

admit convergence, and furthermore they admit a polynomial rate of convergence of the minimal errors. The second argument $p$ denotes the use of the norm of $L_p$. The value $\ell^{\,\mathrm{sett-x}}(r,p)$ is defined as $r^{-1}$ times the best rate of convergence we obtain using only function values for a worst possible choice of a Hilbert or Banach space. That is why $\ell^{\,\mathrm{sett-x}}(r,p)\le 1$, and the larger $\ell^{\,\mathrm{sett-x}}(r,p)$ the better. Hence, if we have

$$\ell^{\,\mathrm{sett-x}}(r,p)=1$$

then the power of standard information is the same as the power of linear information. We will see later that this does happen in some cases. Then standard information yields the same rate of convergence as linear information for the embeddings $I:F\to L_p$ for *all* Hilbert (if $x=H$) or Banach (if $x=B$) spaces without the need of a case to case study for each $F$. This holds in the randomized setting for Hilbert spaces with $p=2$, see Theorem 6, and in the average case setting for Banach spaces equipped with zero mean Gaussian measures and $p=2$, see Theorem 8. It is open if $\ell^{\,\mathrm{sett-x}}(r,p)=1$ may happen in the worst case setting, see Open Problem 1.

On the other hand, if we have

$$\ell^{\,\mathrm{sett-x}}(r,p)=0$$

then the power of standard information is zero as compared to the power of linear information. Finally, if we have

$$\ell^{\,\mathrm{sett-x}}(r,p)\in(0,1)$$

then we know qualitatively how much we may lose by using function values.

The concept of the power function seems to be new. For many values of $p$, especially when $p\ne 2$, this function has not yet been studied. This is especially the case for the randomized and average case settings. That is why we indicate eight open problems related to the power function with the hope that many mathematicians will be interested in solving them and advancing our knowledge about the power of function values.

In this paper, we tried to summarize and supply a few new estimates on the power function. We now briefly indicate a few results presented in the paper.

In the worst case setting for the Hilbert case and $p=2$, we conclude from [6, 8] that

$$\begin{aligned}
\ell^{\,\mathrm{wor-H}}(r,2) &= 0 && \text{for all}\quad r\in(0,\tfrac12],\\
\ell^{\,\mathrm{wor-H}}(r,2) &\in \left[\frac{2r}{2r+1},1\right] && \text{for all}\quad r\in(\tfrac12,\infty).
\end{aligned}$$

Hence, the power of function values is zero for $r\le 1/2$, and almost the same as the power of linear information for large $r$. One of the main open problem is to verify whether $\ell^{\,\mathrm{wor-H}}(r,2)=1$ for all $r>1/2$.

Staying with the worst case and Hilbert spaces but with $p\ne 2$, we conclude from [18] that

$$\ell^{\,\mathrm{wor-H}}(r,p)=0\quad\text{for all}\quad r\in\left(0,\min(\tfrac1p,\tfrac12)\right].$$

For $r>\min(1/p,1/2)$, we do not know anything about the values of $\ell^{\,\mathrm{wor-H}}(r,p)$ except the case $p=\infty$ for which we know from [12] that

$$\ell^{\,\mathrm{wor-H/B}}(r,\infty)\ge 1-\frac{1}{r}.$$

By H/B we mean that we obtain this result for both Hilbert and Banach spaces. Again for large $r$, the power of standard information is almost the same as the power of linear information.

For the worst case and the Banach case, we have

$$\begin{array}{rcll}
\ell^{\,\mathrm{wor-B}}(r,p) & = & 0 & \text{for all} \quad r\in(0,1] \quad \text{and} \quad p\in[1,2],\\
\ell^{\,\mathrm{wor-B}}(r,p) & = & 0 & \text{for all} \quad r\in(0,\tfrac12+\tfrac1p] \quad \text{and} \quad p\in(2,\infty),\\
\ell^{\,\mathrm{wor-B}}(r,p) & \le & 1-\dfrac1r\left(1-\dfrac1p\right) & \text{for all} \quad r>1 \quad \text{and} \quad p\in[1,2],\\
\ell^{\,\mathrm{wor-B}}(r,p) & \le & 1-\dfrac1{2r} & \text{for all} \quad r>1 \quad \text{and} \quad p\in[2,\infty),\\
1-\dfrac1r \le \ell^{\,\mathrm{wor-B}}(r,\infty) & \le & 1-\dfrac1{2r} & \text{for all} \quad r>1,
\end{array}$$

see Theorem 5 in Section 2.3.

Even though we do not know much about the power function in this case, we can conclude that the Hilbert and Banach cases are different since

$$\ell^{\,\mathrm{wor-B}}(r,2) < \ell^{\,\mathrm{wor-H}}(r,2) \quad \text{for all} \quad r\in(\tfrac12,\infty).$$

Surprisingly enough, for the randomized setting with the Hilbert case and for the average case setting with the Hilbert or Banach case we have complete knowledge about the power function for $p=2$ due to [28] and [5]. More precisely, we know that

$$\ell^{\,\mathrm{ran-H}}(r,2) = \ell^{\,\mathrm{avg-H/B}}(r,2) = 1 \quad \text{for all} \quad r>0.$$

More estimates of the power function can be found in the subsequent sections.

## 2 Worst case setting

Let $F$ be a Hilbert or Banach space of functions, defined on a set $\Omega$, such that the linear functionals $f\mapsto f(x)$ are continuous for all $x\in\Omega$. We assume that $F\subset L_p$ and that the embedding $I:F\to L_p$ is continuous[2], where $I(f)=f$. We write $H$ instead of $F$ if $F$ is a Hilbert space.

Let $(c_n)$ be a sequence of nonnegative numbers. Assume first that $(c_n)$ converges to zero. We define its (polynomial) rate of convergence $r(c_n)$ by

$$r(c_n) = \sup\{\,\beta\ge0 \mid \lim_{n\to\infty} c_n n^\beta = 0\}.$$

If $(c_n)$ is not convergent to zero, we set $r(c_n)=0$. Then $r(c_n)$ is well defined for all nonnegative sequences $(c_n)$. For example, the rate of convergence of $n^{-\alpha}$ is $\max(0,\alpha)$.

We approximate functions from $F$ using finitely many arbitrary linear functionals $L\in F^*$ or function values $f(x)$ for some $x\in\Omega$. We define the error of such approximations by taking the worst case setting with respect to the $L_p$ norm. The norm of $L_p$ is denoted by $\|\cdot\|_p$.

We define two classes $\Lambda^{\mathrm{all}}$ and $\Lambda^{\mathrm{std}}$ of information evaluations. We have $\Lambda^{\mathrm{std}}\subseteq\Lambda^{\mathrm{all}}=F^*$ and $\Lambda^{\mathrm{std}}$ consists of linear functionals of the form $L_x(f)=f(x)$ for all $f\in F$, where $x\in\Omega$. We approximate functions from $F$ by algorithms $A_n:F\to L_p$ given by

$$A_n(f) = \phi_n(L_1(f),L_2(f),\dots,L_n(f)),$$

[2] We do not specify $\Omega$ or the underlying measure of $L_p$ since they can be arbitrary.

where $n$ is a nonnegative integer, $\phi_n : \mathbb{R}^n \to L_p$ is an arbitrary mapping, and $L_j \in \Lambda$, where $\Lambda \in \{\Lambda^{\rm all}, \Lambda^{\rm std}\}$. The choice of $L_j$ can be adaptive, that is, $L_j(\cdot) = L_j(\cdot; L_1(f), L_2(f), \dots, L_{j-1}(f))$ may depend on the already computed values $L_1(f), L_2(f), \dots, L_{j-1}(f)$. For $n = 0$, $A_n(f)$ equals some fixed element of the space $L_p$. More details can be found in e.g., [14, 21].

Hence, we consider algorithms that use $n$ linear functionals either from the class $\Lambda^{\rm std}$ or from the class $\Lambda^{\rm all}$. We define the minimal errors as follows.

**Definition 1.** For $n = 0$ and $n \in \mathbb{N} := \{1, 2, \dots\}$, let

$$e_n^{\rm all-wor}(F, L_p) := \inf_{A_n \text{ with } L_j \in \Lambda^{\rm all}} \sup_{\|f\|_F \le 1} \big\| f - A_n(f) \big\|_p$$

and

$$e_n^{\rm std-wor}(F, L_p) := \inf_{A_n \text{ with } L_j \in \Lambda^{\rm std}} \sup_{\|f\|_F \le 1} \big\| f - A_n(f) \big\|_p.$$

For $n = 0$, it is easy to see that the best algorithm is $A_0(f) = 0$ and we obtain

$$e_0^{\rm all-wor}(F, L_p) = e_0^{\rm std-wor}(F, L_p) = \sup_{\|f\|_F \le 1} \|f\|_p = \sup_{\|f\|_F \le 1} \|I(f)\|_p = \|I\|.$$

This is the initial error that can be achieved without computing any linear functional on the functions $f$. Clearly,

$$e_n^{\rm all-wor}(F, L_p) \le e_n^{\rm std-wor}(F, L_p) \quad \text{for all} \quad n \in \mathbb{N}.$$

The sequences $\big(e_n^{\rm all-wor}(F, L_p)\big)$ and $\big(e_n^{\rm std-wor}(F, L_p)\big)$ are both non-increasing but not necessarily convergent to zero.

We want to compare the rates of convergence

$$r^{\rm all-wor}(F, L_p) := r\left(e_n^{\rm all-wor}(F, L_p)\right) \qquad \text{and} \qquad r^{\rm std-wor}(F, L_p) := r\left(e_n^{\rm std-wor}(F, L_p)\right).$$

In particular, we would like to know if it is possible that the sequence $\big(e_n^{\rm all-wor}(F, L_p)\big)$ converges to zero much faster than the sequence $\big(e_n^{\rm std-wor}(F, L_p)\big)$. In many cases it is much easier to analyze the sequence $(e_n^{\rm all-wor}(F, L_p))_{n \in \mathbb{N}}$. It is then natural to ask what can be said about the sequence $(e_n^{\rm std-wor}(F, L_p))_{n \in \mathbb{N}}$.

The main question addressed in this paper is to find or estimate the *power* function defined as $\ell^{\rm wor-x} : (0, \infty) \times [1, \infty] \to [0, 1]$ by

$$\ell^{\rm wor-x}(r, p) := \inf_{F:\, r^{\rm all-wor}(F, L_p) = r} \frac{r^{\rm std-wor}(F, L_p)}{r},$$

where $\mathrm{x} \in \{H, B\}$ and indicates that the infimum is taken over all Hilbert spaces ($\mathrm{x} = H$) or over all Banach spaces ($\mathrm{x} = B$) continuously embedded in $L_p$ for which function values are continuous linear functionals and the rate of convergence is $r$ when we use arbitrary linear functionals.

It is easy to show, and it will be shown later, that the set of spaces $F$ for which $r^{\rm all-wor}(F, L_p) = r$ is not empty and therefore $\ell^{\rm wor-x}$ is well defined. Obviously, $\ell^{\rm wor-x}(r, p) \in [0, 1]$, as already claimed. The power function $\ell^{\rm wor-x}$ measures the ratio between the best rates of convergence of approximations based on function values over those based on arbitrary linear functionals for a worst possible Hilbert or Banach space.

We briefly comment on why we take the infimum over $F$ in the definition of the power function. For some specific spaces $F$, standard information is as powerful as linear information[3]. But this is a property of $F$, not the indication of the power of standard information. By taking the infimum with respect to $F$, we concentrate on the power of standard information as compared to the power of linear information.

Suppose now that we take the minimal $n = n^{\mathrm{wor-all/std}}(\varepsilon, F, L_p)$ for which the minimal worst case error is $\varepsilon$ or $\varepsilon\,\|I\|$. Assume for simplicity that

$$e_n^{\mathrm{all-wor}}(F, L_p) = n^{-r} \quad \text{and} \quad e_n^{\mathrm{std-wor}}(F, L_p) = n^{-\alpha}$$

for some positive $\alpha = r^{\mathrm{std-wor}}(F, L_p) \le r$. Then

$$n^{\mathrm{wor-all}}(\varepsilon, F, L_p) = \left\lceil \varepsilon^{-1/r} \right\rceil \quad \text{and} \quad n^{\mathrm{wor-std}}(\varepsilon, F, L_p) = \left\lceil \varepsilon^{-1/\alpha} \right\rceil.$$

Clearly,

$$\lim_{\varepsilon\to 0} \frac{\ln n^{\mathrm{wor-all}}(\varepsilon, F, L_p)}{\ln n^{\mathrm{wor-std}}(\varepsilon, F, L_p)} = \frac{\alpha}{r} \ge \ell^{\mathrm{wor-x}}(r, p).$$

Hence, if $\ell^{\mathrm{wor-x}}(r,p) = 1$ then function values are as powerful as arbitrary linear functionals. On the other hand, the smaller $\ell^{\mathrm{wor-x}}(r,p)$ the less powerful are function values as compared to arbitrary linear functionals. If $\ell^{\mathrm{wor-x}}(r,p) = 0$ then the polynomial behavior of $n^{\mathrm{all}}(\varepsilon, F, L_p)$ in $\varepsilon^{-1}$ can be drastically changed for $n^{\mathrm{std}}(\varepsilon, F, L_p)$.

**Remark 1.** It is well known that, in some cases, we can restrict ourselves only to linear algorithms. This holds when $p = \infty$ or when $F$ is a Hilbert space. Then the corresponding infima for the minimal worst case errors are attained by

$$A_n(f) = \sum_{j=1}^{n} L_j(f) h_j$$

for some $L_j \in \Lambda \in \{\Lambda^{\mathrm{std}}, \Lambda^{\mathrm{all}}\}$ and $h_j \in L_p$. Much more about the existence of linear optimal error algorithms can be found in e.g., [14].

### 2.1 Double Hilbert Case

In this subsection, we consider the approximation problem defined over a Hilbert space with the error measured also in the Hilbert space $L_2$. That is why the name of this subsection is the double Hilbert case. Approximation in the $L_2$ norm for Hilbert spaces has been studied in many papers. For our problem the most relevant papers are [6], [8] and [27].

Assume that $H$ is a Hilbert space of functions defined on a set $\Omega$. Since we assume that function values are continuous this means that $H$ is a reproducing kernel Hilbert space, $H = H(K)$, where

[3]This holds with $r = \infty$ for all finite dimensional spaces $F$. This also holds for some infinite dimensional Hilbert spaces $F$. For example, take $F$ as the space of piecewise constant functions over, say, $I_j := [1/(j+1), j)$ for $j = 1, 2, \dots$. The inner product of $F$ is chosen such that the functions $e_j$ equal to 1 over $I_j$ are orthonormal. Then the algorithm $A_n(f) = \sum_{j=1}^{n} \langle f, e_j \rangle_F \, e_j$ minimizes the worst case error for all $L_p$ with $p \in [1, \infty)$. The error is $[n(n+1)]^{-1/p}$. Since $\langle f, e_j \rangle_F = f(1/(j+1))$, we may say that this algorithm uses standard information. Therefore $r^{\mathrm{all-wor}}(F, L_p) = r^{\mathrm{std-wor}}(F, L_p) = 2/p$.

$K$ is defined on $\Omega \times \Omega$. Let $L_2 = L_2(\Omega,\mu)$ be the space of $\mu$-square integrable functions with a measure $\mu$ on $\Omega$. Since the embedding $I : H(K) \to L_2(\Omega,\mu)$ is continuous, we have

$$\int_\Omega |f(t)|^2 \,\mathrm{d}\mu(t) < \infty \quad \text{for all} \quad f \in H(K).$$

In particular, we can take $f = K(\cdot,t)$ for arbitrary $t \in \Omega$, since such a function $f$ belongs to $H(K)$. Therefore $W = I^*I : H(K) \to H(K)$, where $I^*$ is defined by $\langle g, I(f)\rangle_{L_2(\Omega,\mu)} = \langle I^*(g), f\rangle_{H(K)}$ for all $f \in H(K)$ and $g \in L_2(\Omega,\mu)$, is given by

$$W(f)\,(x) = \int_\Omega K(x,t)\,f(t)\,\mathrm{d}\mu(t) \quad \text{for all} \quad f \in H(K).$$

The operator $W$ is self-adjoint and positive semi-definite. It is well known that

$$\lim_n e_n^{\mathrm{wor-all}}(H, L_2) = 0$$

if and only if $W$ is compact, see, e.g., [14, Section 4.2.3]. Unfortunately, in general, $W$ needs not be compact and therefore $e_n^{\mathrm{wor-all}}(H, L_2)$ does not have to go to zero. In fact, the sequence $e_n^{\mathrm{wor-all}}(H, L_2)$ can be an arbitrary non-increasing sequence as the following example shows.

**Example 1** (Arbitrary Sequence $e_n^{\mathrm{wor-all}}(H, L_2)$)**.**

Let $(\alpha_n)_{n\in\mathbb{N}}$ be an arbitrary non-increasing sequence of nonnegative numbers. Define $k^*$ as the number of positive $\alpha_n$. If all $\alpha_n$ are positive, we formally set $k^* = \infty$. If $k^*$ is finite let $\mathbb{N}_{k^*} = \{1, 2, \dots, k^*\}$, otherwise let $\mathbb{N}_{k^*} = \mathbb{N}$.

For $k \in \mathbb{N}_{k^*}$, take arbitrary disjoint intervals $I_k$ of positive Lebesgue measure $|I_k|$ such that $\bigcup_{k\in\mathbb{N}_{k^*}} I_k = [0,1]$, and define the functions $e_k : [0,1] \to \mathbb{R}$ by

$$e_k = \frac{\sqrt{\alpha_k}}{\sqrt{|I_k|}}\; 1_{I_k},$$

where $1_{I_k}$ is the indicator function of $I_k$. That is, $e_k(x) = \sqrt{\alpha_k/|I_k|}$ for $x \in I_k$ and $e_k(x) = 0$ for $x \notin I_k$.

Define the Hilbert space $H = \overline{\mathrm{span}}\{e_k \mid k \in \mathbb{N}_{k^*}\}$ equipped with the inner product such that $\langle e_k, e_j\rangle_H = \delta_{k,j}$ for all $k, j \in \mathbb{N}_{k^*}$. This means that $H$ is the space of piecewise constant functions $f : [0,1] \to \mathbb{R}$ such that

$$f = \sum_{k=1}^{k^*} a_k e_k \quad \text{with} \quad a_k = \langle f, e_k\rangle_H \quad \text{and} \quad \|f\|_H = \left(\sum_{k=1}^{k^*} a_k^2\right)^{1/2} < \infty.$$

The Hilbert space $H$ has the reproducing kernel

$$K(x,y) = \sum_{k=1}^{k^*} e_k(x) e_k(y) \quad \text{for all} \quad x, y \in [0,1].$$

Indeed, first of all note that $K$ is well defined since for all $x$ and $y$ the last series has at most one nonzero term. Then $\langle K(\cdot, y_i), K(\cdot, y_j)\rangle_H = K(y_i, y_j)$, and

$$0 \le \left\| \sum_{j=1}^m a_j K(\cdot, y_j) \right\|_H^2 = \sum_{i,j=1}^m a_i a_j K(y_i, y_j).$$

This shows that the matrix $(K(y_i, y_j))_{i,j=1,2,\dots,m}$ is symmetric and positive semi-definite for all $m$ and $y_j$. Clearly,

$$\langle f, K(\cdot, y)\rangle_H = \sum_{k=1}^{k^*} a_k e_k(y) = f(y),$$

and this completes the proof of the fact that $K$ is the reproducing kernel of $H$.

Let $L_2 = L_2([0,1])$ be the usual space of square Lebesgue integrable functions. Note that

$$\|e_k\|_2 = \frac{\alpha_k}{\sqrt{|I_k|}} \left(\int_{I_k} \mathrm{d}t\right)^{1/2} = \alpha_k.$$

Therefore, for any $f \in H$, we have

$$\|I(f)\|_2 = \|f\|_2 = \left(\sum_{k=1}^{k^*} a_k^2 \alpha_k^2\right)^{1/2} \le \alpha_1 \|f\|_H.$$

The last bound is sharp, and therefore $\|I\| = \alpha_1$ showing that $H$ is continuously embedded in $L_2$. The operator $W$ takes now the form

$$W(f) = \sum_{k=1}^{k^*} \langle f, e_k\rangle_2 \; e_k.$$

Note that $W(e_k) = \|e_k\|_2^2 \, e_k = \alpha_k^2 \, e_k$. This means that $(\alpha_k^2, e_k)$ are the eigenpairs of $W$ and

$$W(f) = \sum_{k=1}^{k^*} \alpha_k^2 \; \langle f, e_k\rangle_H \; e_k.$$

It is well known that

$$e_n^{\mathrm{wor-all}}(H, L_2) = \alpha_{n+1} \quad \text{for all} \qquad n \in \mathbb{N},$$

see, e.g., [14, Section 4.2.3]. This proves that the behavior of $e_n^{\mathrm{wor-all}}(H, L_2)$ can be arbitrary and, in general, we do not have convergence of $e_n^{\mathrm{wor-all}}(H, L_2)$ to zero. Clearly, $W$ is compact if and only if $\lim_n \alpha_n = 0$.

In addition, this example also shows that for a given $\beta \ge 0$ we can define a sequence $\alpha_k$ such that $r^{\mathrm{all-wor}}(H, L_2) = \beta$. Indeed, it is enough to take $\alpha_k = k^{-\beta}$. □

We discuss the power function $\ell^{\mathrm{wor-H}}$. We now assume that $r^{\mathrm{all-wor}}(H, L_2) = r > 0$. In particular, we assume that the operator $W$ is compact. Then $W$ has eigenpairs $(\lambda_j, \eta_j)$,

$$W(\eta_j) = \lambda_j \eta_j \quad \text{for all} \;\; j = 1, 2, \dots,$$

with $\langle \eta_j, \eta_k\rangle_H = \delta_{j,k}$. Without loss of generality, we can order the eigenvalues $\lambda_j$ such that $\lambda_1 \ge \lambda_2 \ge \cdots$. For all $f \in H$, we have

$$\langle f, \eta_k\rangle_2 = \langle I(f), I(\eta_k)\rangle_2 = \langle f, W(\eta_k)\rangle_H = \lambda_k \, \langle f, \eta_k\rangle_H \, .$$

In particular, letting $f = \eta_j$, we conclude that the functions $\eta_j$ are also orthogonal in the space $L_2$.

As above, it is well known that

$$e_n^{\mathrm{wor-all}}(H, L_2) = \sqrt{\lambda_{n+1}} \quad \text{for all} \quad n \in \mathbb{N}.$$

If $(e_n^{\mathrm{wor-all}}(H, L_2))$ is convergent to zero then the same also holds for function values, i.e., $(e_n^{\mathrm{wor-std}}(H, L_2))$ is also convergent to zero. Indeed, we can reason as in Section 10.4 of [14] that all linear functionals can be approximated with an arbitrarily small error when we use function values, and then it is enough to remember that the error $\sqrt{\lambda_{n+1}}$ is achieved by a linear algorithm that uses the $n$ linear functionals $\langle f, \eta_j \rangle_{H(K)}$.

We have

$$\mathrm{trace}(W) := \sum_{j=1}^{\infty} \lambda_j = \int_{\Omega} K(x, x)\, \mathrm{d}\mu(x) = \sum_{n=0}^{\infty} \left[ e_n^{\mathrm{wor-all}}(H, L_2) \right]^2$$

and this is finite if $r^{\mathrm{all-wor}}(H, L_2) > \frac{1}{2}$. If $r^{\mathrm{all-wor}}(H, L_2) = \frac{1}{2}$ then $\sum_{n=0}^{\infty} \left[ e_n^{\mathrm{wor-all}}(H, L_2) \right]^2$ may be finite or infinite, and if $r^{\mathrm{all-wor}}(H, L_2) < \frac{1}{2}$ then $\sum_{n=0}^{\infty} \left[ e_n^{\mathrm{wor-all}}(H, L_2) \right]^2$ is infinite.

The result from [8] states that $r^{\mathrm{all-wor}}(H, L_2) = r > \frac{1}{2}$ implies

$$r^{\mathrm{std-wor}}(H, L_2) \ge r - \frac{r}{2r+1} = \frac{2r^2}{2r+1}.$$

The case $\sum_{n=0}^{\infty} \left[ e_n^{\mathrm{wor-all}}(H, L_2) \right]^2 = \infty$ was studied in [6]. It was shown that for any $r \in [0, \frac{1}{2}]$ there is a Hilbert space $H$ such that

$$r^{\mathrm{all-wor}}(H, L_2) = r \quad \text{and} \quad r^{\mathrm{std-wor}}(H, L_2) = 0.$$

These results give us the following bounds on the power function $\ell^{\,\mathrm{wor-H}}(\cdot, 2)$.

**Theorem 1** ([6, 8])**.**

$$\begin{aligned} \ell^{\,\mathrm{wor-H}}(r, 2) &= 0 && \textit{for all} \quad r \in (0, \tfrac{1}{2}], \\ \ell^{\,\mathrm{wor-H}}(r, 2) &\in \left[ \frac{2r}{2r+1}, 1 \right] && \textit{for all} \quad r \in (\tfrac{1}{2}, \infty). \end{aligned}$$

Although we do not know the power function $\ell^{\,\mathrm{wor-H}}(\cdot, 2)$ exactly, we know that there is a jump at $\frac{1}{2}$ since $\ell^{\,\mathrm{wor-H}}(r, 2) \ge 1/2$ for all $r > 1/2$. Note also that for large $r$, the values of $\ell^{\,\mathrm{wor-H}}(r, 2)$ are close to 1. This means that the power of function values for $r \in (0, \frac{1}{2})$ is zero, and is almost optimal for large $r$.

The problem of finding the exact values of $\ell^{\,\mathrm{wor-H}}(r, 2)$ for $r > \frac{1}{2}$ is one of the main open problems in the worst case setting. We know that many people, including the two of us, spent a lot of time trying to solve this problem but so far in vain. That is why we propose an open problem with the hope that it will soon be solved by the reader.

**Open Problem 1.** Suppose that $r > \frac{1}{2}$. Is it true that

$$\ell^{\,\mathrm{wor-H}}(r, 2) = 1\,?$$

If not, what are the values of $\ell^{\,\mathrm{wor-H}}(r, 2)$?

The rate of convergence neglects to distinguish between sequences that differ by a power of logarithms of $n$. Indeed, for $c_n = n^{-r}$ and $b_n = n^{-r}[\ln(n+1)]^\beta$ for a positive $r$ and an arbitrary $\beta$, we have $r(c_n) = r(b_n) = r$ independent of $\beta$. Obviously, for some standard spaces, we would like to know not only the rate but also a power of logarithms. We discuss this point in the next example, where we use the notation

$$c_n \asymp b_n$$

which means that there exist positive numbers $a_1$ and $a_2$ such that $a_1 \le c_n/b_n \le a_2$ for large $n$.

**Example 2** (Sobolev spaces, $p = 2$)**.**
a) For the standard Sobolev spaces $W_2^s([0,1]^d)$ with an arbitrary $s > 0$, which measures the total smoothness of functions, it is well known that

$$e_n^{\mathrm{all-wor}}(W_2^s([0,1]^d), L_2) \asymp n^{-s/d}.$$

Of course, in general, function values are not well defined in $W_2^s([0,1]^d)$. We must assume the embedding condition $2s > d$ and then function values are well defined and they are continuous linear functionals. Furthermore, it is known that

$$e_n^{\mathrm{all-wor}}(W_2^s([0,1]^d), L_2) \asymp e_n^{\mathrm{std-wor}}(W_2^s([0,1]^d), L_2) \asymp n^{-s/d},$$

see, e.g., [14] for a survey of such results.

b) For the Sobolev spaces $W_2^{r,\mathrm{mix}}([0,1]^d)$ with $r > 0$, which measures the smoothness of functions with respect to each variable, it is known that

$$e_n^{\mathrm{all-wor}}(W_2^{r,\mathrm{mix}}([0,1]^d), L_2) \asymp n^{-r} (\log n)^{(d-1)r},$$

see, e.g., [1, 11, 17, 19, 21, 26], where this result can be found in various generalities.

For function values, we must assume that $r > 1/2$, and then the best upper bound is

$$e_n^{\mathrm{std-wor}}(W_2^{r,\mathrm{mix}}([0,1]^d), L_2) = \mathcal{O}\left(n^{-r} (\log n)^{(d-1)(r+1/2)}\right),$$

see [17, 19, 22].

It is *not* known whether this extra power $(d-1)/2$ of logarithms is needed. It would be very interesting to verify whether

$$e_n^{\mathrm{all-wor}}(W_2^{r,\mathrm{mix}}([0,1]^d), L_2) \asymp e_n^{\mathrm{std-wor}}(W_2^{r,\mathrm{mix}}([0,1]^d), L_2)$$

holds also for this example. □

The examples in [6] use very irregular sequences $(e_n^{\mathrm{all-wor}}(H, L_2))$ and hence do not exclude a positive answer to the question in the next open problem.

**Open Problem 2.** Assume that $e_n^{\mathrm{all-wor}}(H, L_2) \asymp n^{-r}\,[\ln(n+1)]^\beta$ with arbitrary $r > 0$ and $\beta \in \mathbb{R}$. Is it true that this implies

$$e_n^{\mathrm{std-wor}}(H, L_2) \asymp e_n^{\mathrm{all-wor}}(H, L_2)?$$

## 2.2 Single Hilbert Case

In this short subsection, we mostly consider the approximation problem defined over a Hilbert space with the error measured in the non-Hilbert space $L_p$ for $p \neq 2$. That is why the name of this subsection is the single Hilbert case.

We report on a recent result of Tandetzky [18] who considered the approximation problem for arbitrary $p \in [1,\infty)$. He proved that for any $r \in (0, \min(\frac{1}{p}, \frac{1}{2})]$ there exists a Hilbert space $H$ continuously embedded in $L_p = L_p([0,1])$ such that

$$r^{\text{all}-\text{wor}}(H, L_p) = r \qquad \text{and} \qquad r^{\text{std}-\text{wor}}(H, L_p) = 0.$$

This result obviously implies that the power function is zero over $(0, \min(\frac{1}{p}, \frac{1}{2})]$. It seems to us that no example is known in the literature for a Hilbert space for which $e_n^{\text{all}-\text{wor}}(H, L_p)$ tends to zero faster than the sequence $e^{\text{std}-\text{wor}}(H, L_p)$ with the additional assumption that $r^{\text{all}-\text{wor}}(H, L_p) > \min(\frac{1}{p}, \frac{1}{2})$. This implies that we do not know the behavior of the power function over $(\min(\frac{1}{p}, \frac{1}{2}), \infty)$. We summarize our partial knowledge of the power function in the following theorem.

**Theorem 2** ([18])**.** *Let $p \neq 2$.*

$$\ell^{\,\text{wor}-\text{H}}(r,p) \quad = \quad 0 \qquad \textit{for all} \quad r \in (0, \min(\tfrac{1}{p}, \tfrac{1}{2})].$$

Only for the case $p = \infty$ do we know a little more about the behavior of the power function. In this case the rates are related as explained in the following theorem.

**Theorem 3** ([12])**.** *Let $F$ be a Hilbert or a Banach space. Then*

$$e_n^{\text{std}-\text{wor}}(F, L_\infty) \leq (1+n)\, e_n^{\text{all}-\text{wor}}(F, L_\infty) \quad \textit{for all} \quad n \in \mathbb{N}. \tag{1}$$

This inequality follows from Proposition 1.2.5, page 16, in [12], where it is stated for the Kolmogorov widths and also applies to the linear or Gelfand widths.

The inequality (1) cannot be improved even if we assume that $F$ is a Hilbert space. This follows from the following example.

**Example 3.** Take $F = H = \mathbb{R}^{n+1}$. That is, $f \in H$ is now defined on $\{1, 2, \ldots, n+1\}$ and can be identified with $f = [f_1, f_2, \ldots, f_{n+1}]$, where $f_i = f(i)$. The space $H$ is equipped with the inner product

$$\langle f, g\rangle_H = \left[\sum_{i=1}^{n+1} f_i\right] \left[\sum_{i=1}^{n+1} g_i\right] + \varepsilon \sum_{i=1}^{n+1} f_i g_i \quad \text{for all} \quad f, g \in H.$$

The unit ball of $H$ is thus

$$B = \left\{ f \in \mathbb{R}^{n+1} \mid \quad \left[\textstyle\sum_{i=1}^{n+1} f_i\right]^2 + \varepsilon \sum_{i=1}^{n+1} f_i^2 \leq 1 \right\}.$$

Then for $\varepsilon \to 0$, we obtain

$$e_n^{\text{std}-\text{wor}}(F, L_\infty) \geq 1.$$

Indeed, knowing $f(x_i)$ for $i = 1, 2, \ldots, n$, with $x_i \in \{1, 2, \ldots, n+1\}$, we take $f$ such that $f(x_i) = 0$. Since we have at most $n$ conditions on $n+1$ components of $f$ then at least one component of $f$ from the unit ball is free and can be taken as $\pm 1/\sqrt{1+\varepsilon}$. This proves that the worst case error of any algorithm is at least $1/\sqrt{1+\varepsilon}$ which in the limit as $\varepsilon$ goes to zero is 1.

Consider the information

$$N(f) = [f_1 - f_2, f_2 - f_3, \dots, f_n - f_{n+1}] \quad \text{for all} \quad f \in H.$$

It is known that the minimal error of all algorithms that use $N$ is the supremum of $\|f\|_H$ for $f \in B$ and $N(f) = 0$. Observe that $N(f) = 0$ implies that $f = [c, c, \dots, c]$. Next, $f \in B$ implies that

$$c^2 \le \frac{1 + \varepsilon/(n+1)}{(n+1)^2}.$$

Hence, again for $\varepsilon \to 0$, we obtain $e_n^{\mathrm{all-wor}}(F, L_\infty) \le 1/(n+1)$. □

Let $r^{\mathrm{all-wor}}(F, L_\infty) = r > 1$. Then the inequality (1) implies that

$$r^{\mathrm{std-wor}}(F, L_\infty) \ge r - 1.$$

Thus, Theorem 3 implies the following behavior of the power function for $p = \infty$.

**Theorem 4.**

$$\ell^{\,\mathrm{wor-H/B}}(r, \infty) \quad \in \quad \left[\frac{r-1}{r}, 1\right] \qquad \textit{for all} \quad r > 1.$$

Hence, for both $p = 2$ and $p = \infty$, we see that for large $r$, the power function is almost one.

We want to guess the behavior of the power function for $r > \min(\frac{1}{p}, \frac{1}{2})$. It can be helpful to see the actual rates of convergence for some standard spaces. In particular, for $p = \infty$, the rates are known for Sobolev spaces.

**Example 4** (Sobolev spaces, $p = \infty$)**.**

a) For the Sobolev spaces $W_2^s([0,1]^d)$ and an arbitrary $s$ for which $2s > d$, it is well known that

$$e_n^{\mathrm{all-wor}}(W_2^s([0,1]^d), L_\infty) \asymp e_n^{\mathrm{std-wor}}(W_2^s([0,1]^d), L_\infty) \asymp n^{-s/d+1/2},$$

see, e.g., [14].

b) For the Sobolev spaces $W_2^{s,\mathrm{mix}}([0,1]^d)$ with $s > 1/2$, it is known that

$$e_n^{\mathrm{all-wor}}(W_2^{s,\mathrm{mix}}([0,1]^d), L_\infty) \asymp e_n^{\mathrm{std-wor}}(W_2^{s,\mathrm{mix}}([0,1]^d), L_\infty) \asymp n^{-s+1/2}(\log n)^{(d-1)s},$$

see [20]. □

Hence, at least for the standard Sobolev spaces the rates are the same even up to logarithmic factors. This again suggests that the power function can be just one for all $r > (\min(\frac{1}{p}, \frac{1}{2}), \infty)$. This is the next open problem.

**Open Problem 3.** Verify whether it is true that for all $p \in [1, \infty]$ we have

$$\ell^{\,\mathrm{wor-H}}(r, p) = \begin{cases} 0 & \text{for all } r \in \big(0, \min(\frac{1}{p}, \frac{1}{2})\big], \\ 1 & \text{for all } r \in \big(\min(\frac{1}{p}, \frac{1}{2}), \infty\big). \end{cases}$$

We end this section with a remark on the rates of convergence for different $p$.

**Remark 2.** It is interesting to compare the sequences

$$e_n^{\text{all-wor}}(H, L_p) \qquad \text{and/or} \qquad e_n^{\text{std-wor}}(H, L_p)$$

for the same $H$ but different $p$. The following example shows that, in general, there exists *no* relation between these sequences. Some relations do exist as shown in [7] but under some additional assumptions about $H$. The following example shows that some assumptions on $H$ are indeed needed, otherwise everything can happen.

Take $L_2 = L_2([0,1])$, $L_\infty = L_\infty([0,1])$ and assume that $[0,1]$ is the disjoint union of intervals $I_k$ of positive length $\lambda_k$ such that $\sum_{k=1}^\infty \lambda_k = 1$. Assume also that

$$\lambda_1 \ge \lambda_2 \ge \cdots$$

and put $e_k = 1_{I_k}$. We define a Hilbert space $H$ by its unit ball

$$B = \left\{ \sum_{k=1}^\infty \alpha_k e_k \;\middle|\; \sum_{k=1}^\infty \frac{\alpha_k^2}{\gamma_k^2} \le 1 \right\},$$

where

$$\gamma_1 \ge \gamma_2 \ge \cdots > 0 \quad \text{with} \quad \lim_{k\to\infty} \gamma_k = 0.$$

Hence for $f = \sum_{k=1}^\infty \alpha_k e_k \in H$, we obtain

$$\|f\|_H^2 = \sum_{k=1}^\infty \frac{\alpha_k^2}{\gamma_k^2} \qquad \text{and} \qquad \|f\|_2^2 = \sum_{k=1}^\infty \alpha_k^2\, \lambda_k, \qquad \|f\|_\infty = \sup_k |\alpha_k|.$$

From this, we easily conclude that the optimal approximation for $L_2$ as well as for $L_\infty$ is given by

$$f = \sum_{k=1}^\infty \alpha_k e_k \;\mapsto\; \sum_{k=1}^n \alpha_k e_k.$$

Note that

$$\alpha_k = \langle f, e_k \rangle_H = f(x_k)\, \lambda_k,$$

where $x_k \in I_k$. This means that the optimal error algorithm for function values and linear functionals is the same, and therefore

$$e^{\text{all-wor}}(H, L_p) = e^{\text{std-wor}}(H, L_p) \quad \text{for} \quad p \in \{2, \infty\}.$$

However,

$$e_n^{\text{all-wor}}(H, L_\infty) = \gamma_{n+1} \qquad \text{and} \qquad e_n^{\text{all-wor}}(H, L_2) = \gamma_{n+1}\sqrt{\lambda_{n+1}}.$$

Since $\{\gamma_n\}$ and $\{\lambda_n\}$ are not related, it is easy to get an example with

$$r^{\text{all-wor}}(H, L_\infty) = 0 \qquad \text{but} \qquad r^{\text{all-wor}}(H, L_2) = \infty.$$

Hence, in general, the difference between the minimal rates for $L_2$ and $L_\infty$ approximation can be extreme.

### 2.3 Banach Case

In this subsection, we study the approximation problem defined over a Banach space that is continuously embedded in $L_p$. As always, we assume that function evaluations are continuous functionals. We establish some bounds on the power functions by recalling known results for Sobolev spaces.

**Example 5** (Sobolev spaces, $1 \le p < \infty$)**.**
For the Sobolev space $W_p^s([0,1]^d)$ for an arbitrary $s > 0$, it is known that

$$e_n^{\text{all-wor}}(W_p^s([0,1]^d), L_p) \asymp n^{-s/d}.$$

Function values are well defined in $W_p^s([0,1]^d)$ only if the embedding condition $s/d > 1/p$ or $s = d$ and $p = 1$ holds. However, we may use the approach suggested in [2] that allows us to consider the case without this embedding condition. Namely, we limit ourselves only to continuous functions by taking

$$F = W_p^s([0,1]^d) \cap C([0,1]^d)$$

with norm

$$\|f\|_F = \|f\|_{W_p^s([0,1]^d)} + \|f\|_{C([0,1]^d)}.$$

Here, $C([0,1]^d)$ is the space of continuous functions equipped with the max norm. Then $F$ is a Banach space for which function values are well defined and function values are continuous linear functionals on this space. Then for $s/d \le 1/p$ and $s/d < 1$ in the case $p = 1$, respectively, it was shown in [2] that

$$e_n^{\text{std-wor}}(F, L_p) \asymp 1. \qquad \square$$

The last example implies that

$$\begin{aligned} \ell^{\,\text{wor-B}}(r,p) &= 0 \quad \text{for all} \quad r \in (0, 1/p] \quad \text{and} \quad 1 < p < \infty, \\ \ell^{\,\text{wor-B}}(r,1) &= 0 \quad \text{for all} \quad r \in (0,1). \end{aligned}$$

We now show that $\ell^{\,\text{wor-B}}(r,p) = 0$ over larger domains of $r$ for a given $p$ by recalling other results for Sobolev spaces.

**Example 6** (Sobolev space $W_1^s([0,1]^d)$, $1 \le p < \infty$)**.**
Consider the approximation problem for the Sobolev space $W_1^s([0,1]^d)$ with error measured in $L_p = L_p([0,1]^d)$. This problem is well defined and convergent for the class $\Lambda^{\text{all}}$ if we assume that $s/d > 1 - 1/p$.
For $p \in [1,2]$, we have

$$e_n^{\text{all-wor}}(W_1^s([0,1]^d), L_p) \asymp n^{-s/d},$$

whereas for $p \in [2,\infty)$, we have

$$e_n^{\text{all-wor}}(W_1^s([0,1]^d), L_p) \asymp n^{-s/d+1/2-1/p},$$

see e.g., [24]. The last relation also holds for $p = \infty$ as will be needed later.
The same results are also valid for the space $F = W_1^s([0,1]^d) \cap C([0,1]^d)$ with the norm

$$\|f\|_F = \|f\|_{W_1^s([0,1]^d)} + \|f\|_{C([0,1]^d)}.$$

For the space $F$, we can consider function values for all $s/d > 1 - 1/p$. For $s/d \le 1$, we have

$$e_n^{\text{std-wor}}(F, L_p) \asymp 1. \qquad \square$$

Let $p \in [1,2]$. The previous example implies that

$$\ell^{\,\text{wor}-\text{B}}(r,p) = 0 \qquad \text{for all} \qquad r \in \left(1 - \frac{1}{p},\ 1\right].$$

For $p \in (1,2]$, we showed before that $\ell^{\,\text{wor}-\text{B}}(r,p) = 0$ for all $r \in (0, 1/p]$. Since $(0,1/p] \cup (1-1/p, 1] = (0,1]$, we obtain

$$\ell^{\,\text{wor}-\text{B}}(r,p) = 0 \qquad \text{for all} \qquad r \in (0,1] \quad \text{and} \quad p \in [1,2].$$

Let $p \in [2,\infty)$. The previous example implies that

$$\ell^{\,\text{wor}-\text{B}}(r,p) = 0 \qquad \text{for all} \qquad r \in \left(\frac{1}{2},\ \frac{1}{2} + \frac{1}{p}\right].$$

Now we show that $\ell^{\,\text{wor}-\text{B}}(r,p) = 0$ also for $p \in [2,\infty)$ and $r \in (0, \frac{1}{2}]$. We increase the space $F = W_1^s([0,1]) \cap C([0,1])$ with the norm

$$\|f\|_F = \|f\|_{W_1^s([0,1])} + \|f\|_{C([0,1])}$$

(for $d = 1$) even more by adding functions from a Hölder class $C^\alpha$, where $0 < \alpha \le 1/2$. Hence we take the space

$$\widetilde{F} = F + C^\alpha$$

with the norm

$$\|f\|_{\widetilde{F}} := \inf\{\|g\|_F + \|h\|_{C^\alpha} \mid \ f = g + h,\ g \in F,\ h \in C^\alpha\}.$$

Since the unit ball of $\widetilde{F}$ is larger than that of $F$ we still have $e_n^{\text{std}-\text{wor}}(\widetilde{F}, L_p) \asymp 1$ for $s \le 1$. It is well known that $e_n^{\text{all}-\text{wor}}(C^\alpha([0,1]), L_p) \asymp n^{-\alpha}$ and the same holds for $\widetilde{F}$ if $\alpha \le s - 1/2 + 1/p$. Hence for $p \in [2,\infty)$, we obtain

$$\ell^{\,\text{wor}-\text{B}}(r,p) = 0 \qquad \text{for all} \qquad r \in \left(0,\ \frac{1}{2} + \frac{1}{p}\right].$$

We learnt some properties of the power function by using known results for Sobolev spaces $W_{p_1}^s([0,1]^d)$ in the case $s/d \le 1/p_1$ so that function values did not even supply convergence. Since we needed to assume that $s/d > 1/p_1 - 1/p$, the case $p = \infty$ could not be covered.

We now recall some results for Sobolev spaces when the embedding condition is satisfied and when there is a difference in the convergence rates between function values and arbitrary linear functionals.

**Example 7** (Sobolev space $W_1^s([0,1]^d)$, $1 \le p \le \infty$)**.**
Consider the approximation problem for the Sobolev space $W_1^s([0,1]^d)$ with error measured in $L_p$. We now assume that $s/d \ge 1$. Then function values are well defined and are continuous linear functionals. Furthermore,

$$e_n^{\text{std}-\text{wor}}(W_1^s([0,1]^d), L_p) \asymp n^{-s/d+1-1/p},$$

see, e.g., the survey of such results in Section 4.2.4 of [14] or [23, 24]. □

The last two examples imply the following estimates of the power function. For all $r > 1$ and $p \in [1, 2]$, we have

$$\ell^{\,\mathrm{wor-B}}(r, p) \le 1 - \frac{1}{r}\left(1 - \frac{1}{p}\right),$$

and for all $r > 1$ and $p \in [2, \infty]$, we have

$$\ell^{\,\mathrm{wor-B}}(r, p) \le 1 - \frac{1}{2r}.$$

We summarize the properties of the power function established in this section in the following theorem. The only case where we have a positive lower bound is the case $p = \infty$, see Theorem 4.

**Theorem 5.**

$$\begin{array}{rcll}
\ell^{\,\mathrm{wor-B}}(r,p) & = & 0 & \textit{for all} \quad r \in (0,1] \quad \textit{and} \quad p \in [1,2], \\
\ell^{\,\mathrm{wor-B}}(r,p) & = & 0 & \textit{for all} \quad r \in (0, \frac{1}{2} + \frac{1}{p}] \quad \textit{and} \quad p \in (2,\infty), \\
\ell^{\,\mathrm{wor-B}}(r,p) & \le & 1 - \dfrac{1}{r}\left(1 - \dfrac{1}{p}\right) & \textit{for all} \quad r > 1 \quad \textit{and} \quad p \in [1,2], \\
\ell^{\,\mathrm{wor-B}}(r,p) & \le & 1 - \dfrac{1}{2r} & \textit{for all} \quad r > 1 \quad \textit{and} \quad p \in [2,\infty), \\
1 - \dfrac{1}{r} \le \ell^{\,\mathrm{wor-B}}(r,\infty) & \le & 1 - \dfrac{1}{2r} & \textit{for all} \quad r > 1.
\end{array}$$

It is interesting to note that although we do not know the exact values of the power functions in the Hilbert and Banach cases, we can check that they are different at least for $p = 2$. Indeed, from Theorems 1 and 5, we have

$$\begin{array}{rcll}
\ell^{\,\mathrm{wor-B}}(r,2) & = & \ell^{\,\mathrm{wor-H}}(r,2) & \text{for all} \quad r \in (0, \frac{1}{2}], \\
\ell^{\,\mathrm{wor-B}}(r,2) = 0 & < & \frac{1}{2} \le \ell^{\,\mathrm{wor-H}}(r,2) & \text{for all} \quad r \in (\frac{1}{2}, 1], \\
\ell^{\,\mathrm{wor-B}}(r,2) \le 1 - \dfrac{1}{2r} & < & \dfrac{2r}{2r+1} \le \ell^{\,\mathrm{wor-H}}(r,2) & \text{for all} \quad r \in (1,\infty).
\end{array}$$

This shows that at least for $p = 2$ the power of function values for the Hilbert case is larger than for the Banach case for all $r > \frac{1}{2}$.

Obviously, it would be desirable to find the exact values of the power function $\ell^{\,\mathrm{wor-B}}(r, p)$ for all $r \in (0, \infty)$ and $p \in [1, \infty]$. However, it could be a very difficult problem. Hence, as maybe a less difficult problem, we would like to check the following property of the power function.

**Open Problem 4.** For $p \in [1, \infty]$, find the supremum $a^*(p)$ of $a$ for which

$$\ell^{\,\mathrm{wor-B}}(r, p) = 0 \quad \text{for all} \quad r \in (0, a].$$

We only know that $a^*(p) \ge 1$ for all $p \in [1, \infty)$.

We already indicated that the power functions for the Hilbert and Banach cases are different for $p = 2$. It would be of interest to check if this holds for all $p$.

**Open Problem 5.** Find all $p \in [1, \infty]$ for which

$$\ell^{\,\mathrm{wor-B}}(\cdot, p) \ne \ell^{\,\mathrm{wor-H}}(\cdot, p).$$

Similar to Example 3, we present an example of a Banach space $F$ where the ratio

$$\frac{e_n^{\text{std-wor}}(F, L_p)}{e_n^{\text{all-wor}}(F, L_p)}$$

is large for $p > 1$ and a fixed $n$.

**Example 8.** Take $F = \ell_1^{n+1}$, i.e., $F = \mathbb{R}^{n+1}$ with the $\ell_1$ norm. Then we obtain

$$e_n^{\text{std-wor}}(F, L_p) = (n+1)^{1-1/p} e_n^{\text{all-wor}}(F, L_p), \tag{2}$$

since $e_n^{\text{std-wor}}(F, L_p) = 1$ and $e_n^{\text{all-wor}}(F, L_p) = (n+1)^{1/p-1}$. The upper bound in the last statement follows again with the information $N(x) = (x_2 - x_1, x_3 - x_2, \dots, x_{n+1} - x_n)$ while the lower bound follows from the fact that the unit ball of $\ell_1^{n+1}$ contains a $\ell_p^{n+1}$ ball of radius $(n+1)^{1/p-1}$.

Again this ratio $(n+1)^{1-1/p}$ as in (2) can be obtained with a Hilbert space and actually we can take the same spaces as in Example 3, i.e., we define in $H = \mathbb{R}^{n+1}$ the scalar product

$$\langle f, g \rangle_H = \left[\sum_{i=1}^{n+1} f_i\right] \left[\sum_{i=1}^{n+1} g_i\right] + \varepsilon \sum_{i=1}^{n+1} f_i g_i \quad \text{for all} \quad f, g \in H,$$

and consider the limit where $\varepsilon > 0$ tends to zero. $\square$

We end this section with another open problem.

**Open Problem 6.** Find the supremum of $e_n^{\text{std-wor}}(F, L_p)/e_n^{\text{all-wor}}(F, L_p)$ over all Banach and/or Hilbert spaces. So far, we know that

$$\sup_F \frac{e_n^{\text{std-wor}}(F, L_p)}{e_n^{\text{all-wor}}(F, L_p)} \geq (n+1)^{1-1/p}, \tag{3}$$

and equality holds if $p = \infty$.

## 3 Randomized setting

We approximate the embedding operator $I : F \to L_p$ in the randomized setting. We now briefly define this setting. The reader may find more on this subject, e.g., in [14, 15, 21].

We approximate $I$ by algorithms $A_n$ that use $n$ values of linear functionals on the average and each linear functional is chosen randomly with respect to a probability distribution.

More precisely, the algorithm $A_n$ is of the following form

$$A_n(f, \omega) = \phi_{n,\omega}\left(L_{1,\omega_1}(f), L_{2,\omega_2}(f), \dots, L_{n(\omega),\omega_{n(\omega)}}(f)\right), \tag{4}$$

and the number $n(\omega)$ of functionals can also be random. Here $\omega = [\omega_1, \omega_2, \dots]$, and the linear functionals $L_{j,\omega_j}$ are random functionals distributed according to a probability distribution on elements $\omega_j$ which may depend on $j$ as well as on the values already computed, i.e., on $L_{i,\omega_i}(f)$ for $i = 1, 2, \dots, j-1$. The mapping $\phi_{n,\omega} : \mathbb{R}^{n(\omega)} \to L_p$ is a random mapping, and

$$\mathbb{E}_\omega\, n(\omega) \leq n.$$

We also allow adaptive choices of the functionals $L_{j,\omega_j}$. That is, $L_{j,\omega_j}$ may depend on the already selected functionals and the values $L_{1,\omega_1}(f), L_{2,\omega_2}(f), \dots, L_{j-1,\omega_{j-1}}(f)$.

Without loss of generality, we assume that $A_n(f,\cdot)$ is measurable, and define the randomized error of $A_n$ as

$$e^{\rm ran}(A_n) = \sup_{\|f\|_F \le 1} \left(\mathbb{E}_\omega \|I(f) - A_n(f,\omega)\|_p^2\right)^{1/2}.$$

Again, we compare such algorithms with algorithms that are based on function values, i.e., each $L_{j,\omega_j}$ is now of the form $L_{j,\omega_j}(f) = f(t_{j,\omega_j})$ and

$$A_n(f,\omega) = \phi_{n,\omega}\left(f(t_{1,\omega_1}), f(t_{2,\omega_2}), \dots, f(t_{n(\omega),\omega_{n(\omega)}})\right). \tag{5}$$

Hence, we consider algorithms that use $n$ linear functionals either from the class $\Lambda^{\rm std}$ or the class $\Lambda^{\rm all}$. We define the minimal errors as follows.

**Definition 2.** For $n \in \mathbb{N}_0$, let

$$e_n^{\rm all-ran}(F,L_p) = \inf\left\{e^{\rm ran}(A_n) \mid \ L_j \in \Lambda^{\rm all} \text{ and } A_n \text{ as in (4)}\right\},$$

and

$$e_n^{\rm std-ran}(F,L_p) = \inf\left\{e^{\rm ran}(A_n) \mid \ L_j \in \Lambda^{\rm std} \text{ and } A_n \text{ as in (5)}\right\}.$$

As in the worst case setting, for $n=0$ it is easy to see that the best algorithm is $A_0 = 0$ and obtain

$$e_0^{\rm all-ran}(F,L_p) = e_0^{\rm std-ran}(F,L_p) = \sup_{\|f\|_F \le 1} \|f\|_p = \sup_{\|f\|_F \le 1} \|I(f)\|_p = \|I\|.$$

This is the initial error that can be achieved without computing any linear functional on the functions $f$. Clearly,

$$e_n^{\rm all-ran}(F,L_p) \le e_n^{\rm std-ran}(F,L_p) \quad \text{for all} \quad n \in \mathbb{N}.$$

The sequences $\left(e_n^{\rm all-ran}(F,L_p)\right)$ and $\left(e_n^{\rm std-ran}(F,L_p)\right)$ are both non-increasing but not necessarily convergent to zero.

As in the worst case setting, we want to compare the rates of convergence

$$r^{\rm all-ran}(F,L_p) = r\left(e_n^{\rm all-ran}(F,L_p)\right) \qquad \text{and} \qquad r^{\rm std-ran}(F,L_p) = r\left(e_n^{\rm std-ran}(F,L_p)\right).$$

In particular, we would like to know if it is possible that the sequence $\left(r^{\rm all-ran}(F,L_p)\right)$ converges much faster than the sequence $\left(r^{\rm std-ran}(F,L_p)\right)$. The main question addressed in this section is to find or estimate the *power* function defined as $\ell^{\rm ran-x} : (0,\infty) \times [1,\infty] \to [0,1]$ by

$$\ell^{\rm ran-x}(r,p) := \inf_{F:\, r^{\rm all-ran}(F,L_p)=r} \frac{r^{\rm std-ran}(F,L_p)}{r},$$

where $\mathrm{x} \in \{H,B\}$ indicates that the infimum is taken over all Hilbert spaces ($\mathrm{x}=H$) or over all Banach spaces ($\mathrm{x}=B$) continuously embedded in $L_p$ and the rate of convergence is $r$ when we use arbitrary linear functionals. In the randomized setting, we do *not* need to assume that function values are continuous linear functionals.

### 3.1 Double Hilbert Case

In this subsection, we consider the approximation problem defined over a Hilbert space with the error measured also in the Hilbert space $L_2$. It may be surprising but the results in the double Hilbert case are complete due to [28], and there is no need to discuss different cases depending on the values of $r$.

**Theorem 6** ([28]). *Let $I : H \to L_2(\Omega)$ be a continuous embedding from a Hilbert space $H$ into $L_2(\Omega)$. Then*

$$r^{\mathrm{all-ran}}(H, L_2) = r^{\mathrm{std-ran}}(H, L_2).$$

*Therefore*

$$\ell^{\mathrm{ran-H}}(r, 2) = 1 \qquad \textit{for all} \quad r > 0.$$

We add that it was known before, see [13, 25], that also

$$r^{\mathrm{all-ran}}(H, L_2) = r^{\mathrm{all-wor}}(H, L_2).$$

This means that the power of function values in the randomized setting is the same as the power of arbitrary linear functionals in the worst case setting, which in turn is the same as in the randomized setting.

### 3.2 Other Cases

For $p > 2$, we know examples from the literature where the rate $r^{\mathrm{all-ran}}(H, L_p)$ is larger than the rate $r^{\mathrm{std-ran}}(H, L_p)$. Namely take $I : W_2^r([0,1]) \to L_p([0,1])$. Then with $\Lambda^{\mathrm{all}}$ one can achieve the order $n^{-r}$ (with additional log terms in the case $p = \infty$, but the order is still $r$), see [10]. For $\Lambda^{\mathrm{std}}$ the optimal order is $n^{-r+1/2-1/p}$, see [2]. The authors of [2, 10] studied the case of integer $r$, but the results can be extended via interpolation to all $r > 1$. Therefore, we obtain

$$\ell^{\mathrm{ran-H}}(r, p) \le \frac{r - 1/2 + 1/p}{r} \qquad \text{if} \quad r \ge 1 \quad \text{and} \quad p > 2.$$

We summarize these estimates of the power function in the following theorem.

**Theorem 7.** *Let $p > 2$. Then*

$$\ell^{\mathrm{ran-H}}(r, p) \quad \le \quad 1 - \frac{1/2 - 1/p}{r} \qquad \textit{for all} \quad r \ge 1.$$

Sobolev embeddings in the randomized setting were studied by several authors, including [2, 3, 4, 10, 12, 21, 25]. For our purpose, the most important papers are [2, 10] and the paper [3] for the interpolation argument.

For the embedding $I : W_2^r([0,1]) \to L_\infty([0,1])$ the rate is improved by $1/2$ if we switch from the class $\Lambda^{\mathrm{std}}$ to the class $\Lambda^{\mathrm{all}}$. This gap of $1/2$ is the largest possible under some additional conditions, see [7, 9]. Let us add in passing that the same gap of $1/2$ appears for $\Lambda^{\mathrm{all}}$ between the worst case and the randomized setting.

The Hilbert case for $p \in [1, 2)$ as well as the Banach case for all $p \in [1, \infty]$ have not yet been studied. We pose this as an open problem.

**Open Problem 7.** Study the power function in the randomized setting for the Hilbert case with $p \in [1,2)$ and for the Banach case for all $p \in [1,\infty]$. In particular, determine the supremum $a^*(p)$ of $a$ for which

$$\ell^{\,\mathrm{ran-H/B}}(r,p) = 0 \quad \text{for all} \quad r \in (0,a].$$

## 4 Average case setting with a Gaussian measure

In the average case setting, we assume that $I : F \to L_p(\Omega)$ is continuously embedded and function evaluations are continuous functionals on $F$. As far as we know, only the case $p = 2$ was studied and we report the known results from [5] for this case.

We assume that $F$ is a separable Hilbert/Banach space equipped with a zero mean Gaussian measure $\mu$. As in the worst case setting, we consider deterministic algorithms, and due to general results, see [21], it is enough to compare linear algorithms

$$A_n(f) = \sum_{k=1}^{n} L_k(f) g_k \quad \text{and} \quad A_n(f) = \sum_{k=1}^{n} f(x_k) g_k,$$

where $g_k \in L_2(\Omega)$. The average case error of an algorithm is defined by

$$e^{\mathrm{avg}}(A) := \left( \int_F \|f - A(f)\|_p^2 \, \mathrm{d}\mu(f) \right)^{1/p}.$$

As in the other settings, we define the minimal $n$th average case errors $e_n^{\mathrm{all-avg}}(F, L_p)$, $e_n^{\mathrm{std-avg}}(F, L_p)$ and the power function $\ell^{\,\mathrm{avg-H/B}}$. That is, for

$$r^{\mathrm{all/std-avg}}(F, L_p) = r(e_n^{\mathrm{all/std-avg}}(F, L_p))$$

we have

$$\ell^{\,\mathrm{avg-x}}(r,p) := \inf_{F:\, r^{\mathrm{all-avg}}(F,L_p) = r} \frac{r^{\mathrm{std-avg}}(F, L_p)}{r}.$$

As always, $\mathrm{x} \in \{H, B\}$ and we take the infimum over separable Hilbert ($\mathrm{x} = H$) or Banach ($\mathrm{x} = B$) spaces equipped with zero mean Gaussian measures that are continuously embedded in $L_p$ and for which function values are continuous linear functionals as well as the rate of convergence is $r$ when arbitrary linear functionals are used.

As already mentioned, results are known only for $p = 2$. Then the cases of the Hilbert and Banach spaces are the same due to the presence of Gaussian measures. This follows from the fact that even if $F$ is a separable Banach space then the minimal errors for the class $\Lambda^{\mathrm{all}}$ depend on the Gaussian measure $\nu = \mu\, I^{-1}$ given by

$$\nu(M) = \mu\left(\{f \in F \mid \ I(f) \in M\}\right)$$

for a Borel set $M$ of $L_2$. The measure $\nu$ is also a zero mean Gaussian measure whose covariance operator $C_\nu : L_2 \to L_2$ is given by

$$\langle C_\nu f_1, f_2 \rangle_{L_2} = \int_{L_2} \langle f, f_1 \rangle_{L_2} \langle f, f_2 \rangle_{L_2} \, \mathrm{d}\nu(f) \quad \text{for all} \quad f_1, f_2 \in L_2.$$

The operator $C_\nu$ is self adjoint, positive semi-definite, compact and has a finite trace. That is, its ordered eigenvalues $\lambda_j$ have a finite sum. It is known that

$$e_n^{\mathrm{all-avg}}(F, L_2) = \left( \sum_{j=n+1}^{\infty} \lambda_j \right)^{1/2}.$$

As in the randomized setting for the double Hilbert space, the results on the power function are complete and there is no need to discuss different cases of $r$.

**Theorem 8** ([5])**.** *Let $I : F \to L_2(\Omega)$ be a continuous embedding from a separable Banach space $F$ equipped with a zero mean Gaussian measure $\mu$ into $L_2(\Omega)$. Then*

$$r^{\mathrm{all-avg}}(F, L_2) = r^{\mathrm{std-avg}}(F, L_2).$$

*Therefore*

$$\ell^{\,\mathrm{avg-H/B}}(r, 2) = 1 \qquad \textit{for all} \quad r > 0.$$

Of course it would be interesting to study the power function for other values of $p$. This is posed as our last open problem.

**Open Problem 8.** Study the power function in the average case setting for $p \neq 2$. In particular, verify whether a similar result as Theorem 8 holds.

**Acknowledgment**

We appreciate comments on this paper from Stefan Heinrich, Anargyros Papageorgiou, Joseph F. Traub, Grzegorz W. Wasilkowski and two anonymous referees. We especially thank Stefan Heinrich for pointing out some errors in our previous manuscript.

E.N. was partially supported by the DFG-Priority Program 1324. H.W. was partially supported by the National Science Foundation.

## References

[1] E. M. Galeev, Linear widths of Hölder-Nikolskii classes of periodic functions of several variables, Math. Notes **59**, 133–146, 1996.

[2] S. Heinrich, Randomized approximation of Sobolev embeddings, in: *Monte Carlo and Quasi-Monte Carlo Methods 2006*, A. Keller, S. Heinrich, H. Niederreiter (eds.), 445–459, Springer, Berlin, 2008.

[3] S. Heinrich, Randomized approximation of Sobolev embeddings II, *J. Complexity* **25**, 455–472, 2009.

[4] S. Heinrich, Randomized approximation of Sobolev embeddings III, *J. Complexity* **25**, 473–507, 2009.

[5] F. Hickernell, G. W. Wasilkowski and H. Woźniakowski, Tractability of linear multivariate problems in the average case setting, in: *Monte Carlo and Quasi-Monte Carlo Methods 2006*, A. Keller, S. Heinrich, H. Niederreiter (eds.), 461–494, Springer, Berlin, 2008.

[6] A. Hinrichs, E. Novak and J. Vybíral, Linear information versus function evaluations for $L_2$-approximation, *J. Approx. Th.* **153**, 97–107, 2008.

[7] F. Y. Kuo, G. W. Wasilkowski and H. Woźniakowski, Multivariate $L_\infty$ approximation in the worst case setting over reproducing kernel Hilbert spaces, *J. Approx. Th.* **152**, 135–160, 2008.

[8] 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. Th.* **158**, 97–125, 2009.

[9] F. Y. Kuo, G. W. Wasilkowski and H. Woźniakowski, On the power of standard information for $L_\infty$ approximation in the randomized case setting, *BIT Numer. Math.* **49**, 543–564, 2009.

[10] P. Mathé, Random approximation of Sobolev embeddings, *J. Complexity* **7**, 261–281, 1991.

[11] C. A. Micchelli and G. Wahba, Design problems for optimal surface interpolation, in *Approximation Theory and Applications*, Z. Ziegler ed., pp. 329–347, Academic Press, New York, 1981.

[12] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis*, LNiM **1349**, Springer-Verlag, Berlin, 1988.

[13] E. Novak, Optimal linear randomized methods for linear operators in Hilbert spaces, *J. Complexity* **8**, 22–36, 1992.

[14] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems*, Volume I: Linear Information, European Math. Soc., Zürich, 2008.

[15] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems*, Volume II: Standard Information for Functionals, EMS, Zürich, 2010, to appear.

[16] A. Pietsch, *Operator Ideals*, North Holland, 1980.

[17] W. Sickel and T. Ullrich, Spline interpolation on sparse grids, to appear in Applicable Analysis.

[18] R. Tandetzky, Approximation of functions from a Hilbert space using function values or general linear information, in progress, 2010.

[19] V. N. Temlyakov, *Approximation of Periodic Functions*, Nova Science, New York, 1993.

[20] V. N. Temlyakov, On approximate recovery of functions with bounded mixed derivative. *J. Complexity* **9**, 41–59, 1993.

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

[22] H. Triebel, *Bases in Function Spaces, Sampling, Discrepancy, Numerical Integration*, EMS Publ. House, Zürich, 2010.

[23] J. Vybíral, Sampling numbers and function spaces, *J. Complexity* **23**, 773–792, 2008.

[24] J. Vybíral, Widths of embeddings in function spaces, *J. Complexity* **24**, 545–570, 2008.

[25] G. W. Wasilkowski, Randomization for continuous problems, *J. Complexity* **5**, 195–218, 1989.

[26] H. Woźniakowski, Tractability and strong tractability of multivariate tensor product problems, *J. of Computing and Information* **4**, 1–19, 1994.

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

[28] G. W. Wasilkowski and H. Woźniakowski, The power of standard information for multivariate approximation in the randomized case setting, *Math. Comp.* **76**, 965–988, 2007.

Erich Novak
Mathematisches Institut, Universität Jena
Ernst-Abbe-Platz 2, 07740 Jena, Germany
`novak@mathematik.uni-jena.de`
`http://users.minet.uni-jena/∼novak`

Henryk Woźniakowski
Department of Computer Science, Columbia University
New York, NY 10027, USA, and
Institute of Applied Mathematics, University of Warsaw
ul. Banacha 2, 02-097 Warszawa, Poland
henryk@cs.columbia.edu
http://www.cs.columbia.edu/∼henryk