Linear information versus function evaluations for L_2-approximation

Aicke Hinrichs
Mathematisches Institut, Universität Jena
Ernst-Abbe-Platz 2, 07740 Jena, Germany
email: hinrichs@mathematik.uni-jena.de

Erich Novak
Mathematisches Institut, Universität Jena
Ernst-Abbe-Platz 2, 07740 Jena, Germany
email: novak@mathematik.uni-jena.de

Jan Vybíral
Mathematisches Institut, Universität Jena
Ernst-Abbe-Platz 2, 07740 Jena, Germany
email: vybiral@mathematik.uni-jena.de

February 25, 2008

Abstract

We study algorithms for the approximation of functions, the error is measured in an L_2 norm. We consider the worst case setting for a general reproducing kernel Hilbert space of functions. We analyze algorithms that use standard information consisting in n function values and we are interested in the optimal order of convergence. This is the maximal exponent b for which the worst case error of such an algorithm is of order n^{-b}.

Let p be the optimal order of convergence of all algorithms that may use arbitrary linear functionals, in contrast to function values only. So far it was not known whether $p > b$ is possible, i.e., whether the approximation numbers or linear widths can be essentially smaller than the sampling numbers. This is (implicitly) posed as an open problem in the recent paper Kuo, Wasilkowski, Woźniakowski (2007) where the authors prove that $p > 1/2$ implies $b \geq 2p^2/(2p+1) > p - 1/2$. Here we prove that the case $p = 1/2$ and $b = 0$ is possible, hence general linear information can be exponentially better than function evaluation. Since the case $p > 1/2$ is quite different, it is still open whether $b = p$ always holds in that case.

1 Introduction

We assume that μ is a measure on a set D and consider the space $L_2 = L_2(D, \mu)$. This is our target space. We also have a Hilbert space H of functions defined on D such that

- function values $f \mapsto f(x)$ are continuous (with respect to the H-norm);
- the identity (embedding)
$$I : H \to L_2 \tag{1}$$
is a well defined compact operator.

Hence H is a reproducing kernel Hilbert space imbedded in L_2. The approximation problem

$$\mathrm{APP} : H \to L_2, \qquad \mathrm{APP}(f) = f \tag{2}$$

is a well defined continuous linear operator. Let F be the unit ball of H. Then the approximation numbers or linear widths $a_n(F)$ are defined as follows. For a continuous linear algorithm

$$S_n(f) = \sum_{i=1}^{n} L_i(f) g_i \tag{3}$$

the (worst case) error is defined by

$$e(S_n) = \sup_{f \in F} \|f - S_n(f)\|_2.$$

Then $a_n(F)$ is given by

$$a_n(F) = \inf_{S_n} e(S_n).$$

The class of all continuous linear functionals is called Λ^{all} and hence the approximation numbers correspond to using information from Λ^{all}. In many applications not all algorithms (3) are feasible and algorithms

$$S_n(f) = \sum_{i=1}^{n} f(x_i) g_i \tag{4}$$

based on function values (standard information, denoted by Λ^{std}) are preferred. The sampling numbers $g_n(F)$ are defined by

$$g_n(F) = \inf_{S_n} e(S_n),$$

where now the infimum runs only over all S_n of the form (4).

Often the $g_n(F)$ are only "slightly" larger than the $a_n(F)$. To have precisely posed questions and, in some cases, answers, we consider the concept of "order of convergence" and define

$$p^{\mathrm{all}}(F) = \sup \left\{ \alpha \ge 0 \,:\, \lim_{n \to \infty} a_n(F) \cdot n^{\alpha} = 0 \right\}$$

as well as

$$p^{\mathrm{std}}(F) = \sup \left\{ \alpha \geq 0 \, : \, \lim_{n \to \infty} g_n(F) \cdot n^\alpha = 0 \right\}.$$

We know no example from the literature with $p^{\mathrm{all}}(F) > p^{\mathrm{std}}(F)$. It is known from Kuo, Wasilkowski, Woźniakowski (2007) that the gap cannot be too large. One of the results of these authors is that $p^{\mathrm{all}}(F) > 1/2$ implies

$$p^{\mathrm{std}}(F) \geq p^{\mathrm{all}}(F) - \frac{1}{2 + 1/p^{\mathrm{all}}}. \tag{5}$$

The main result of this paper is the construction of an example F with

$$p^{\mathrm{all}}(F) = 1/2 \qquad \text{and} \qquad p^{\mathrm{std}}(F) = 0. \tag{6}$$

Together with (5) this implies that the gap 1/2 between $p^{\mathrm{std}}(F)$ and $p^{\mathrm{all}}(F)$ is maximal.

Remark 1. *One could think of replacing* (1) *by an arbitrary compact operator*

$$S : H \to L_2, \tag{7}$$

also the target space L_2 could be replaced by another Hilbert space. Then, however, the results are completely different. *Linear information now can be* much *better than function values. Examples and an explanation of this will follow in Remark 3.*

Remark 2. *Here we discuss a related problem and another interpretation of our example with* (6).

For the approximation numbers of APP : $H \to L_2$ *we consider arbitrary linear mappings of the form*

$$S_n(f) = \sum_{i=1}^n \langle f, v_i \rangle \, w_i \tag{8}$$

with $v_i, w_i \in L_2$. Information about f is given by the functionals $f \mapsto \langle f, v_i \rangle$. Assume now that only mappings

$$S_n(f) = \sum_{i=1}^n \langle f, b_{k_i} \rangle \, w_i \tag{9}$$

are allowed, where the b_i form a given and fixed complete orthonormal system of L_2. This means that the class Λ^{all} of all functionals is restricted to another class Λ^{restr}, but the w_i in (9) *are still arbitrary. We ask whether approximations* (8) *are "much" better than those of the form* (9).

It turns out that the inequality of Kuo, Wasilkowski, Woźniakowski (2007) is still true, hence $p^{\mathrm{all}}(F) > 1/2$ implies

$$p^{\mathrm{restr}}(F) \geq p^{\mathrm{all}}(F) - \frac{1}{2 + 1/p^{\mathrm{all}}}. \tag{10}$$

Moreover, our example also covers this case since we have

$$p^{\mathrm{all}}(F) = 1/2 \qquad \textit{and} \qquad p^{\mathrm{restr}}(F) = 0. \tag{11}$$

A somehow dual problem was studied by Donoho [2], Temlyakov [4] and others: these authors assume that the "approximation space" can be chosen only in a restricted way. Hence the w_i form an orthonormal system and $S_n(f)$ is of the form $\sum_{i=1}^n c_i w_{k_i}$, where the c_i and the k_i may depend in an arbitrary way on f.

With (9) *we still study linear algorithms since the k_i are chosen for the whole space H, not individually for a given f.*

2 The finite dimensional case

We start with the finite dimensional case. Some technical problems disappear, now function values are defined and continuous even on $L_2 = \mathbb{R}^N = \ell_2^N$. Observe also that function evaluations correspond to scalar products with respect to a particular orthonormal system. Hence it is clear that in this case our main problem coincides with the problem mentioned in Remark 2.

We assume that D has N elements and consider the mapping

$$\mathrm{APP} : \mathbb{R}^N \to \mathbb{R}^N = \ell_2^N \tag{12}$$

on an ellipsoid $F \subset \mathbb{R}^N$ of the form

$$F = \{f \in \mathbb{R}^N \mid f = \sum_{i=1}^N x_i e_i,\ \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} \le 1\}. \tag{13}$$

Here we assume that the e_i form a complete ON-system and the singular values are ordered,

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_N \ge 0. \tag{14}$$

Of course we have $x_i = \langle f, e_i \rangle$, with the scalar product in ℓ_2^N. The x_i are the coordinates of f with respect to the complete ON-system $\{e_i\}$. Observe that F is the unit ball of a reproducing kernel Hilbert space $H = \mathbb{R}^N$, the kernel is given by

$$K(x,y) = \sum_{i=1}^N \sigma_i^2 e_i(x) e_i(y).$$

Each f is a mapping from $D = \{1, 2, \ldots, N\}$ to $\mathbb{R}$ and the function evaluations are the mappings

$$f \mapsto f_i = \langle f, b_i \rangle . \tag{15}$$

Here the $\{b_i\}$ form the standard basis of $\mathbb{R}^N = \ell_2^N$, of course this is another complete ON-system of the target space ℓ_2^N. The approximation numbers are given by

$$a_n(F) = \sigma_{n+1} \tag{16}$$

and it is clear what to do: the optimal approximation is

$$S_n^*(f) = \sum_{i=1}^{n} \langle f, e_i \rangle \; e_i. \tag{17}$$

The optimal information $f \mapsto \langle f, e_i \rangle$ (for $i = 1, 2, \dots n$) clearly depends on the set F since we use the "eigenvectors" e_i. In the case of standard information we have to use approximations of the form

$$S_n(f) = \phi(f_{i_1}, f_{i_2}, \dots, f_{i_n}), \tag{18}$$

i.e., information of the form $f \mapsto \langle f, b_i \rangle$. Since F is a Hilbert space, the optimal ϕ can always be chosen linear, see, e.g., [5, Chapter 4].

Remark 3. *The difference between* (1) *and the more general case* (7) *is easy to see in this finite dimensional case: In the first case the allowed information functionals are of the form*

$$f \mapsto L_i(S(f)) = L_i(f)$$

that are orthogonal with respect to the target space L_2, i.e., we can compute the projection of $S(f)$ onto vectors in L_2 that are orthogonal.

This is not the case in the more general case (7)*, examples are known where approximation numbers are much smaller than sampling numbers (see [1, 6, 9]) . One example is given by the Sobolev spaces $H^s(\Omega)$, where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain. Consider the embedding $I : H^s(\Omega) \to H^{-t}(\Omega)$ where $s, t > 0$ and $s > d/2$. Then $a_n \asymp n^{-(s+t)/d}$ while $g_n \asymp n^{-s/d}$.*

It seems that the numbers g_n are "large" if the b_i are "almost orthogonal" to the e_k. Hence we consider the following example.

We assume that the matrix which transforms $b_1, b_2, \dots, b_N$ into $e_1, e_2, \dots, e_N$ is a Hadamard matrix. Then we have formulas of the form

$$b_k = N^{-1/2} \cdot (\pm e_1 \pm e_2 \cdots \pm e_N) \tag{19}$$

and also

$$e_k = N^{-1/2} \cdot (\pm b_1 \pm b_2 \cdots \pm b_N). \tag{20}$$

We want to be more specific: We assume that N is of the form $N = 2^m$ and that the transformation $\{e_k\}_k \to \{b_k\}_k$ (and vice versa) is given by a Walsh-Hadamard matrix. Let

$$H^0 = (1), \qquad H^1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad H^{k+1} = \begin{pmatrix} H^k & H^k \\ H^k & -H^k \end{pmatrix}. \tag{21}$$

Then we have

$$N^{-1/2} \, H^m \, e_k = b_k \qquad \text{and} \qquad N^{-1/2} \, H^m \, b_k = e_k, \quad k = 1, 2, \dots, N. \tag{22}$$

Consider first the simplest case, $n = 1$. Since the signs (plus or minus) do not matter, we may assume

$$b_1 = N^{-1/2} \cdot (e_1 + e_2 + \cdots + e_N). \tag{23}$$

To compute the radius of information or, equivalently, the first sampling number g_1, we have to maximize $\sum_{k=1}^N x_k^2$ under the conditions

$$\sum_{k=1}^{N} \frac{x_k^2}{\sigma_k^2} = 1 \quad \text{and} \quad \sum_{k=1}^{N} x_k = 0. \tag{24}$$

The result of this extremal problem is g_1^2.

Define

$$c^* = \left(\sum_{k=1}^{N} \sigma_k^2 \right)^{-1}. \tag{25}$$

Then there exists an f^* of the form

$$f^* = e_1 - c^* \sum_{k=1}^{N} \pm \sigma_k^2 e_k \tag{26}$$

with information 0. If c^* is small then f^* is close to e_1 and the error is large. If c^* tends to 0 then $g_1(F)$ tends to the initial error σ_1.

Now we deal with a general (fixed) number n of information functionals. We assume that a whole sequence

$$\sigma_1 \geq \sigma_2 \geq \ldots$$

is given with

$$\sum_{k=1}^{\infty} \sigma_k^2 = \infty$$

but we still are in the (finite) Hadamard case: $N = 2^m$ is large but finite and we also consider the case $N \to \infty$. But formally $N = 2^m$ is finite and then of course we only have the finite sub-sequence

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N > 0.$$

Assume that

$$k_1, k_2, \ldots, k_n$$

(between 1 and N) are given. To estimate the radius of information and $g_n(F)$ we are looking for an f of the form

$$f = e_1 - c \sum_{k \in I} \pm \sigma_k^2 e_k \tag{27}$$

such that the information for f is zero, that is

$$\langle f, g_{k_i} \rangle = 0, \qquad i = 1, 2, \ldots, n,$$

and f is "close to e_1", hence $g_n(F)$ is "almost equal" to the initial error σ_1. To define I and to obtain results, we need the following on the matrices H^m.

The matrix H^2 has the following property: Pick any two different rows $k_1, k_2 \in \{1, 2, 3, 4\}$. Then the 8 pairs

$$\pm(H^2_{k_1,1}, H^2_{k_2,1}), \ldots, \pm(H^2_{k_1,4}, H^2_{k_2,4})$$

yield the whole set $\{-1, 1\}^2$ and each $z \in \{-1, 1\}^2$ has (exactly) two such representations.

To formalize this for larger n, it is convenient to use a different form of the Walsh-Hadamard matrices with columns and rows permuted which is borrowed from Walsh analysis. Let $G = \{+1, -1\}^m$ be the group of m-tuples of signs equipped with coordinatewise multiplication. The Rademacher functions $r_1, \ldots, r_m$ on G are just the coordinate functionals given by $r_h(s) = s_h$ for $s = (s_1, \ldots, s_m) \in G$. For a subset $A \subset \{1, \ldots, m\}$, the Walsh function w_A is defined as

$$w_A(s) = \prod_{h \in A} r_h(s) = \prod_{h \in A} s_h.$$

Let $N = 2^m$. Then the $N \times N$-matrix

$$\big(w_A(s)\big)_{A \subset \{1,\ldots,m\}, s \in G}$$

is just the Walsh-Hadamard matrix H^m up to the order of the rows and columns. To pick a specific order, we map the row and column indices to the set $\{1, \ldots, N\}$ via the maps

$$s \;\mapsto\; 1 + \sum_{h=1}^{m} \frac{1 - s_h}{2}\, 2^{m-h} \tag{28}$$

$$A \;\mapsto\; 1 + \sum_{h \in A} 2^{h-1}. \tag{29}$$

By slight abuse of notation, we again denote the resulting matrix with H^m.

The structural result needed for our purposes is contained in the following Lemma.

Lemma 1. *For k=0,...,m, define*

$$M_k = \{s \in G \;:\; s_h = 1 \textit{ for } h = 1, \ldots, k\}.$$

For $A_1, \ldots, A_n \subset \{1, \ldots, m\}$, let

$$M = \Big\{s \in G \;:\; w_{A_i}(s) = \prod_{h \in A_i} s_h = 1 \textit{ for } i = 1, \ldots, n\Big\}.$$

Then

$$\# M \cap M_k \geq 2^{m-n-k}. \tag{30}$$

For the proof of this lemma, we need the following

Lemma 2. *Let $M \subset G$ be a subgroup of G and let $A \subset \{1, \dots, m\}$. Then*

$$M_A = \{s \in M \ : \ w_A(s) = \prod_{h \in A} s_h = 1\}$$

satisfies either $M = M_A$ or $\# M_A = \# M/2$.

Proof. Obviously, M_A is a subgroup of M. Assume that $M \neq M_A$ and choose $s^o \in M \setminus M_A$. Then, for any $s \in M \setminus M_A$, we have $s^o s \in M_A$. Hence $\# M \setminus M_A \leq \# M_A$. Moreover, for any $s \in M_A$, we have $s^o s \in M \setminus M_A$. Hence also $\# M \setminus M_A \geq \# M_A$. This shows that

$$\# M = \# M_A + \# M \setminus M_A = 2 \# M_A.$$

$\square$

Proof of Lemma 1. Define $A_{n+i} = \{i\}$ for $i = 1, \dots, k$. Then

$$M \cap M_k = \{s \in G \ : \ \prod_{h \in A_i} s_h = 1 \text{ for } i = 1, \dots, n+k\}.$$

Since $\# G = 2^m$, successive application of Lemma 2 to the chain of subgroups starting with $M = G$

$$\{s \in G \ : \ \prod_{h \in A_i} s_h = 1 \text{ for } i = 1, \dots, \ell\}$$

for $\ell = 1, \dots, n+k$ gives

$$\# M \cap M_k \geq \frac{2^m}{2^{n+k}} = 2^{m-n-k}.$$

$\square$

Using the identifications (28) and (29), the subsets $A_1, \dots, A_n$ are mapped to row indices $k_1, \dots, k_n \in \{1, \dots, N\}$ and the elements $s \in M$ are mapped to column indices ℓ_j such that $H^m_{k_i, \ell_j} = 1$. Moreover, $s \in M_k$ translates into $\ell \leq 2^{m-k}$ for the corresponding column index. Ordering the column indices $1 = \ell_1 < \ell_2 < \ldots < \ell_r \leq N$ which correspond to the elements $s \in M \cap M_k$, the inequalities (30) for $k = 0, \dots, m-n$ are equivalent to $\ell_{2^k} \leq 2^{k+n}$ for $k = 0, \dots, m-n$. Hence we obtain

Lemma 3. *Let $1 \leq k_1 < k_2 < \ldots < k_n \leq N = 2^m$. Let $1 = \ell_1 < \ell_2 < \ldots < \ell_r \leq N$ be the indices of the columns of H^m for which*

$$H^m_{k_i, \ell_j} = 1 \quad \textit{for } i = 1, \dots, n.$$

Then $r \geq 2^{m-n}$ and $\ell_{2^k} \leq 2^{k+n}$ for $k = 0, 1, \dots, m-n$.

Using this with $N = 2^m = 2^{n+t}$, we obtain

Lemma 4. *Assume that the information consists in the function values with the numbers*

$$k_1, k_2, \dots, k_n$$

(between 1 and N). Then there exist $2^t - 1$ numbers $\ell_1, \ell_2, \dots, \ell_{2^t-1}$ (different from 1) such that the information evaluated for e_1 coincides with the information evaluated for e_{ℓ_i} for each i. In addition we can arrange that

$$\ell_1 \le 2^{n+1}, \qquad \ell_2, \ell_3 \le 2^{n+2}, \qquad \ell_4, \ell_5, \ell_6, \ell_7 \le 2^{n+3} \tag{31}$$

and so on.

Hence we get the zero information for a vector of the form

$$f = e_1 - c \sum_{i=1}^{2^t-1} \sigma_{\ell_i}^2 e_{\ell_i}. \tag{32}$$

The number c is chosen in such a way that

$$c \sum_{i=1}^{2^t-1} \sigma_{\ell_i}^2 = 1.$$

Because of our assumption that $\sum_k \sigma_k^2 = \infty$, the number c tends (for given n and $t \to \infty$) to zero, the vector f tends to e_1 and the $g_n(F)$ tend to the initial error σ_1. Hence we obtain the following result.

Theorem 1. *Assume that a sequence*

$$1 = \sigma_1 \ge \sigma_2 \ge \dots$$

is given with

$$\sum_{k=1}^{\infty} \sigma_k^2 = \infty. \tag{33}$$

Assume further that a number n_0 and $\varepsilon_0 > 0$ are given. Then there exists an example with

$$a_{n_0}(F) = \sigma_{n_0+1}$$

and

$$g_{n_0}(F) \ge 1 - \varepsilon_0.$$

In this sense there does not exist any reasonable upper bound for the $g_n(F)$ if (33) holds.

3 Main result

Our main result is the following infinite dimensional example with similar properties.

Theorem 2. *Let $\sigma_n = n^{-1/2}$ and $\tau_n = (1+\log_2 n)^{-1/2}$ for $n \in \mathbb{N}$. Then there exists a sequence space H such that for its unit ball F the following holds*

- *$a_n(F) = \sigma_{n+1}$ for all natural numbers n,*
- *$g_n(F) \geq \frac{\sqrt{2}}{2} \cdot \tau_n$ for infinitely many natural numbers n.*

We start the construction with the following Lemma.

Lemma 5. *Let $K \in \mathbb{N}$ and $n \in \mathbb{N}$ be arbitrary natural numbers. Set $\varrho_k = \sigma_{k+K} = (k+K)^{-1/2}$ and $t = \frac{2(K+2^n)}{K}$. If F is the $N = 2^{n+t}$-dimensional Hadamard example with respect to $\varrho_1, \ldots, \varrho_N$, n and t as described above, then*

$$g_n(F) \geq \frac{\sqrt{2}}{2} \cdot \varrho_1.$$

Proof. Let $k_1, k_2, \cdots, k_n$ be the sampling points of the information and let $\ell_1, \ell_2, \ldots, \ell_{2^t-1}$ be the corresponding natural numbers (different from 1) as constructed in Lemma 4. We set

$$f = e_1 - c \sum_{i=1}^{2^t-1} \varrho_{\ell_i}^2 e_{\ell_i}, \quad \text{where} \quad c = \left(\sum_{i=1}^{2^t-1} \varrho_{\ell_i}^2 \right)^{-1}.$$

Then f is a vector with zero information and

$$\frac{||f||_2^2}{||f||_H^2} = \frac{1 + c^2 \sum_{i=1}^{2^t-1} \varrho_{\ell_i}^4}{\frac{1}{\varrho_1^2} + c^2 \sum_{i=1}^{2^t-1} \frac{\varrho_{\ell_i}^4}{\varrho_{\ell_i}^2}} \geq \varrho_1^2 \cdot \frac{\sum_{i=1}^{2^t-1} \varrho_{\ell_i}^2}{\sum_{i=1}^{2^t-1} \varrho_{\ell_i}^2 + \varrho_1^2}. \tag{34}$$

This estimate, combined with

$$\sum_{i=1}^{2^t-1} \varrho_{\ell_i}^2 = \sum_{j=1}^{t} \sum_{i=2^{j-1}}^{2^j-1} \varrho_{\ell_i}^2 \geq \sum_{j=1}^{t} 2^{j-1} \cdot \frac{1}{K+2^{n+j}} \geq \frac{t}{2} \cdot \frac{1}{K+2^n} \geq \frac{1}{K} > \varrho_1^2 \tag{35}$$

gives

$$g_n^2(F) \geq \frac{1}{2} \cdot \varrho_1^2.$$

$\square$

The infinite-dimensional example will now be constructed inductively. In the first step, we set $N_1 = 1$ and consider the 1-dimensional Hadamard example (which is of course trivial).

Now, let us assume, that the first j building blocks with dimensions $N_1, N_2, \ldots, N_j$ have already been constructed. We denote by H_j the corresponding sequence spaces and by F_j their unit balls. We set $D_j = N_1 + N_2 + \cdots + N_j$, $K = D_j$ and $n = 2^{D_j}$ and apply Lemma 5.

It follows that if $t = \frac{2(K+2^n)}{K}$ and $N_{j+1} = 2^{t+n}$, then

$$g_n(F_{j+1}) \geq \frac{\sqrt{2}}{2} \cdot \sigma_{D_j+1} = \frac{\sqrt{2}}{2} \cdot \frac{1}{\sqrt{D_j+1}}.$$

The infinite-dimensional sequence space H is then defined as a direct sum of all the Hilbert spaces H_j:

$$H = \bigoplus_{j=1}^{\infty} H_j$$

and F is its unit ball. We observe, that

$$a_{D_j}(F) = g_{D_j}(F) = \frac{1}{\sqrt{D_j+1}}, \qquad j \in \mathbb{N}$$

and

$$a_n(F) = \frac{1}{\sqrt{n+1}} = \frac{1}{\sqrt{2^{D_j}+1}}, \qquad g_n(F) \geq g_n(F_{j+1}) \geq \frac{\sqrt{2}}{2} \cdot \frac{1}{\sqrt{D_j+1}}, \quad j \in \mathbb{N}, \quad n = 2^{D_j}.$$

Hence, for each $n = D_j$, we get $a_n(F) = g_n(F) = \sigma_{n+1}$ and for each $n = 2^{D_j}$, we obtain $g_n(F) \geq \frac{\sqrt{2}}{2} \cdot \tau_n$.

Observe that, very roughly,

$$D_{j+1} \approx 2^{2^{2^{D_j}}},$$

hence the D_j increase very rapidly.

Remark 4. *This example may be easily generalized in the following way. To every sequence $1 = \sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ with $\sum_n \sigma_n^2 = \infty$ and every sequence $\tau_1 \geq \tau_2 \geq \cdots \geq 0$ with $\lim_{n\to\infty} \tau_n = 0$, there is a sequence space H such that for its unit ball F the following holds:*

- $a_n(F) = \sigma_{n+1}$ *for all natural numbers* n,
- $g_n(F) \geq \tau_n$ *for infinitely many natural numbers* n.

Remark 5. *We end this paper with two additional remarks.*

a) The construction in this section can also be done in the case $\sum_n \sigma_n^2 < \infty$. If e.g. $\sigma_n = n^{-\alpha}$ with $\alpha > 1/2$ then it only follows that $g_n(F)$ is larger than $a_n(F)$ by a constant factor $c_\alpha > 1$ for infinitely many n with $c_\alpha \to \infty$ for $\alpha \to 1/2$.

b) In our example with $p^{\mathrm{all}}(F) = 1/2$ and $p^{\mathrm{std}}(F) = 0$ we used the sequence space ℓ_2 instead of, say, $L_2([0,1])$. This is not essential, however, since we could easily translate our example using piecewise constant functions in $L_2([0,1])$.

c) In this paper we only consider deterministic algorithms. The randomized setting is studied in [8].

Acknowledgment

The problem studied in this paper was presented to us by Henryk Woźniakowski during his long stay in Jena, from November 2006 till July 2007. Henryk also read an early draft of this paper and gave many important comments.

References

[1] S. Dahlke, E. Novak, W. Sickel, Optimal approximation of elliptic problems by linear and nonlinear mappings I. *J. Complexity* **22**, 29–49, 2006.

[2] D. L. Donoho, Unconditional bases are optimal bases for data compression and for statistical estimation, Appl. Comput. Harmon. Anal. **1**, 100–115, 1993.

[3] F. Y. Kuo, G. W. Wasilkowski and H. Woźniakowski, On the power of standard information for multivariate approximation in the worst case setting, submitted for publication.

[4] V. N. Temlyakov, Nonlinear methods of approximation, *Found. Comput. Math.* **3**, 33–107, 2003.

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

[6] J. Vybiral, Sampling numbers and function spaces, *J. Complexity* **23**, 773–792, 2007.

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

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

[9] A. G. Werschulz, *The Computational Complexity of Differential and Integral Equations: an Information-Based Approach*, Oxford University Press, New York, 1991.