Forum of Mathematics, Sigma (2025), **Vol. 13**:e152 1–23
doi:10.1017/fms.2025.10101

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

On the power of adaption and randomization

David Krieg [1], Erich Novak [2] and Mario Ullrich [3]

[1]Faculty of Computer Science and Mathematics, University of Passau, 94032 Passau, Germany;
E-mail: david.krieg@uni-passau.de (corresponding author).
[2]Institute of Mathematics, Friedrich Schiller University, 07743 Jena, Germany; E-mail: erich.novak@uni-jena.de.
[3]Institute for Analysis & Department of Quantum Computing, Johannes Kepler University, 4040 Linz;
E-mail: mario.ullrich@jku.at.

Received: 26 March 2025; **Revised:** 31 July 2025; **Accepted:** 31 July 2025

2020 Mathematical Subject Classification: *Primary* – 47N40; *Secondary* – 41A25, 41A45, 41A46

Abstract

We present bounds on the maximal gain of adaptive and randomized algorithms over nonadaptive, deterministic ones for approximating linear operators on convex sets. If the sets are additionally symmetric, then our results are optimal. For nonsymmetric sets, we unify some notions of n-widths and s-numbers, and show their connection to minimal errors. We also discuss extensions to nonlinear widths and approximation based on function values, and conclude with a list of open problems.

Contents

1. Introduction and summary

Let X and Y be (real) Banach spaces, $S \in \mathcal{L}(X,Y)$, that is, a continuous linear mapping between X and Y, and $F \subset X$. We usually assume that F is convex, and sometimes also that F is symmetric. The goal is to approximate $S(f)$ for arbitrary $f \in F$ by an algorithm $A_n\colon\ F \longrightarrow Y$ that has access to the values of at most n linear functionals (aka measurements) applied to f; see Section 2 for precise definitions. Here, we ask the following question:

How much can be gained by choosing the functionals
adaptively and/or randomly?

In this paper, we present several upper bounds on the largest possible gain. In the case that F is not only convex, but also symmetric, we can apply known relations of minimal worst-case errors and s-numbers as well as inequalities between different s-numbers. In the case that F is only convex, much less is known and new concepts are required. Such nonsymmetric model classes F appear quite naturally, for example, if the problem instances $f \in F$ are non-negative, monotone or convex functions. They may behave very differently compared to symmetric classes, as we discuss below. We also consider the maximal gain if only one of the two features, that is, adaption or randomization, is allowed, and present an upper bound if the measurements are given by n function evaluations instead of arbitrary linear functionals.

Let us now describe the state of the art and our main results in more detail. We start by discussing the power of adaption: How much better are algorithms that are allowed to choose information successively depending on already observed information, compared to those that apply the same n measurements to all inputs? This is sometimes called the "adaption problem." Note that we compare all algorithms that use the same amount of information, regardless of their computational cost.

In the deterministic setting, if F is additionally symmetric, it is known that the answer is *almost nothing*. More precisely, the minimal worst-case error that can be achieved with adaptive algorithms improves upon the one achievable with nonadaptive algorithms by a factor of at most two, see [2, 9, 15, 49, 51, 67, 68]. For nonsymmetric sets, it was proved in [48] that the largest possible gap between those errors is bounded above by $4(n+1)^2$.

For a long time, it was not known whether adaption helps for randomized algorithms if the input set F is convex and symmetric. The problem was posed in [49] and restated in [51, Open Problem 20]. This open problem was recently solved in the affirmative by Stefan Heinrich [22, 23, 24, 25] who studied (parametric) integration and approximation in mixed $\ell_p(\ell_q)$-spaces using standard information (function evaluations). We stress that in this paper we mainly consider arbitrary linear information, hence the setting is different.

For randomized algorithms using arbitrary linear information, the paper [38] shows that one may gain by adaption a factor of main order $n^{1/2}$ for the embedding $S\colon \ell_1^m \to \ell_2^m$ if F is the unit ball of ℓ_1^m. It is proved in [39] that the same gain occurs for the embedding $S\colon \ell_2^m \to \ell_\infty^m$ and one may even gain a factor of main order n for the embedding $S\colon \ell_1^m \to \ell_\infty^m$. In these results, the dimension m is chosen in (exponential) dependence of n and hence the problem S depends on n. Both papers also show how one can obtain from this a single infinite-dimensional problem, where adaption gives a speed-up of the respective main order for all $n \in \mathbb{N}$ by using a construction similar to the one proposed in [25].

In this paper, we give upper bounds for the maximal gain of randomized adaptive algorithms (the most general kind) over deterministic nonadaptive algorithms (the least general kind). We denote the corresponding n-th minimal worst-case errors for approximating S over F by $e_n^{\mathrm{ran}}(S,F)$ and $e_n^{\text{det-non}}(S,F)$, see Section 2. Our main result reads as follows; see Theorem 5.1 for a slightly stronger version and its proof.

Theorem 1.1. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X,Y)$. For every convex $F \subset X$ and $n \in \mathbb{N}$, we have*

$$e_{2n-1}^{\text{det-non}}(S,F) \;\le\; 12\, n^{3/2} \left(\prod_{k<n} e_k^{\mathrm{ran}}(S,F) \right)^{1/n}.$$

In special cases, the following improvements hold:

a) *if F is symmetric, we can replace $n^{3/2}$ with n,*
b) *if Y is a Hilbert space, we can replace $n^{3/2}$ with n,*

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

c) *if F is symmetric and Y a Hilbert space, replace $n^{3/2}$ with $n^{1/2}$,*
d) *if X is a Hilbert space and F its unit ball, we can replace $n^{3/2}$ with $n^{1/2}$ if we additionally replace the index $2n-1$ with $4n-1$.*

Although these bounds are of a nonasymptotic nature, see Corollary 5.2, they might be most easily understood in terms of the polynomial rate of convergence. For this, one has to realize that the geometric mean on the right-hand side has the same polynomial rate of convergence as the error numbers e_n^{ran}. Hence, we find that adaption and randomization improve the rate of convergence by no more than 1 in the symmetric case and by no more than $3/2$ in the nonsymmetric case. This maximal improvement is further reduced by $1/2$ if either the input or the target space is a Hilbert space. In the case that X and Y are Hilbert spaces and F the unit ball of X, then there is no gain (up to constants), see [46] and Lemma 4.3.

By recalling the aforementioned results from [38, 39], we see that our results for the polynomial rate of convergence are sharp in the case of symmetric classes F. We summarize the new state of the art for the adaption and randomization problem in Table 1. The same results hold for the adaption problem in the randomized setting. See also Section 6.1 and Table 2 for an individual discussion of adaption and randomization. For comparison, recall that adaption gives no speed-up for deterministic algorithms for all convex and symmetric classes F.

A crucial tool in our analysis are inequalities between *s-numbers* of operators, see, for example, [55, 56], and between variants of those numbers for the nonsymmetric case, see Section 3.

Indeed, the *Gelfand numbers* c_n characterize the error $e_n^{\text{det-non}}$ of deterministic and nonadaptive algorithms up to a factor of two. On the other hand, it is known that the *Bernstein numbers* b_n are a lower bound for the error of deterministic and adaptive algorithms, see, for example, [48]. More recently, based on earlier results of [21], it has been proven in [36, 37] that also the error e_n^{ran} of adaptive randomized algorithms is bounded below by the Bernstein numbers. Hence, one can obtain bounds for the ratio $e_n^{\text{det-non}}/e_n^{\text{ran}}$ from corresponding bounds involving c_n and b_n, which is the approach of this paper.

For the symmetric case, such bounds follow from the already available estimates on the maximal difference between arbitrary s-numbers, see [55, 56] and the recent paper [70]. For the nonsymmetric case, we will use similar concepts and proof ideas. In particular, we will introduce the Hilbert width h_n as a substitute of the Hilbert numbers, that is, the smallest s-numbers, and prove bounds between c_n and h_n similar to our Theorem 1.1, see Theorem 3.3.

There are many questions which remain unanswered, even despite all the recent progress on the matter of adaption and randomization. For instance, Table 1 neglects any logarithmic factors and it is probably a very hard problem to determine the correct behavior of the maximal gain including logarithmic factors, even in the symmetric case. In the nonsymmetric case, we do not even know the right polynomial order of the maximal gain. Moreover, what is the maximal gain of nonadaptive randomized algorithms over nonadaptive deterministic algorithms? We give a list of open problems in Section 6.

Possibly the most interesting open problem is the following: How do the results change if we switch from algorithms that use arbitrary linear functionals to algorithms that are only allowed to use function evaluations? (In information-based complexity this type of information is called *standard information*.) We guess that the results are, under suitable conditions, quite similar, but so far have not found the right ideas for a proof.

Table 1. *Maximal gain in the rate of convergence of adaptive randomized over nonadaptive deterministic algorithms using linear information. The same table applies for the comparison of adaptive randomized with nonadaptive randomized algorithms.*

Y \ F	unit ball of a Hilbert space	convex & symmetric	only convex
Hilbert space	no gain	1/2	≤ 1
Banach space	1/2	1	$\leq 3/2$

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

There are results on this question, but mostly for particular S and F. The techniques of our paper can be easily adapted to standard information in the case of uniform approximation on convex subsets of $B(D)$, the space of bounded functions on a set D. That is, we consider $X = Y = B(D)$, equipped with the sup-norm on D, and $S = \mathrm{APP}_\infty$ being the identity on $B(D)$.

We obtain that algorithms that only use function evaluations obey the same upper bounds as given in Theorem 1.1, see below and Section 6.3. Here, we only present the interesting special case that F is convex and symmetric. In this case, it is known that we can restrict ourselves to linear algorithms, see [9] or [51, Thm. 4.8]. Using this, we obtain bounds on the *linear sampling numbers*. For $F \subset B(D)$, those are defined by

$$g_n^{\mathrm{lin}}(\mathrm{APP}_\infty, F) \;:=\; \inf_{\substack{x_1,\dots,x_n\in D\\ \varphi_1,\dots,\varphi_n\in B(D)}} \; \sup_{f\in F} \Big\| f - \sum_{i=1}^n f(x_i)\,\varphi_i \Big\|_{B(D)} .$$

One might argue that linear sampling algorithms are the simplest type of algorithms, which are not only nonadaptive, deterministic, and linear but also only employ very restrictive (but natural) information.

The following theorem bounds the error of linear sampling algorithms with the error of general algorithms which may be nonlinear, randomized, adaptive, and based on arbitrary linear information.

Theorem 1.2. *Let D be a set, F be a convex and symmetric subset of $B(D)$, and APP_∞ be the identity on $B(D)$. Then, for all $n \in \mathbb{N}$, we have*

$$g_{2n-1}^{\mathrm{lin}}(\mathrm{APP}_\infty, F) \;\le\; 6n \left(\prod_{k<n} e_k^{\mathrm{ran}}(\mathrm{APP}_\infty, F) \right)^{1/n} .$$

If F is the unit ball of a Hilbert space, we can replace the factor n with $n^{1/2}$ if we additionally replace the index $2n-1$ with $4n-1$.

Theorem 1.2 is optimal in the sense that the factor n cannot be replaced by a lower-order term. This follows again by considering the embedding $S : \ell_1^m \to \ell_\infty^m$ as discussed in [39]. See Section 6.3 for some details, extensions, as well as remarks on this setting. Theorem 1.2 is proven by Theorem 6.4, a common generalization of Theorems 1.1 and 1.2.

2. Algorithms and minimal errors

In general, a deterministic algorithm $A_n\colon F \to Y$ is an arbitrary mapping of the form $A_n = \varphi_n \circ N_n$ with $N_n\colon F \to \mathbb{R}^n$ being the *information mapping*, and $\varphi_n\colon \mathbb{R}^n \to Y$ the *reconstruction mapping*. We mostly pose no restriction at all on the mappings φ_n and focus on the form of N_n; see also Section 6. The most general form we consider is that an information mapping is given recursively by

$$N_n(f) = (N_{n-1}(f),\, L_n(f)),$$

where the choice of the n-th linear functional $L_n = L_n(\cdot, N_{n-1}(f))$ may depend on the first $n-1$ measurements. This is called an *adaptive* choice of information, and we denote the collection of all such algorithms by $\mathcal{A}_n^{\mathrm{det}}(F,Y)$, or just $\mathcal{A}_n^{\mathrm{det}}$.

An algorithm is called *nonadaptive* if $N_n = (L_1, \dots, L_n)$, that is, the same functionals are used for every input, and we denote by $\mathcal{A}_n^{\text{det-non}}$ the corresponding class of algorithms.

Let us add that the assumption that measurements are given by linear functionals is very common in numerical analysis and approximation theory. However, also other concepts are possible. We shortly discuss this in Section 6.2.

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

For an algorithm $A_n \in \mathcal{A}_n^*$ with $* \in \{\mathrm{det}, \mathrm{det\text{-}non}\}$, a mapping $S\colon X \to Y$ and a set $F \subset X$, we define the *worst-case error* of A_n for approximating S over F by

$$e(A_n, S, F) := \sup_{f \in F} \|S(f) - A_n(f)\|_Y .$$

(Note that we omit the Y in $\|\cdot\|_Y$ when no confusion is possible.)

Randomized algorithms are random variables whose realizations are deterministic algorithms as described above.

A randomized algorithm is a family of deterministic algorithms $A_n = (A_n^\omega)_{\omega\in\Omega} \subset \mathcal{A}_n^{\mathrm{det}}(F, Y)$ which is indexed by a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For technical reasons, we assume that the mapping $(f, \omega) \mapsto \|S(f) - A_n^\omega(f)\|_Y$ is $(\mathcal{B}_F \otimes \mathcal{A},\, \mathcal{B}_Y)$-measurable, where $\mathcal{B}_Y$ denotes the Borel σ-algebra on Y, the set F is assumed to be convex, and $\mathcal{B}_F$ denotes the Borel σ-algebra of the topology associated with F, that is, with respect to the semi-norm whose unit ball is the convex and symmetric set $F - F$. Then, formally, the desirable statement $\mathcal{A}_n^{\mathrm{det}} \subset \mathcal{A}_n^{\mathrm{ran}}$ is not correct since we do not assume that a deterministic algorithm has to be measurable. See [51, Section 4.3.3] and Section 6 for a discussion of this technicality.

We denote the class of all such (possibly adaptive) algorithms by $\mathcal{A}_n^{\mathrm{ran}}(F, Y)$ and let $\mathcal{A}_n^{\mathrm{ran\text{-}non}}(F, Y)$ be the class of randomized algorithms whose realizations are nonadaptive. Again, we may omit the dependence on F and Y. We define the *worst-case error* of a randomized algorithm $A_n \in \mathcal{A}_n^{\mathrm{ran}}(F, Y)$ for approximating S over F by

$$e(A_n, S, F) := \sup_{f \in F} \mathbb{E}\, \|S(f) - A_n(f)\|_Y .$$

In order to compare the power of the just introduced types of algorithms, we now define the *n-th minimal worst-case error* for approximating S over F by

$$e_n^*(S, F) := \inf_{A_n \in \mathcal{A}_n^*} e(A_n, S, F),$$

where $* \in \{\mathrm{det}, \mathrm{det\text{-}non}, \mathrm{ran}, \mathrm{ran\text{-}non}\}$.

The respective concepts can indeed lead to very different minimal errors. Several examples, remarks, and open problems will be presented in Section 6.

3. Widths and s-numbers

Widths have a long tradition in approximation theory and there is a whole range of widths of sets within normed spaces. See, for example, Tikhomirov [66] and Ismagilov [26] for early treatments, and Pinkus [58] or Lorentz et al. [40] for books on the subject. A somehow competing concept are *s-numbers* of operators which play an important role in operator theory and geometry of Banach spaces, see Pietsch [55, 56]. A short account of their history and potential differences can be found in [56, 6.2.6], see also [13, 20]. Some of these widths and numbers have an obvious relation to algorithms, and hence to information-based complexity, while others are seemingly unrelated. We will discuss some known relations in Section 4. However, we first study the relation of the relevant numbers among each other.

We start by providing a common generalization of the above concepts. That is, we introduce various s-numbers of a mapping $S \in \mathcal{L}(X, Y)$ on a subset $F \subset X$. Alternatively, one may call them widths of a set $F \subset X$ with respect to a mapping $S \in \mathcal{L}(X, Y)$.

The original definitions of the corresponding widths of sets $F \subset X$ are obtained by considering the s-numbers of the identity id_X on $F \subset X$ (or the width of F w.r.t. id_X), while s-numbers of the operator S are recovered by considering $F = B_X$ (or the width of B_X w.r.t. S).

Here and in the following, the (closed) unit ball of X is denoted by B_X and the continuous dual space of X by X'.

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

We define the *Gelfand numbers* of $S \in \mathcal{L}(X,Y)$ on $F \subset X$ by

$$\begin{aligned} c_n(S,F) &:= \inf_{L_1,\dots,L_n \in X'} \sup_{\substack{f,g\in F:\\ L_k(f)=L_k(g)}} \frac{1}{2}\big\|S(f)-S(g)\big\| \\ &= \inf_{\substack{M\subset X \text{ closed}\\ \operatorname{codim}(M)\le n}} \sup_{\substack{f,g\in F:\\ f-g\in M}} \frac{1}{2}\big\|S(f)-S(g)\big\|. \end{aligned}$$

In particular, $c_0(S,F) = \frac{1}{2}\mathrm{diam}\big(S(F)\big)$. Note that in the theory of s-numbers, there is usually an index shift of one and "s_n" is only considered for $n \ge 1$ (such that $s_1(S) = \|S\| = \frac{1}{2}\mathrm{diam}\big(S(B_X)\big)$). We use a different convention here because n is used for the amount of information. It is well-known, and we will present the details in Section 4, that the Gelfand numbers c_n are closely related to $e_n^{\text{det-non}}$.

Other quantities that will serve as lower bounds for all minimal errors are the *Bernstein numbers* of S on F, which are defined by

$$\begin{aligned} b_n(S,F) \;:=\; \sup_{\substack{\dim(V)=n+1\\ S \text{ injective on } V}} \sup\big\{r>0 \,:\, & g+B\subset F \text{ for some } g\in F \\ & \text{and a ball } B \text{ of radius } r \text{ in } (V,\|\cdot\|_S)\big\}. \end{aligned}$$

Here, we consider the norm on the linear space V that is induced by S, that is, $\|x\|_S := \|Sx\|_Y$. If F is convex and symmetric, it suffices to consider balls centered at the origin in the above definition. We note again that these numbers coincide with the classical Bernstein widths if S is the identity on X. In the special case that F is a bounded subset of X, it is not hard to verify that we have the handy formula

$$b_n(S,F) \;=\; \sup_{\substack{V\subset X \text{ affine}\\ \dim(V)=n+1}} \sup_{g\in F\cap V} \inf_{f\in V\cap(X\setminus F)} \|S(f)-S(g)\|.$$

Remark 3.1. The paper [48] considers instead the Bernstein widths of the set $S(F)$ in Y, that is, the radius of the largest $(n+1)$-dimensional ball contained in $S(F)$. This coincides with the Bernstein numbers of S on F as defined above if S is injective. If S is not injective, the widths of the set $S(F)$ may be larger.

We will see in Section 4, that one obtains bounds on $e_n^{\text{det-non}}/e_n^{\text{ran}}$ from corresponding bounds involving c_n and b_n, which is the approach of this paper. For this, we want to employ proof ideas that have already been used for bounding the maximal difference between s-numbers, see, for example, [55, 56]. Inspired by Hilbert numbers, see [4], which are the smallest s-numbers, we introduce the *Hilbert numbers* of $S \in \mathcal{L}(X,Y)$ on $F \subset X$ by

$$\begin{aligned} h_n(S,F) \;:=\; \sup\Big\{ & c_n\big(BSA, B_{\ell_2}\big) :\; B\in\mathcal{L}(Y,\ell_2) \text{ with } \|B\|\le 1, \\ & A\in\mathcal{L}(\ell_2,X) \text{ and } x\in F \text{ with } A(B_{\ell_2})+x\subset F\Big\}. \end{aligned}$$

In this definition, we can replace c_n with b_n since both numbers coincide for operators $T \in \mathcal{L}(\ell_2,\ell_2)$; they are both equal to the $(n+1)$-st singular value of T, see, for example, [53].

One of the key ingredients to our results will be bounds between these numbers. First, note that they are related in the same way as the corresponding s-numbers, see [4, 53].

Proposition 3.2. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X,Y)$. For every convex $F \subset X$ and $n \in \mathbb{N}_0$, we have*

$$h_n(S,F) \;\le\; b_n(S,F) \;\le\; c_n(S,F).$$

Equalities hold if X and Y are Hilbert spaces and $F = B_X$.

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

Proof of Proposition 3.2. In order to prove $b_n \geq h_n$, let $B \in \mathcal{L}(Y, \ell_2)$ with $\|B\| \leq 1$ as well as $A \in \mathcal{L}(\ell_2, X)$ and $g \in X$ with $A(B_{\ell_2}) + g \subset F$. For any $\beta < b_n\big(BSA, B_{\ell_2}\big)$, there exists an $(n+1)$-dimensional linear space $V \subset \ell_2$ such that BSA is injective on V and for any $v \in V$ it holds that $\|BSAv\|_2 \leq \beta$ implies $v \in B_{\ell_2}$. First, we observe that the injectivity of BSA implies that S is injective on $W = A(V)$ and that W is $(n+1)$-dimensional. Moreover, let $f \in W$ with $\|Sf\|_Y \leq \beta$. Choose $v \in V$ with $f = Av$. We have

$$\|BSAv\|_2 \leq \|SAv\|_Y = \|Sf\|_Y \leq \beta$$

and hence $v \in B_{\ell_2}$. By assumption, this implies $f + g \in F$ for all such f. Hence, F contains an $\|\cdot\|_S$-ball of radius β in W and we have $b_n(S, F) \geq \beta$. Taking the supremum over all β gives

$$b_n(S, F) \ \geq \ b_n(BSA, B_{\ell_2})$$

and taking the supremum over all B, A, and g as above gives

$$b_n(S, F) \ \geq \ h_n(S, F).$$

In order to show $c_n \geq b_n$, let $\beta < b_n(S, F)$ be arbitrary and let $V \subset X$ be an $(n+1)$-dimensional subspace such that S is injective on V as well as $m \in F$ such that $h \in V$ and $\|Sh\|_Y \leq \beta$ imply $m + h \in F$. Now, for all $L_1, \ldots, L_n \in X'$, there must be some $h \in V \setminus \{0\}$ with $L_i(h) = 0$ for all $i \leq n$. We choose h such that $\|Sh\|_Y = \beta$, which implies $f = m + h \in F$ and $g = m - h \in F$. Note that $L_i(f) = L_i(g)$ for all $i \leq n$. Moreover, $\frac{1}{2}\|Sf - Sg\| = \beta$ and hence $c_n(S, F) \geq \beta$. □

Here, we prove the following reverse inequalities, which are reminiscent to the corresponding bounds for s-numbers, see Remark 3.4.

Theorem 3.3. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X, Y)$. For every convex $F \subset X$ and $n \in \mathbb{N}$, we have*

$$c_{n-1}(S, F) \ \leq \ \left(\prod_{k=0}^{n-1} c_k(S, F) \right)^{1/n} \ \leq \ n^{3/2} \left(\prod_{k=0}^{n-1} h_k(S, F) \right)^{1/n}.$$

In special cases, the following improvements hold:

a) *if F is symmetric, we can replace the exponent $3/2$ with 1,*
b) *if Y is a Hilbert space, we can replace the exponent $3/2$ with 1,*
c) *if F is symmetric and Y a Hilbert space, replace $3/2$ with $1/2$,*
d) *if X is a Hilbert space and F its unit ball, we can replace the exponent $3/2$ with $1/2$ if we also replace all c_k with c_{2k}.*

Proof of Theorem 3.3. Let $S \in \mathcal{L}(X, Y)$ and $F \subset X$ be convex.

General case: We first show that, for fixed $\varepsilon > 0$, we can find $f_0, g_0, f_1, g_1, \ldots \in F$ and $L_0, L_1, \ldots \in X'$ such that, with $p_k := \frac{f_k - g_k}{2}$, we have $L_j(p_k) = 0$ for $j < k$ and $(1+\varepsilon)L_k(p_k) > c_k(S, F)$ for $k = 0, 1, \ldots$. See [48, p. 132] for a similar proof.

The proof is by induction. Let $k \in \mathbb{N}_0$ and assume that f_j, g_j and L_j for $j < k$ are already found. Define

$$M_k \ := \Big\{ p \in X \colon \ L_j(p) = 0 \quad \text{for} \quad j < k \Big\}.$$

Since $\operatorname{codim} M_k \leq k$, we can choose $f_k, g_k \in F$ with $p_k \in M_k$ and

$$(1+\varepsilon)\, \|Sp_k\| \ \geq \ c_k(S, F). \tag{3.1}$$

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

By the Hahn-Banach theorem, there is $\lambda_k \in B_{Y'}$ with $\lambda_k(Sp_k) = \|Sp_k\|$ and hence

$$\lambda_k(Sp_k) \;\geq\; (1+\varepsilon)^{-1}\, c_k(S,F). \tag{3.2}$$

We finish the induction step by setting $L_k = \lambda_k \circ S \in X'$.

For $n \in \mathbb{N}$, we now define $g = \frac{1}{n}\sum_{i<n} \frac{f_i+g_i}{2} \in F$ and the operators

$$A(\xi) \;:=\; \frac{1}{n}\sum_{i<n} \xi_i p_i \in X, \quad \xi = (\xi_i)_{i<n} \in \ell_2^n, \tag{3.3}$$

and

$$B(y) \;:=\; \frac{1}{\sqrt{n}}\big(\lambda_i(y)\big)_{i<n} \in \ell_2^n, \qquad y \in Y,$$

and consider the mapping $S_n := BSA$. We observe that $\|B\| \leq 1$ and, for all $\xi \in [-1,1]^n$, due to convexity, it holds

$$A(\xi) + g \;=\; \frac{1}{n}\sum_{i<n}\left(\frac{1+\xi_i}{2} f_i + \frac{1-\xi_i}{2} g_i\right) \in F.$$

In particular, $A(B_{\ell_2^n}) + g \subset F$. This gives, for any $k < n$,

$$c_k(S_n, B_{\ell_2^n}) \leq h_k(S,F).$$

Our bounds are obtained by considering the determinant of $S_n\colon\ \ell_2^n \to \ell_2^n$. Since S_n is generated by the triangular matrix $n^{-3/2}(L_j(p_i))_{i,j<n}$, we have

$$\det(S_n) \;\geq\; \prod_{k<n} \frac{c_k(S,F)}{n^{3/2}(1+\varepsilon)}.$$

On the other hand, the determinant is multiplicative and equals the product of the singular values, which in turn equal the Gelfand widths $c_k(S_n, B_{\ell_2^n})$. Therefore, we obtain

$$\prod_{k<n} \frac{c_k(S,F)}{n^{3/2}(1+\varepsilon)} \;\leq\; \det(S_n) \;=\; \prod_{k<n} c_k(S_n, B_{\ell_2}) \;\leq\; \prod_{k<n} h_k(S,F).$$

Taking the infimum over all $\varepsilon > 0$ and using $c_k(S,F) \geq c_{n-1}(S,F)$ for $k < n$, we obtain

$$c_{n-1}(S,F) \;\leq\; n^{3/2}\left(\prod_{k<n} h_k(S,F)\right)^{1/n}.$$

<u>F symmetric</u>: If F is additionally symmetric, then one has $p_i = \frac{f_i - g_i}{2} \in F$ such that we can redefine A by $A(\xi) := \frac{1}{\sqrt{n}}\sum_{i<n} \xi_i p_i$ to have $A(B_{\ell_2}) \subset F$. Hence, we can continue with the triangular matrix $S_n = n^{-1}(L_j(p_i))_{i,j<n}$ to obtain the improved bound.

<u>Y Hilbert space</u>: If Y is a Hilbert space, we can choose the functionals L_k from the Hahn-Banach theorem explicitly as

$$L_k \;:=\; \left\langle \,\cdot\,,\, \frac{Sp_k}{\|Sp_k\|_Y} \right\rangle,$$

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

where $\langle \cdot, \cdot \rangle$ is the inner product in Y. Hence, by the definition of the sets M_k, we see that the Sp_k for $k \le n$ are pairwise orthogonal. We can hence skip the factor $n^{-1/2}$ in the definition of B and put $B(y) := (\lambda_i(y))_{i<n}$ while preserving the property $\|B\| \le 1$. Thus, also in this case, we can continue with the triangular matrix $S_n = n^{-1}(L_j(p_i))_{i,j<n}$.

F symmetric, Y Hilbert space: We combine the modifications from the previous two cases and work with the matrix $S_n = n^{-1/2}(L_j(p_i))_{i,j<n}$.

F unit ball of a Hilbert space: The proof is similar to the general case. We show by induction that for fixed $\varepsilon > 0$, there are orthogonal vectors $p_0, p_1, \ldots \in B_X$ and $L_0, L_1, \ldots \in X'$ such that $L_j(p_k) = 0$ for $j < k$ and $(1+\varepsilon)L_k(p_k) \ge c_{2k}(S, B_X)$ for $k = 1, 2, \ldots$

Assume that p_j, L_j for $j < k$ are already found, and define

$$M_n := \Big\{ p \in X\colon\ L_j(p) = 0 \text{ and } \langle p_k, p \rangle = 0 \text{ for } j < k \Big\}.$$

Since $\operatorname{codim} M_k \le 2k$, we can choose $p_k \in B_X$ with $p_k \in M_k$ and

$$(1+\varepsilon)\|Sp_k\| \ \ge\ c_{2k}(S, B_X).$$

By the Hahn-Banach theorem, there is $\lambda_k \in B_{Y'}$ with $\lambda_k(Sp_k) = \|Sp_k\|$.

For $n \in \mathbb{N}$, we define the operators $A \in \mathcal{L}(\ell_2^n, X)$ and $B \in \mathcal{L}(Y, \ell_2^n)$ by

$$A(\xi) := \sum_{i<n} \xi_i p_i, \qquad B(y) := \frac{1}{\sqrt{n}} \big(\lambda_i(y)\big)_{i<n}$$

such that $\|A\| \le 1$ and $\|B\| \le 1$. The mapping $S_n := BSA$ is generated by the triangular matrix $n^{-1/2}(L_j(p_i))_{i,j<n}$, where $L_j := \lambda_j \circ S$. This gives

$$\prod_{k<n} \frac{c_{2k}(S, B_X)}{n^{1/2}(1+\varepsilon)} \ \le\ \det(S_n) \ =\ \prod_{k<n} c_k(S_n, B_{\ell_2^n}) \ \le\ \prod_{k<n} h_k(S, B_X).$$

Taking the infimum over $\varepsilon > 0$ and using $c_{2k} \le c_{2n-2}$ for $k < n$, we get

$$c_{2n-2}(S, F) \ \le\ n^{1/2} \left(\prod_{k<n} h_k(S, F) \right)^{1/n}. \qquad \square$$

Remark 3.4. a) Note the "oversampling" in Theorem 3.3 in the case that the input space is a Hilbert space, where we consider c_{2n-2} on the left hand side. We do not know if this is necessary.

b) We mentioned above that the Gelfand and Hilbert numbers coincide with the corresponding s-numbers if F is the unit ball of X. In this case, Theorem 3.3 was known, see [55, 2.10.7] and [54, Theorem 11.12.3] or [70] for a streamlined presentation.

It may be desirable to compare c_n directly with h_n instead of the geometric mean of $h_0, \ldots, h_n$. Under the regularity condition that h_k / h_{2k} is bounded, such a comparison is obtained from the following lemma.

Lemma 3.5. *Let $n \in \mathbb{N}$ be even and $z_1 \ge \ldots \ge z_n > 0$. Moreover, let $c > 0$ with $z_k \le c\, z_{2k}$ for all $k \le n/2$. Then*

$$\left(\prod_{k=1}^{n} z_k \right)^{1/n} \ \le\ c^4\, z_n.$$

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

Proof of Lemma 3.5. Choose $\nu \in \mathbb{N}_0$ such that $n/2 < 2^\nu \le n$. For $2^j \le k \le n$, we get

$$z_k \le z_{2^j} \le c^{\nu-j} z_{2^\nu} \le c^{\nu-j} z_{n/2} \le c^{\nu-j+1} z_n$$

so that

$$\left(\prod_{k=1}^{n} z_k\right)^{1/n} \le \left(\prod_{j=0}^{\nu} \prod_{2^j \le k < 2^{j+1}} c^{\nu-j+1}\right)^{1/n} \cdot z_n = c^\kappa\, z_n$$

with

$$\kappa = \frac{1}{n}\sum_{j=0}^{\nu}(\nu - j + 1)2^j = \frac{2^{\nu+1}}{n}\sum_{i=1}^{\nu+1} i2^{-i} \le 4. \qquad \square$$

Lemma 3.5 is a fitting estimate for sequences of polynomial decay: If $z_n = n^{-\alpha}$, then $c = 2^\alpha$ is independent of n. For sequences of super-polynomial decay, it might be better to use the simple estimate

$$\left(\prod_{k=1}^{n} z_k\right)^{1/n} \le \sqrt{z_1 \cdot z_{n/2}}. \tag{3.4}$$

In the case that Y is a Hilbert space, there also is the following alternative bound which works for individual n without any regularity condition as in Lemma 3.5. On the downside, the upper bound is in terms of the (possibly larger) Bernstein numbers instead of the Hilbert numbers.

Theorem 3.6. *Let X be a Banach space, H be a Hilbert space and $S \in \mathcal{L}(X, H)$. For every convex $F \subset X$ and $n \in \mathbb{N}_0$, we have*

$$c_n(S, F) \le (n+1) \cdot b_n(S, F).$$

We can replace $(n+1)$ by $\sqrt{n+1}$ if F is additionally symmetric.

Proof of Theorem 3.6. We take g_k, f_k and $p_k = \frac{f_k - g_k}{2}$ from the proof of Theorem 3.3 (general case), and put $r = \frac{c_{n-1}}{1+\varepsilon}$. The n-dimensional space V spanned by $p_0, \ldots, p_{n-1}$ with the norm $\|\cdot\|_S$ is a Hilbert space. The vectors p_k have norm at least r and, as observed in the proof of Theorem 3.3, they are orthogonal in V. If F is convex and symmetric, we have $\pm p_k \in F$ and so F contains a ball of radius $\frac{r}{\sqrt{n}}$ in V. This proves the claim, that is, $b_{n-1}(S, F) \ge \frac{c_{n-1}(S,F)}{\sqrt{n}}$.

In the nonsymmetric case, we already observed that $A(B_{\ell_2^n}) + g \subset F$ with A and g as in (3.3), so that F contains a ball of radius $\frac{r}{n}$. $\square$

The paper [59] contains bounds similar to Theorem 3.6 with the Gelfand widths replaced by the Kolmogorov widths. Since the target space is a Hilbert space, the Kolmogorov widths are larger than the Gelfand widths, see, for example, [58, Prop. 5.2] and Section 6. This means that the bounds of Theorem 3.6 are known up to constants. We presented the proof anyway since the result follows with little effort from our other observations.

Due to their relations with the previously defined minimal worst-case errors (as discussed in the next section), the Gelfand, Bernstein, and Hilbert numbers as considered above are of particular interest to us. Nonetheless, in Section 6, we will mention some other types of widths that may be of independent interest and discuss how Theorem 3.3 applies to these widths.

4. Widths versus minimal errors

In this section, we discuss how the Gelfand, Bernstein, and Hilbert numbers are related to minimal errors and hence obtain bounds between the different types of minimal errors.

First, note that the Gelfand numbers characterize the minimal worst-case error of deterministic algorithms up to a factor of two. This is a special case of a classical result in information-based complexity, see, e.g, [51, Sec. 4.1]. In our setting, the result reads as follows.

Proposition 4.1. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X,Y)$. For every $F \subset X$ and $n \in \mathbb{N}_0$, we have*

$$c_n(S,F) \le e_n^{\text{det-non}}(S,F) \le 2\, c_n(S,F).$$

If F is convex and symmetric then

$$c_n(S,F) \le e_n^{\text{det}}(S,F) \le e_n^{\text{det-non}}(S,F) \le 2\, c_n(S,F).$$

We turn to the relation of Bernstein numbers and minimal errors. It is known, see [48], that $b_n(S,F)$ may serve as a lower bound for the error of adaptive deterministic algorithms.

Proposition 4.2. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X,Y)$. For every $F \subset X$ and $n \in \mathbb{N}_0$, we have*

$$e_n^{\text{det}}(S,F) \ge b_n(S,F).$$

Proof of Proposition 4.1 and 4.2. The technique of the proof is the same for both results and is well known (but a factor of 2 is missing in Proposition 1 of [48]) and hence we concentrate on Proposition 4.2. Let $A_n = \varphi_n \circ N_n$ be an algorithm based on the information $N_n : F \to \mathbb{R}^n$ that might be adaptive. We fix the nonadaptive and linear mapping $N_n^* = (L_1^*, \dots, L_n^*) : F \to \mathbb{R}^n$ that is taken for the midpoint g of a ball $g + B \subset F$. The mapping N_n^* cannot be injective and there exists a point $\tilde{f}$ on the sphere of B with $N^*(\tilde{f} + g) = N^*(g) = N^*(g - \tilde{f})$, hence $N(g + \tilde{f}) = N(g - \tilde{f})$. Then A_n cannot distinguish between the two inputs and we obtain the lower bound. □

In the case that X and Y are Hilbert spaces and F is the unit ball of X, it is shown in [46] that the Bernstein numbers also yield lower bounds for randomized adaptive algorithms. Here we use a slightly different error criterion and hence formulate a lemma.

Lemma 4.3. *Let H and G be Hilbert spaces and $S \in \mathcal{L}(H,G)$. For every $n \in \mathbb{N}_0$, we have*

$$e_n^{\text{ran}}(S,B_H) \ge \frac{1}{2}\, b_{2n-1}(S,B_H) = \frac{1}{2}\, e_{2n-1}^{\text{det-non}}(S,F). \tag{4.1}$$

With a different constant, Lemma 4.3 is implied by [21, Cor. 2].

Proof of Lemma 4.3. Let $0 < b < b_{2n-1}(S,F)$. There exists a subspace $V \subset H$ with dimension $2n$ such that $\|Sf\|_G \ge b\|f\|_H$ for all $f \in V$. Let $W := S(V)$. We choose $R \in \mathcal{L}(\mathbb{R}^{2n}, V)$ and $Q \in \mathcal{L}(W, \mathbb{R}^{2n})$, where $\mathbb{R}^{2n}$ is considered with the Euclidean norm, such that $\|R\| \le 1$ and $\|Q\| \le b^{-1}$ and such that QSR equals the identity id_{2n} on $\mathbb{R}^{2n}$. If A_n is a randomized algorithm for S with error less than $b/2$, then QA_nR is a randomized algorithm for id_{2n} with error less than $1/2$. It hence suffices to prove the lower bound $1/2$ in the case $S = \mathrm{id}_{2n}$.

We let P_{2n} be the uniform distribution on the sphere of $\mathbb{R}^{2n}$. An application of Fubini's theorem (known as Bakhvalov's proof technique, see [3] or [51, Section 4.3.3]) gives

$$e_n^{\text{ran}}(S,F) \ge \inf_{A_n} \int \|f - A_n(f)\| \,\mathrm{d}P_{2n}(f),$$

where the infimum runs over all deterministic and measurable algorithms $A_n \in \mathcal{A}_n^{\text{det}}$. Hence let $A_n = \varphi \circ N_n$ be a measurable deterministic algorithm with adaptively chosen information $N_n = (L_1, \dots, L_n)$ and let f be distributed according to P_{2n}. Assume that the functionals L_i are chosen orthonormal; this

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

is no restriction. For each y in the unit ball of $\mathbb{R}^n$, the information $N_n(f) = y$ defines a sphere $\mathbb{S}_y$ of radius $r_y = \sqrt{1 - \|y\|^2}$. We have

$$\int \|f - A_n(f)\| \,\mathrm{d}P_{2n}(f) = \int \int \|f - \varphi(y)\| \,\mathrm{d}\mu_y(f) \,\mathrm{d}\nu(y)$$

where μ_y is the uniform distribution on $\mathbb{S}_y$ and ν is the distribution of $N_n(f)$. The inner integral is minimized if $\varphi(y)$ equals the center of $\mathbb{S}_y$, so that we have

$$\int \|f - A_n(f)\| \,\mathrm{d}P_{2n}(f) \geq \int r_y \,\mathrm{d}\nu(y) \geq \int r_y^2 \,\mathrm{d}\nu(y).$$

From the symmetry of P_{2n} it follows that ν does not depend on N_n. We choose $N_n(f) = (f_1, \ldots, f_n)$ and get

$$\int \|f - A_n(f)\| \,\mathrm{d}P_{2n}(f) \geq \int \sum_{i=n+1}^{2n} f_i^2 \,\mathrm{d}P_{2n}(f) = \frac{1}{2}.$$

The last identity holds since the f_i^2 are identically distributed so that their expected value equals $1/(2n)$. See Proposition 3.2 for the equality $b_n = e_n^{\text{det-non}}$. □

More recently, it has been shown in [36, 37] that also for Banach spaces X and Y, it holds that

$$e_n^{\text{ran}}(S, F) \geq \frac{1}{30} b_{2n-1}(S, F). \tag{4.2}$$

The result of [37] is proven only in the symmetric case but it remains valid if F is only convex. On the other hand, the result (4.1) for the Hilbert case easily implies the following.

Proposition 4.4. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X, Y)$. For every convex $F \subset X$ and $n \in \mathbb{N}$, we have*

$$e_n^{\text{ran}}(S, F) \geq \frac{1}{2} h_{2n-1}(S, F).$$

Proof of Proposition 4.4. Let $A_n \in \mathcal{A}_n^{\text{ran}}(X, Y)$ and let $B \in \mathcal{L}(Y, \ell_2)$ with $\|B\| \leq 1$ as well as $A \in \mathcal{L}(\ell_2, X)$ and $g \in F$ with $A(B_{\ell_2}) + g \subset F$. Then we have $BA_nA \in \mathcal{A}_n^{\text{ran}}(\ell_2, \ell_2)$. This algorithm uses information of the form $L_k' = L_k A \in \ell_2'$, if $L_k \in X'$ is the information used by A_n. Note that A is continuous in the norm induced by F due to $A(B_{\ell_2}) \subset F - F$ and hence the algorithm is measurable. By (4.1), we have

$$e(BA_nA, BSA, B_{\ell_2}) \geq \frac{1}{2} b_{2n-1}(BSA, B_{\ell_2}).$$

On the other hand,

$$\begin{aligned} e(BA_nA, BSA, B_{\ell_2}) &= e(BA_n(A+g), BS(A+g), B_{\ell_2}) \\ &\leq e(BA_n, BS, F) \leq e(A_n, S, F), \end{aligned}$$

where we used $(A+g)(B_{\ell_2}) \subset F$ in the first and $\|B\| \leq 1$ in the second inequality. So,

$$e(A_n, S, F) \geq \frac{1}{2} b_{2n-1}(BSA, B_{\ell_2}).$$

Taking the supremum over all B, A and g as above gives the result. □

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

We point out that the Hilbert numbers can be much smaller than the Bernstein numbers. For example, if S is the identity on ℓ_1 and F the unit ball of ℓ_1, then the Bernstein numbers are equal to one, see [53], while the Hilbert numbers are of order $n^{-1/2}$, see [55, 2.9.19]. So in general, one should prefer the bound (4.2) over Proposition 4.4. However, since our upper bounds are in terms of the Hilbert numbers anyway, we will obtain a better constant in the overall comparison if we work with Proposition 4.4 instead of (4.2).

5. The main result

We now arrive at our main result, Theorem 1.1, which we present here in a slightly stronger form.

Theorem 5.1. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X,Y)$. For every convex $F \subset X$ and $n \in \mathbb{N}$, we have*

$$\left(\prod_{k<2n} e_k^{\text{det-non}}(S,F)\right)^{1/(2n)} \le 2^{7/2}\, n^{3/2} \left(\prod_{k<n} e_k^{\text{ran}}(S,F)\right)^{1/n}.$$

In special cases, the following improvements hold:

a) *if F is symmetric, we can replace $n^{3/2}$ with n,*
b) *if Y is a Hilbert space, we can replace $n^{3/2}$ with n,*
c) *if F is symmetric and Y a Hilbert space, replace $n^{3/2}$ with $n^{1/2}$,*
d) *if X is a Hilbert space and F its unit ball, we can replace $n^{3/2}$ with $n^{1/2}$ if we also replace the range $k < 2n$ with $k < 4n$.*

Proof of Theorems 1.1 and 5.1. A successive application of Proposition 4.1, Theorem 3.3, the monotonicity of the Hilbert widths, and Proposition 4.4 (in the weaker form $h_{2k} \le 2e_k^{\text{ran}}$ for all $k \in \mathbb{N}_0$) gives

$$\begin{aligned}\prod_{k<2n} e_k^{\text{det-non}}(S,F) &\le 2^{2n} \cdot \prod_{k<2n} c_k(S,F) \le 2^{5n} n^{3n} \cdot \prod_{k<2n} h_k(S,F)\\ &\le 2^{5n} n^{3n} \cdot \prod_{k<n} h_{2k}(S,F)^2 \le 2^{7n} n^{3n} \cdot \prod_{k<n} e_k^{\text{ran}}(S,F)^2.\end{aligned}$$

The modifications in the special cases are obvious. □

Since estimates in terms of geometric means might be unfamiliar to the reader, we present a corollary of Theorem 1.1 which is reminiscent of Carl's inequality.

Corollary 5.2. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X,Y)$. For every convex $F \subset X$, $n \in \mathbb{N}$ and $\alpha > 0$, we have*

$$e_{2n-1}^{\text{det-non}}(S,F) \le C_\alpha\, n^{-\alpha+3/2} \cdot \sup_{k<n} \big((k+1)^\alpha\, e_k^{\text{ran}}(S,F)\big),$$

where $C_\alpha \le 12^{\alpha+1}$. In accordance with the special cases given in Theorem 1.1, the exponent $3/2$ can be replaced with 1 or $1/2$.

Proof of Corollary 5.2. If K denotes the supremum on the right hand side, then $e_k^{\text{ran}}(S,F) \le K(k+1)^{-\alpha}$ for all $k < n$. Now the statement follows from Theorem 1.1 and Lemma 3.5 with $z_n = Kn^{-\alpha}$. □

We also write explicitly the implication for the polynomial rate of convergence, which has been presented in the introduction as Table 1. The polynomial rate of convergence of a sequence $(z_n) \subset [0,\infty)$ is defined by

$$\operatorname{rate}(z_n) := \sup\left\{\alpha > 0 \,\middle|\, \exists C \ge 0 : \forall n \in \mathbb{N} : z_n \le C n^{-\alpha}\right\}. \tag{5.1}$$

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

We only give the result for the symmetric case, where the bounds are sharp up to logarithmic factors. It should be obvious enough what the corresponding results in the nonsymmetric case look like.

Corollary 5.3. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X,Y)$. For every convex and symmetric $F \subset X$, we have*

$$\operatorname{rate}\Big(e_n^{\text{det-non}}(S,F)\Big) \;\geq\; \operatorname{rate}\Big(e_n^{\text{ran}}(S,F)\Big) - 1.$$

If either Y is a Hilbert space or F is the unit ball of a Hilbert space X, we even have

$$\operatorname{rate}\Big(e_n^{\text{det-non}}(S,F)\Big) \;\geq\; \operatorname{rate}\Big(e_n^{\text{ran}}(S,F)\Big) - 1/2.$$

Moreover, in each of these cases, equality can occur.

Proof of Corollary 5.3. The inequalities are implied by Corollary 5.2. Equality occurs for the examples from [39, Cor. 4.2], [38, Rem. 3.4], and [39, Rem. 4.3], respectively. □

6. Examples and related problems

In this section we give further details and extensions of our result and discuss related problems. In particular, we analyze the individual influence of adaption and randomization on the minimal worst-case error, and present those examples which exhibit the largest gain known to us. In addition, we show how our results apply to other *nonlinear widths* and to approximation based on standard information, that is, function evaluations, as shown in Theorem 1.2. We also present a list of open problems.

6.1. The individual power of adaption and randomization

By the results of the previous sections, we know how much adaption and randomization can help if they are allowed *together*. That is, we have a good understanding of the maximal gain from $\mathcal{A}_n^{\text{det-non}}$ to $\mathcal{A}_n^{\text{ran}}$. In the symmetric case, we even know that our bounds are optimal up to logarithmic factors. However, our knowledge about the individual power of randomization or adaption still has several gaps.

Let us first talk about upper bounds. Intuitively, it is clear that the gain of randomization or adaption alone cannot be larger than the gain of adaption and randomization together. Let us make this a corollary.

Corollary 6.1. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X,Y)$. For every convex and bounded $F \subset X$ and $n \in \mathbb{N}$, we have*

$$e_{2n-1}^{*}(S,F) \;\leq\; Cn^{3/2}\left(\prod_{k<n} e_k^{\square}(S,F)\right)^{1/n},$$

where $*, \square \in \{\text{det}, \text{det-non}, \text{ran}, \text{ran-non}\}$ and C is a universal constant. If X or Y is a Hilbert space or if F is symmetric, the improvements of Theorem 1.1 apply.

This corollary is not as obvious as it seems at first glance. The problem is that, due to the assumed measurability of randomized algorithms, we do *not* have the relation $\mathcal{A}_n^{\text{det-non}} \subset \mathcal{A}_n^{\text{ran-non}}$. What we do have is the relation $\mathcal{A}_n^{\text{det-non-mb}} \subset \mathcal{A}_n^{\text{ran-non}}$, where $\mathcal{A}_n^{\text{det-non-mb}}$ denotes the class of all $(\mathcal{B}_F, \mathcal{B}_Y)$-measurable deterministic and nonadaptive algorithms with the corresponding minimal worst-case error denoted by $e_n^{\text{det-non-mb}}$. The issue is fixed by the following lemma, which shows that measurable deterministic and nonadaptive algorithms are (roughly) as good as arbitrary deterministic nonadaptive algorithms, that is, there is no real difference between $\mathcal{A}_n^{\text{det-non-mb}}$ and $\mathcal{A}_n^{\text{det-non}}$.

Lemma 6.2. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X,Y)$. For every bounded and convex $F \subset X$ and $n \in \mathbb{N}_0$, we have*

$$e_n^{\text{det-non-mb}}(S,F) \;\leq\; 8\, e_n^{\text{det-non}}(S,F).$$

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

Table 2. *Maximal gain in the rate of convergence between different classes of algorithms using linear information.*

Gain from to / for	$\mathcal{A}_n^{\text{det-non}}$		$\mathcal{A}_n^{\text{ran-non}}$	$\mathcal{A}_n^{\text{det}}$
	$\mathcal{A}_n^{\text{det}}$	$\mathcal{A}_n^{\text{ran-non}}$	$\mathcal{A}_n^{\text{ran}}$	
F convex+symmetric	0	$\left[\frac{1}{2}, 1\right]$	1	1
F convex	$\left[\frac{1}{2}, \frac{3}{2}\right]$	$\left[\frac{1}{2}, \frac{3}{2}\right]$	$\left[1, \frac{3}{2}\right]$	$\left[1, \frac{3}{2}\right]$

Lemma 6.2 is proven in [41, Thm. 11(v)] for the case that F is the unit ball of the space X. (In fact, it is shown that continuous algorithms are almost optimal.) We show below how it can be transferred to general convex classes. It is open whether the factor 8 can be removed and to what extent measureable algorithms are as good as nonmeasurable algorithms also in other settings, see [51, Section 4.3.3] and [50].

Proof of Lemma 6.2. Without loss of generality we assume that $0 \in F$; the error numbers do not change if we shift F. Then we have $F \subset F - F$. The class $F - F$ is convex, bounded and symmetric, and hence the unit ball of a norm on X. By [41, Thm. 11(v)], it holds that

$$e_n^{\text{det-non-mb}}(S, F) \;\leq\; e_n^{\text{det-non-mb}}(S, F - F) \;\leq\; 2\, e_n^{\text{det-non}}(S, F - F).$$

On the other hand, [51, Lemma 4.3] and Proposition 4.1 give

$$e_n^{\text{det-non}}(S, F - F) \;\leq\; 2\, c_n(S, F - F) \;=\; 4\, c_n(S, F) \;\leq\; 4\, e_n^{\text{det-non}}(S, F).$$ □

We can now prove Corollary 6.1.

Proof of Corollary 6.1. All the minimal errors are bounded from below by $\frac{1}{2} \cdot h_{2n-1}(S, F)$. For $\mathcal{A}_n^{\text{ran}}$ and $\mathcal{A}_n^{\text{ran-non}}$, this follows from Proposition 4.4. For $\mathcal{A}_n^{\text{det}}$ and $\mathcal{A}_n^{\text{det-non}}$, it follows from [48, Prop. 1] and Proposition 3.2. On the other hand, all the minimal errors are bounded from above by $16 \cdot c_n(S, F)$. For $\mathcal{A}_n^{\text{det}}$ and $\mathcal{A}_n^{\text{det-non}}$, this follows from Proposition 4.1. Lemma 6.2 implies that the upper bound also holds for $A_n^{\text{det-non-mb}}$ and thus for $\mathcal{A}_n^{\text{ran}}$ and $\mathcal{A}_n^{\text{ran-non}}$. Hence, the statement follows from Theorem 3.3. □

We summarize the state of the art for the maximal gain between the different classes of algorithms, see Table 2. For this, let us define the "maximal gain function"

$$\operatorname{gain}(*, \square, \triangle) \;:=\; \sup_{\substack{S \in \mathcal{L}(X,Y) \\ F \subset X \text{ is } \triangle}} \Big(\operatorname{rate}\Big(e_n^{\square}(S, F)\Big) - \operatorname{rate}\Big(e_n^{*}(S, F)\Big)\Big),$$

where the rate function is defined in (5.1), $*, \square \in$ {det, det-non, ran, ran-non} and $\triangle \in$ {convex, convex+symmetric}.

The upper bounds in Table 2 are given by Corollary 6.1 and Proposition 4.1. We now turn to the lower bounds on the (individual) gain of adaption and randomization. For this, we collect specific examples:

- F convex+symmetric:
  1. *Power of adaption, deterministic* ($\mathcal{A}_n^{\text{det-non}} \to \mathcal{A}_n^{\text{det}}$):
     There is no gain in the rate of convergence. As stated in Proposition 4.1, we have $e_n^{\text{det-non}}(S, F) \leq 2 \cdot e_n^{\text{det}}(S, F)$ for any $S \in \mathcal{L}$, see for example [9, 49, 51, 68]. An example where adaptive algorithms are slightly better can be found in [28].
  2. *Power of randomization, nonadaptive* ($\mathcal{A}_n^{\text{det-non}} \to \mathcal{A}_n^{\text{ran-non}}$):
     Randomization of nonadaptive algorithms can yield a gain of 1/2 for certain Sobolev embeddings. The simplest case is from $W_2^k([0, 1])$ into $L_\infty([0, 1])$, where the optimal rate with deterministic

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

algorithms is $n^{-k+1/2}$ for $k > 1/2$. This is a classical result of [62]. Using nonadaptive randomized algorithms, one can get the upper bound $n^{-k}\log n$, see [42] and [6, 14, 21] for further results and extensions.

3. *Power of adaption, randomized* ($\mathcal{A}_n^{\text{ran-non}} \to \mathcal{A}_n^{\text{ran}}$):
   The paper [38] shows that one may gain a factor $(n/\log n)^{1/2}$ for the embedding $S\colon \ell_1^m \to \ell_2^m$ and suitable (large) m and in [39] it is proved that one may gain, up to logarithmic terms, a factor of polynomial order n for the embedding $S\colon \ell_1^m \to \ell_\infty^m$ with appropriate m. Note that this example (or more precisely, an infinite-dimensional version of it) shows the optimality of Corollary 5.3 and the factor n in Theorem 1.2. We do not know if a gain of 1 can also occur in the transitions $\mathcal{A}_n^{\text{det-non}} \to \mathcal{A}_n^{\text{ran-non}}$.
4. *Power of randomization, adaptive* ($\mathcal{A}_n^{\text{det}} \to \mathcal{A}_n^{\text{ran}}$):
   If we employ Example (1), as well as $\mathcal{A}_n^{\text{det-non-mb}} \subset \mathcal{A}_n^{\text{ran-non}}$ and Lemma 6.2, we can take the examples from (3), to obtain the same gain.

- F convex:

5. *Power of adaption, deterministic* ($\mathcal{A}_n^{\text{det-non}} \to \mathcal{A}_n^{\text{det}}$):
   Consider $S = \text{id} \in \mathcal{L}(\ell_\infty, \ell_\infty)$, that is, approximation in ℓ_∞, for inputs from
   $$F = \{x \in \ell_\infty \mid x_i \geq 0, \sum x_i \leq 1, x_k \geq x_{2k}, x_k \geq x_{2k+1}\}.$$
   Then one can prove a lower bound $c(\sqrt{n}\log n)^{-1}$ for nonadaptive algorithms while a simple adaptive algorithm (using "function values," that is, values of coordinates of x) gives the upper bound $(n+3)^{-1}$; see [47] for details. This shows a gain of 1/2.In the case of standard information, that is, function values, see also Section 6.3, there are more extreme examples, where adaption yields a gain up to the order n, see again [47].
6. *Remaining Cases*:
   Clearly, the examples given in the symmetric case also apply here. This gives the remaining lower bounds of Table 2. We do not know any example of nonsymmetric F, where the gain of adaptive over nonadaptive randomized algorithms, or of randomized over deterministic algorithms (adaptive or not), is larger than the corresponding gain in the symmetric case.

Let us highlight a few open problems indicated by Table 2 that we assume to be of particular interest. In particular, note that it is still possible that the maximal gain of adaption and randomization is 1 also for nonsymmetric convex sets.

Open problems.

1. Bounds for individual n: Verify whether $e_{2n}^{\text{det-non}}(S,F) \leq C\, n\, e_n^{\text{ran}}(S,F)$ for some $C > 0$ and all convex and symmetric F. See also [53, Thm. 8.6]. Does a similar bound hold even for all convex classes F?
2. Power of adaption: Is there some $S \in \mathcal{L}$ and convex F such that
   $$e_{2n}^{\text{det}}(S,F) \leq C\, n^{-1}\, e_n^{\text{det-non}}(S,F)$$
   for all $n \in \mathbb{N}$?
3. Power of randomization: Is there some $S \in \mathcal{L}$ and convex F such that
   $$e_{2n}^{\text{ran-non}}(S,F) \leq C\, n^{-1}\, e_n^{\text{det-non}}(S,F)$$
   for all $n \in \mathbb{N}$?
4. Power of randomization for symmetric sets: Is there some $S \in \mathcal{L}$ and convex and symmetric F such that
   $$e_{2n}^{\text{ran-non}}(S,F) \leq C\, n^{-\alpha}\, e_n^{\text{det-non}}(S,F)$$
   for all $n \in \mathbb{N}$ and some $\alpha > \frac{1}{2}$?

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

6.2. Other widths

Our main emphasis was to compare different classes of algorithms. However, the study of widths is clearly not only of interest in IBC. These notions are used in many areas, including approximation theory, geometry, and the theory of Banach spaces.

We indicate how our results apply here, and give further references.

6.2.1. Kolmogorov widths

Possibly most prominent among the widths are the *Kolmogorov widths*

$$d_n(S,F) := \inf_{\substack{M\subset Y\\ \dim(M)\le n}} \sup_{f\in F} \inf_{g\in M} \|S(f)-g\|,$$

which describe how well $S(f)$ for $f \in F$ can be approximated by elements from an affine linear subspace. Such a *best approximation* can generally not be found by an algorithm, and so, d_n is usually not a suitable benchmark for algorithms. Still, it is a natural "geometric" quantity.

If $F = B_X$, then it is known that the Kolmogorov numbers d_n are *dual* to the Gelfand numbers, and that the Hilbert numbers are self-dual, see [4, 53, 55]. Similar statements might hold for general F, and this could be used to extend our results to d_n. However, one may also repeat the proofs of our results almost verbatim for d_n. By this, we obtain for $S \in \mathcal{L}(X,Y)$ and convex $F \subset X$ that

$$d_n(S,F) \le (n+1)^\alpha \left(\prod_{k=0}^{n} h_k(S,F)\right)^{1/(n+1)} \tag{6.1}$$

with $\alpha = 1$ if F is symmetric and $\alpha = 3/2$ otherwise. The symmetric case is proven in [70]; the modifications for the nonsymmetric case are analogous. We omit the details.

Of course, the same bound holds with the larger b_n in place of the smaller h_n. The ratio of Kolmogorov and Bernstein widths was studied at least since the paper [43] from Mityagin and Henkin (1963). They proved that $d_n(S,F) \le (n+1)^2\, b_n(S,F)$ for convex and symmetric F, and conjectured that one has indeed $d_n(S,F) \le (n+1)\, b_n(S,F)$. See also [48] for the nonsymmetric case. The above considerations show that this old conjecture is true, at least for regular sequences and up to constants. Note again that this was known, see [56].

Similar problems appear in geometry, where often different notions are used, such as successive radii, or inner and outer radii. In this context, their ratio was mainly considered for sets in Hilbert spaces, see [18, 19, 52, 59], where one can find further references. Results for general norms can be found, for example, in [19, Theorem 5.1]. These bounds are improved by the inequalities above.

Remark 6.3. For S being the identity on a Hilbert space, it is conjectured that the regular simplex provides the largest gap in the nonsymmetric case and the regular cube or the regular cross-polytope provides the largest gap in the symmetric case. The Kolmogorov and Bernstein widths of these sets are completely known [5, 60], but the proven general bounds are slightly weaker.

6.2.2. Linear widths

If we replace the requirement that the approximation space is linear by the requirement that the approximation procedure is linear, we end up with the *approximation numbers of S on F*, or *linear widths of F with respect to S*, that is,

$$a_n(S,F) := \inf_{\substack{L_1,\dots,L_n\in X'\\ \varphi_0,\dots,\varphi_n\in Y}} \sup_{f\in F} \left\| S(f) - \varphi_0 - \sum_{i=1}^{n} L_i(f)\,\varphi_i \right\|.$$

This corresponds to the minimal worst-case error of affine algorithms that use at most n pieces of linear information.

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

If F is the unit ball of X, then it is known from [53, Thm. 8.4] (based on [27]) that

$$a_n(S, B_X) \le (1 + \sqrt{n})\, c_n(S, B_X). \tag{6.2}$$

Note that $a_n(S, B_X)$ and $c_n(S, B_X)$ are equal if X is a Hilbert space or if Y has the metric extension property, see [53] or [9, 41] for extensions. Together with Theorem 3.3, inequality (6.2) implies

$$a_n(S, B_X) \le 2\,(n+1)^{3/2} \left(\prod_{k=0}^{n} h_k(S, B_X) \right)^{1/(n+1)}.$$

The exponent $3/2$ can be replaced by 1 in the aforementioned cases or in the case that Y is a Hilbert space since then we have the identity $a_n(S, B_X) = d_n(S, B_X)$. Again, for this symmetric case, the result is essentially known, see [56, 6.2.3.14], and we only remove an oversampling constant compared to the known estimate. It is a major open problem, whether the exponent $3/2$ can be reduced to 1, see [57, Open Problem 5].

This problem is of particular interest in the theory of s-numbers, where F is assumed to be the unit ball of X. Note that the approximation numbers form the largest scale of s-numbers while the Hilbert numbers are the smallest scale of s-numbers, see [55, Ch. 2].

6.2.3. Nonlinear widths

In connection with nonlinear approximation, also several types of *nonlinear widths* appear in the literature, see, for example, [7, 10, 11, 61, 63, 64, 73]. Let us introduce two of them to illustrate the relation to our setting. First, the *manifold widths* of $F \subset X$ with respect to $S \in \mathcal{L}(X, Y)$ are defined by

$$\delta_n(S, F) := \inf_{\substack{N \in C(X,\mathbb{R}^n) \\ \varphi \in C(\mathbb{R}^n, Y)}} \sup_{f \in F} \big\| S(f) - \varphi(N(f)) \big\|,$$

where $C(X, Y)$ denotes the class of continuous mappings from X to Y. Moreover, the *continuous co-widths* of $F \subset X$ w.r.t. $S \in \mathcal{L}(X, Y)$ are

$$\widetilde{c}_n(S, F) := \inf_{N \in C(X,\mathbb{R}^n)} \sup_{\substack{f, g \in F: \\ N(f) = N(g)}} \frac{1}{2} \big\| S(f) - S(g) \big\|.$$

These numbers correspond (up to a factor 2) to minimal errors for approximating S over F based on n nonadaptive continuous measurements. Comparing these definitions to minimal errors and Gelfand numbers, we see that this approach is more general in the sense that the information mapping (or parameter selection map) N is not built from linear functionals, but less general in the sense that a discontinuous adaptive choice of the one-dimensional measurements (like, e.g., a bisection method) is not allowed. These quantities seem to appear in the literature only in the special case of $S \in \mathcal{L}(X, X)$ being the identity. We naturally extend the definitions, but only comment on this special case in the following.

First, note that another important (and very early) concept are the Aleksandrov widths, see [1, 11], which replace $\mathbb{R}^n$ by more general n-dimensional *complexes*. However, it is shown in [11] that all these quantities are equivalent up to constants and oversampling.

Second, it is clear from the definitions that these widths are smaller than $e_n^{\text{det-non}}$ and c_n, respectively. Moreover, it is shown in [10] that they are lower-bounded by the Bernstein widths b_n. So, we have everything we need to apply our technique: The upper bounds in Theorem 3.3 and in Table 1 are also applicable to the maximal gain in the rate of convergence when passing from linear to arbitrary continuous measurement maps $N\colon F \to \mathbb{R}^n$, at least for S being the identity.

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

In fact, combining Theorem 3.3, in the form of Corollary 5.2, with [10, Thm. 3.1] (and noting that they use the notation "d_n" for "δ_n"), we obtain, for example, for all convex $F \subset X$ that

$$c_{2n}(\mathrm{id}_X, F) \ \le\ C_\alpha\, n^{-\alpha+3/2} \cdot \sup_{k<n}\, (k+1)^\alpha\, \delta_k(\mathrm{id}_X, F), \tag{6.3}$$

where $C_\alpha \le 16^{\alpha+1}$. Again, the exponent $3/2$ can be replaced by 1 for symmetric F and, taking Section 6.2.1 into account, this bound also holds with d_{2n} in place of c_{2n}.

For a very different (and surprising for us) result on adaptive continuous measurements that shows an exponential speed-up, see [30].

6.3. Sampling numbers and other classes of information

As already indicated in the introduction, our proof technique can also be employed for sampling recovery in the uniform norm. In fact, our upper bound from Theorem 1.1 also holds if the class of deterministic, nonadaptive algorithms is further restricted to those using only certain restricted information.

For this, we consider a set of linear functionals $\Lambda \subset X'$, which we call the *admissible information*. The n-th minimal worst-case error for approximating S over F with information from Λ is defined by

$$e_n^{\text{det-non}}(S, F, \Lambda) \ :=\ \inf_{\substack{L_1,\dots,L_n\in\Lambda\\ \varphi\colon\ \mathbb{R}^n\to Y}}\ \sup_{f\in F} \|S(f) - \varphi(L_1(f),\dots,L_n(f))\|.$$

Moreover, we call $\mathcal{N} \subset X'$ a *norming set* of X if

$$\|f\| = \sup\{\,|\lambda(f)|\colon\ \lambda \in \mathcal{N}\} \qquad \text{for all} \quad f \in X.$$

With this, we obtain the following generalization of Theorem 1.1.

Theorem 6.4. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X,Y)$. Moreover, let $\Lambda_S \subset \Lambda \subset X'$, where Λ_S is of the form $\Lambda_S = \{\lambda \circ S\colon\ \lambda \in \mathcal{N}\}$ and $\mathcal{N} \subset Y'$ is a norming set of Y. Then, for every convex $F \subset X$ and $n \in \mathbb{N}$, we have*

$$\begin{aligned} e_{2n-1}^{\text{det-non}}(S, F, \Lambda) \ &\le\ \left(\prod_{k<2n} e_k^{\text{det-non}}(S, F, \Lambda)\right)^{1/(2n)} \\ &\le\ 12\, n^{3/2} \left(\prod_{k<n} e_k^{\text{ran}}(S, F)\right)^{1/n}. \end{aligned}$$

The corresponding improvements from Theorem 1.1 apply if F is additionally symmetric or the unit ball of a Hilbert space X.

Proof of Theorem 6.4. First note that the bound

$$e_n^{\text{det-non}}(S, F, \Lambda) \ \le\ 2\, c_n(S, F, \Lambda) \tag{6.4}$$

from Proposition 4.1 also holds for the generalized Gelfand numbers

$$c_n(S, F, \Lambda) \ :=\ \inf_{L_1,\dots,L_n\in\Lambda}\ \sup_{\substack{f,g\in F:\\ L_k(f)=L_k(g)}} \frac{1}{2}\left\|S(f) - S(g)\right\|$$

see [51, Sec. 4.1]. Hence it suffices to bound the modified Gelfand numbers in terms of the Hilbert numbers as in Theorem 3.3.

We proceed as in the proof of Theorem 3.3. By the definition of a norming set $\mathcal{N} \subset Y'$, we can choose the functionals λ_k in (3.2) from $\mathcal{N}$, arguing with a slightly smaller ε in the previous inequality.

Hence, the functionals L_k defining M_k are from Λ_S and we can choose f_k, g_k such that (3.1) holds with the modified Gelfand widths. This shows that the assertion in the very beginning of the proof of Theorem 3.3 holds with $c_k(S, F)$ replaced by $c_k(S, F, \Lambda)$.

The same replacement is possible for the corresponding assertion in the case that F is the unit ball of a Hilbert space. We can therefore copy the rest of the proof of Theorem 3.3 with $c_k(S, F)$ replaced by $c_k(S, F, \Lambda)$. □

Recall that the minimal errors on the right-hand side in Theorem 6.4 are for the (much bigger) class of randomized, adaptive algorithms that have access to arbitrary linear functionals, see Section 2. We do not consider the case that Y is a Hilbert space since we believe that norming sets in Hilbert spaces are too large to yield an interesting generalization of Theorem 1.1. Let us also note that our proof of Theorem 1.1 only employs information of the form Λ_S. As such, Theorem 1.1 cannot catch the optimal behavior of $e_n^{\text{det-non}}(S, F)$ if the latter decay faster than $e_n^{\text{det-non}}(S, F, \Lambda_S)$.

We discuss a few examples:

1. *Linear information*:
   The case studied in Theorem 1.1 corresponds to $\mathcal{N} = B_{Y'}$ and $\Lambda = X'$ which satisfy the assumptions of the theorem. The class $B_{Y'}$ is a norming set by the Hahn-Banach theorem.
2. *Uniform approximation (and proof of Theorem 1.2)*:
   We consider $X = Y = B(D)$ and $S = \mathrm{APP}_\infty$, that is, the identity on $B(D)$, and observe that $\Lambda^{\text{std}} := \{\delta_x : x \in D\}$ with the Dirac functionals $\delta_x(f) = f(x) = Sf(x)$ is a norming set of $B(D)$. Since $g_n(S, F) = e_n^{\text{det-non}}(S, F, \Lambda^{\text{std}})$, see [9] or [51, Thm. 4.8], we obtain Theorem 1.2 from Theorem 6.4. A special case is the space of bounded sequences $\ell_\infty = B(\mathbb{N})$, where Λ^{std} consists of evaluations of coordinates. Note that the factor 12 from Theorem 6.4 can be replaced with 6 in Theorem 1.2 since the factor 2 in (6.4) can be removed in this case, see again [51, Sec. 4.1]. A related bound on the sampling numbers in $B(D)$, without the geometric mean and in terms of *entropy numbers*, was recently obtained in [71].
3. *C^k-approximation*:
   We can also apply Theorem 6.4 to function recovery in $C^k(D)$, the space of k-times continuously differentiable functions on a compact domain $D \subset \mathbb{R}^d$. That is, we consider the identity S on the space $X = Y = C^k(D)$, which we equip with the norm

$$\|f\|_{C^k(D)} := \max_{|\alpha| \le k} \max_{x \in D} \left| \frac{\partial^{|\alpha|} f(x)}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}} \right|.$$

   Theorem 6.4 applies for the class Λ of point evaluations of derivatives up to order k, that is, for

$$\Lambda = \left\{ \delta_x \circ \frac{\partial^{|\alpha|}}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}} \,\middle|\, x \in D,\ |\alpha| \le k \right\},$$

   which is a norming set on $C^k(D)$.

We discuss two implications of Theorem 1.2.

Firstly, Theorem 1.2 contributes to the question on the power of adaption and randomization if only function values are available. Namely, considering algorithms for uniform approximation that only use standard information, Theorem 1.2 gives that adaption and randomization cannot lead to a speed-up larger than one, that is,

$$\mathrm{rate}\Big(e_n^{\text{det-non}}(\mathrm{APP}_\infty, F, \Lambda^{\text{std}})\Big) \;\ge\; \mathrm{rate}\Big(e_n^{\text{ran}}(\mathrm{APP}_\infty, F, \Lambda^{\text{std}})\Big) - 1$$

for any convex and symmetric $F \subset B(D)$. This is in analogy to our result for linear information, see Corollary 5.3.

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press

Secondly, Theorem 1.2 also contributes to the question on the power of standard information compared to arbitrary linear information. If we consider randomized algorithms for uniform approximation, Theorem 1.2 gives that a restriction to standard information causes a loss in the rate of convergence of no more than one, that is,

$$\operatorname{rate}\Big({e_n}^{\mathrm{ran}}(\mathrm{APP}_\infty, F, \Lambda^{\mathrm{std}})\Big) \;\geq\; \operatorname{rate}\Big({e_n}^{\mathrm{ran}}(\mathrm{APP}_\infty, F)\Big) \;-\; 1$$

for any convex and symmetric $F \subset B(D)$. The analogous result for deterministic algorithms has been proven in [45]. This has recently been improved to

$$\operatorname{rate}\Big({e_n}^{\mathrm{det}}(\mathrm{APP}_\infty, F, \Lambda^{\mathrm{std}})\Big) \;\geq\; \operatorname{rate}\Big({e_n}^{\mathrm{det}}(\mathrm{APP}_\infty, F)\Big) \;-\; 1/2, \tag{6.5}$$

see [32]. Note that we even have equality of the rates if F is the unit ball of a certain kind of reproducing kernel Hilbert space, see [16, 31].

It is an interesting open problem whether (6.5) also holds in the randomized setting, and to what extent the results above hold for more general problems $S \in \mathcal{L}(X, Y)$.

Remark 6.5 (Sampling numbers in L_2). Another problem where several new bounds have been obtained recently is the case that $S = \mathrm{APP}_2$, that is, the embedding of X into the space $Y = L_2$. In this case, there are various upper bounds for the error of nonadaptive algorithms based on function values in terms of the Kolmogorov numbers $d_n(S, F)$, see [12, 34, 35, 44, 65, 69] for deterministic and [8, 29, 72] for randomized algorithms. On the other hand, the bound (6.1) and Lemma 4.3 give an upper bound on $d_n(S, F)$ in terms of the error of adaptive randomized algorithms. Hence, we may derive several bounds on the maximal gain of adaption and/or randomization for the problem of sampling recovery in L_2.

We only mention the special case that F is the unit ball of a reproducing kernel Hilbert space $X = H$ with finite trace. Using that $\operatorname{rate}(e_n^{\text{det-non}}(\mathrm{APP}_2, B_H, \Lambda^{\mathrm{std}})) = \operatorname{rate}(c_n(\mathrm{APP}_2, B_H))$ from [34, Corollary 1] together with Lemma 4.3 and Proposition 3.2, we obtain that

$$\operatorname{rate}\Big(e_n^{\text{det-non}}(\mathrm{APP}_2, B_H, \Lambda^{\mathrm{std}})\Big) \;=\; \operatorname{rate}\Big(e_n^{\mathrm{ran}}(\mathrm{APP}_2, B_H)\Big).$$

That is, linear sampling algorithms are optimal (in the sense of order) among arbitrary adaptive, randomized algorithms that may use general linear information.

Remark 6.6 (Exponential decay). For many classes F of smooth functions, the n-th minimal error has a super-polynomial decay and Theorem 1.2 together with (3.4) implies a bound of the form

$$g_{cn}^{\mathrm{lin}}(\mathrm{APP}_\infty, F) \;\leq\; {e_n}^{\mathrm{ran}}(\mathrm{APP}_\infty, F)$$

for all $n \geq n_0$, where $n_0 \in \mathbb{N}$ and $c \geq 1$ are (relatively small) constants. One such example is given by reproducing kernel Hilbert spaces with a Gaussian kernel, see, for example, [17, Thm 1.1]. This means that, for all such examples, there is no need for sophisticated algorithms that use randomization, adaption or general linear information, at least from the viewpoint of information complexity. In comparison to deterministic and nonadaptive algorithms that only use function evaluations, at most a factor c can be gained. A similar result for L_2-approximation can be obtained along the lines of Remark 6.5, see also [33].

Acknowledgments. We gratefully acknowledge the support of the Leibniz Center for Informatics, where several discussions on this research were held during the Dagstuhl Seminar *Algorithms and Complexity for Continuous Problems* (Seminar ID 23351).

Competing interest. The authors have no competing interests to declare.

Financial support. This research was funded in whole or in part by the Austrian Science Fund (FWF) grant M 3212-N. For open access purposes, the authors have applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission. MU is supported by the Austrian Federal Ministry of Education, Science and Research via the Austrian Research Promotion Agency (FFG) through the project FO999921407 (HDcode) funded by the European Union via NextGenerationEU.

References

[1] P. Aleksandrov, *Combinatorial Topology*, Vol. 1 (Graylock Press, Rochester, NY, 1956).

[2] N. S. Bakhvalov, 'On the optimality of linear methods for operator approximation in convex classes of functions', *USSR Comput. Maths. Math. Phys.* **11** (1971), 244–249.

[3] N. S. Bakhvalov, 'On the approximate calculation of multiple integrals', *J. Complexity* **31** (2015), 502–516.

[4] W. Bauhardt, 'Hilbert-Zahlen von Operatoren in Banachräumen', *Math. Nachr.* **79** (1977), 181–187.

[5] R. Brandenberg, 'Radii of regular polytopes', *Discrete Comput. Geom.* **33** (2005), 43–55.

[6] G. Byrenheid, R. J. Kunsch and V. K. Nguyen, 'Monte Carlo methods for uniform approximation on periodic Sobolev spaces with mixed smoothness', *J. Complexity* **46** (2018), 90–102.

[7] A. Cohen, R. DeVore, G. Petrova and P. Wojtaszczyk, 'Optimal stable nonlinear approximation', *Found. Comput. Math.* **22** (2022), 607–648.

[8] A. Cohen and M. Dolbeault, 'Optimal pointwise sampling for L_2 approximation', *J. Complexity* **68** (2022), 101602.

[9] J. Creutzig and P. Wojtaszczyk, 'Linear vs. nonlinear algorithms for linear problems', *J. Complexity* **20** (2004), 807–820.

[10] R. A. DeVore, R. Howard and C. Micchelli, 'Optimal nonlinear approximation', *Manuscripta Math.* **63** (1989), 469–478.

[11] R.A. DeVore, G. Kyriazis, D. Leviatan and V. M. Tikhomirov, 'Wavelet compression and nonlinear n-widths', *Adv. Comput. Math.* **1** (1993), 197–214.

[12] M. Dolbeault, D. Krieg and M. Ullrich, 'A sharp upper bound for sampling numbers in L_2', *Appl. Comput. Harmon. Anal.* **63** (2023), 113–134.

[13] D. E. Edmunds and J. Lang, 'Gelfand numbers and widths', *J. Approx. Theory* **166** (2013), 78–84.

[14] G. Fang and L. Duan, 'The complexity of function approximation on Sobolev spaces with bounded mixed derivative by linear Monte Carlo methods', *J. Complexity* **24** (2008), 398–409.

[15] S. Gal and C. A. Micchelli, 'Optimal sequential and non-sequential procedures for evaluating a functional', *Appl. Anal.* **10** (1980), 105–120.

[16] J. Geng and H. Wang, 'On the power of standard information for tractability for L_∞ approximation of periodic functions in the worst case setting', *J. Complexity* **80** (2024), 101790.

[17] T. Karvonen and Y. Suzuki, 'Approximation in Hilbert spaces of the Gaussian and related analytic kernels', Preprint (2022), arXiv:2209.12473.

[18] B. González Merino, 'On the ratio between successive radii of a symmetric convex body', *Math. Ineq. Appl.* **16** (2013), 569–576.

[19] B. González Merino, 'Improving bounds for the Perel'man–Pukhov quotient for inner and outer radii', *J. Convex Anal.* **24** (2017), 1099–1116.

[20] S. Heinrich, 'On the relation between linear n-widths and approximation numbers', *J. Approx. Theory* **58** (1989), 315–333.

[21] S. Heinrich, 'Lower bounds for the complexity of Monte Carlo function approximation', *J. Complexity* **8** (1992), 277–300.

[22] S. Heinrich, 'Randomized complexity of parametric integration and the role of adaption I. *Finite dimensional case', J. Complexity* **81** (2024), 101821.

[23] S. Heinrich, 'Randomized complexity of parametric integration and the role of adaption II. Sobolev spaces', *J. Complexity* **82** (2024), 101823.

[24] S. Heinrich, 'Randomized complexity of vector-valued approximation', in A. Hinrichs, P. Kritzer and F. Pillichshammer (eds), *Monte Carlo and Quasi-Monte Carlo Methods. MCQMC 2022*, Springer Proc. Math. Statist., Vol. 460 (Springer, Cham, 2022).

[25] S. Heinrich, 'Randomized complexity of mean computation and the adaption problem', *J. Complexity* **85** (2024), 101872.

[26] R. S. Ismagilov, 'Diameters of sets in normed linear spaces and approximation of functions by trigonometric polynomials', *Russ. Math. Surveys* **29**(3) (1974), 169–186.

[27] M. Kadets and S. Snobar, 'Certain functionals on the Minkowski compactum', *Math. Notes* **10** (1971), 694–696.

[28] M. A. Kon and E. Novak, 'The adaption problem for approximating linear operators', *Bull. Amer. Math. Soc.* **23** (1990), 159–165.

[29] D. Krieg, 'Optimal Monte Carlo methods for L_2-approximation', *Constr. Approx.* **49** (2019), 385–403.

[30] D. Krieg, E. Novak and M. Ullrich, 'How many continuous measurements are needed to learn a vector?', Preprint (2025), arXiv:2412.06468.

[31] D. Krieg, K. Pozharska, M. Ullrich and T. Ullrich, 'Sampling recovery in L_2 and other norms', to appear in *Math. Comp.*, Preprint (2025), arXiv:2305.07539.

[32] D. Krieg, K. Pozharska, M. Ullrich and T. Ullrich, 'Sampling projections in the uniform norm', *J. Math. Anal. Appl.* **553**(2) (2026), 129873.

[33] D. Krieg, P. Siedlecki, M. Ullrich and H. Woźniakowski, 'Exponential tractability of L_2-approximation with function values', *Adv. Comput. Math.* **49** (2023), 18.

[34] D. Krieg and M. Ullrich, 'Function values are enough for L_2-approximation', *Found. Comput. Math.* **21**(4) (2021), 1141–1151.

[35] D. Krieg and M. Ullrich, 'Function values are enough for L_2-approximation: Part II', *J. Complexity* **66** (2021), 101569.

[36] R. J. Kunsch, 'Bernstein Numbers and Lower Bounds for the Monte Carlo Error', in R. Cools and D. Nuyens (eds), *Monte Carlo and Quasi-Monte Carlo Methods*, Springer Proc. Math. Statist., Vol. 163 (Springer, Cham, 2016).

[37] R. J. Kunsch, *High-dimensional Function Approximation: Breaking the Curse with Monte Carlo Methods*, Ph. D. dissertation, Preprint (2017), arXiv:1704.08213.

[38] R. J. Kunsch, E. Novak and M. Wnuk, 'Randomized approximation of summable sequences – adaptive and non-adaptive', *J. Approx. Theory* **304** (2024), 106056.

[39] R. J. Kunsch and M. Wnuk, 'Uniform approximation of vectors using adaptive randomized information', Preprint (2024), arXiv:2408.01098.

[40] G. G. Lorentz, M. v. Golitschek and Yu. Makovoz, *Constructive Approximation, Advanced Problems*, Grundlehren 304 (Springer, Berlin, 1996).

[41] P. Mathé, 's-numbers in Information-Based Complexity', *J. Complexity* **6** (1990), 41–66.

[42] P. Mathé, 'Random approximation of Sobolev embeddings', *J. Complexity* **7** (1991), 261–281.

[43] B. S. Mityagin and G. M. Henkin, 'Inequalities between n-diameters', in *Proc. Seminar on Functional Analysis* 7 (Voronezh, 1963), 97–103.

[44] N. Nagel, M. Schäfer and T. Ullrich, 'A new upper bound for sampling numbers', *Found. Comput. Math.* **22**(2) (2022), 445–468.

[45] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lecture Notes in Mathematics, Vol. 1349 (Springer, Berlin, 1988).

[46] E. Novak, 'Optimal linear randomized methods for linear operators in Hilbert spaces', *J. Complexity* **8**(1) (1992), 22–36.

[47] E. Novak, 'Optimal recovery and n-widths for convex classes of functions', *J. Approx. Theory* **80** (1995), 390–408.

[48] E. Novak, 'The adaption problem for nonsymmetric convex sets', *J. Approx. Theory* **82** (1995), 123–134.

[49] E. Novak, 'On the power of adaption', *J. Complexity* **12** (1996), 199–237.

[50] E. Novak and K. Ritter, 'A stochastic analog to Chebyshev centers and optimal average case algorithms', *J. Complexity* **5** (1989), 60–79.

[51] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems*. *Volume I: Linear Information*, EMS Tracts in Mathematics, Vol. 6 (European Mathematical Society, Zürich, 2008).

[52] G. Ya . Perel'man, 'On the k-radii of a convex body', *Siberian Math. J.* **28** (1987), 665–666.

[53] A. Pietsch, 's-numbers of operators in Banach spaces', *Studia Math.* **51** (1974), 201–223.

[54] A. Pietsch, *Operator Ideals*, North-Holland Mathematical Library, Vol. 20 (Elsevier, Amsterdam, 1980).

[55] A. Pietsch, *Eigenvalues and s-Numbers*, Cambridge Studies in Advanced Mathematics, Vol. 13 (Cambridge University Press, Cambridge, 1987).

[56] A. Pietsch, *History of Banach Spaces and Linear Operators* (Birkhäuser, Boston, MA, 2007).

[57] A. Pietsch, 'Long-standing open problems of Banach space theory: My personal top ten', *Quaest. Math.* **32**(3) (2009), 321–337.

[58] A. Pinkus, *n-Widths in Approximation Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (Springer, Berlin, 1985).

[59] S. V. Pukhov, 'Inequalities for the Kolmogorov and Bernstein widths in Hilbert space', *Math. Notes* **25** (1979), 320–326. [Translated from Mat. Zametki 25(4) (1979), 619–628.]

[60] S. V. Pukhov, 'Kolmogorov diameters of a regular simplex', *Moscow Univ. Math. Bull.* **35** (1980), 38–41.

[61] J. W. Siegel, 'Sharp lower bounds on the manifold widths of Sobolev and Besov spaces', *J. Complexity* **85** (2024), 101884.

[62] M. E. Solomjak and V. M. Tichomirov, 'Some geometric characteristics of the embedding map from W_p^α to C', *Izv. Vyssh. Uchebn. Zaved. Mat.* **10** (1967), 76–82.

[63] M. I. Stesin, 'On Aleksandrov diameters of balls', *Dokl. Akad. Nauk SSSR* **217**(1) (1974), 31–33.

[64] M. I. Stesin, 'Aleksandrov diameters of finite-dimensional sets and classes of smooth functions', *Dokl. Akad. Nauk SSSR* **220**(6) (1975), 1278–1281.

[65] V. N. Temlyakov, 'On optimal recovery in L_2', *J. Complexity* **65** (2021), 101545.

[66] V. M. Tikhomirov, 'Diameters of sets in function spaces and the theory of best approximations', *Russ. Math. Surveys* **15**(3) (1960), 75–111.

[67] J. Traub and H. Woźniakowski, *A General Theory of Optimal Algorithms* (Academic Press, New York, 1980).

[68] J. Traub, G. Wasilkowski and H. Woźniakowski, *Information-Based Complexity* (Academic Press, Boston, MA, 1988).

[69] M. Ullrich, 'On the worst-case error of least squares algorithms for L_2-approximation with high probability', *J. Complexity* **60** (2020), 1–14.

[70] M. Ullrich, 'On inequalities between s-numbers', *Adv. Oper. Theory* **9** (2024), 82.

[71] M. Ullrich, 'Sampling and entropy numbers in the uniform norm', to appear in J. Complexity.

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

[73] D. Yarotsky, 'Error bounds for approximations with deep ReLU networks', *Neural Networks* **94** (2017), 103–114.

https://doi.org/10.1017/fms.2025.10101 Published online by Cambridge University Press