Linear vs. Nonlinear Algorithms for linear problems

Jakob Creutzig*, P. Wojtaszczyk†

Institute of Mathematics, Informatics and Mechanics
Warsaw University
ul Banacha 3, 02–097 Warszawa
email: jakob@creutzig.de, pwojtaszczyk@mimuw.edu.pl

June 16, 2003

Abstract

We compare linear and non–linear approximations for linear problems. Let X be a linear space and Y a normed space. Let $S : X \to Y$ and $N : X \to \mathbb{R}^n$ be linear mappings and consider the worst–case setting over some balanced convex set $F \subseteq X$. We compare the minimal error $\mathrm{err}^{\mathrm{lin}}(S, N, F)$ achievable by linear algorithms processing N with the minimal error $\mathrm{err}(S, N, F)$ achievable by arbitrary algorithms using N. For *bounded* linear problems, P. Mathé showed that

$$\inf_N \mathrm{err}^{\mathrm{lin}}(S, N, F) \ \le \ (1 + n^{1/2}) \cdot \inf_N \mathrm{err}(S, N, F) \ ,$$

where the infimum is taken over all bounded linear mappings $N : X \to \mathbb{R}^n$. We generalize this result as follows: If the target space Y is complete, then for any linear S,N we have

$$\mathrm{err}^{\mathrm{lin}}(S, N, F) \ \le (1 + n^{1/2}) \cdot \mathrm{err}(S, N, F) \ .$$

This and some similar results can easily be derived from a general relation of this problem to extension properties of normed spaces, and the manifold and powerful results available for this problem. This allows a unified treatment of the above estimate together with the results of Smolyak and Packel, who showed that linear algorithms are optimal for some Y. The results are also partially extended to noisy information with uniformly bounded noise.

1 Linear problems and algorithms

We study the approximation of linear problems. Let X be a linear space and Y a normed space, and let $S : X \to Y$ be a linear operator. Given a balanced and

*Research supported by DFG research grant CR-142/1-1

†Research partially supported by KBN grant 5P03A 03620 located at the Institute of Mathematics of the Polish Academy of Sciences

convex subset $F \subseteq X$, we measure errors in the worst case setting over F, i.e., for an approximation $\tilde{S} : X \to Y$ we set

$$\operatorname{err}(\tilde{S}, F) \ := \ \sup_{x \in F} \|S(x) - \tilde{S}(x)\|_Y \ .$$

This idea is essentially equivalent to study sets F induced by a restriction operator (see e.g. [15, NR 4.5.1:1]). That is, we may assume that there is a linear mapping $T : X \to Z$ into some normed space Z such that $F = \{x \in X : \|T(x)\|_Z \leq 1\}$. Consequently, we assume this and write $\operatorname{err}(\tilde{S}, T) = \operatorname{err}(\tilde{S}, F)$.

The concept of information–based complexity restricts the allowed approximation methods $\tilde{S}$ by the amount of partial information about $x \in X$ used by the algorithm. We are first interested in non–adaptive information, i.e., information of the form

$$N : X \to \mathbb{R}^n, \quad N(x) = [\lambda_1(x), \dots, \lambda_n(x)] \in \mathbb{R}^n \ ,$$

where $\lambda_1, \dots, \lambda_n : X \to \mathbb{R}$ are fixed linear functionals. Given such N and some $x \in X$, we wish to approximate the solution $S(x)$, using only the information provided by $N(x)$. (In other words, only approximations $\tilde{S} = \varphi \circ N$ for some mapping $\varphi : \mathbb{R}^n \to Y$ are allowed.) The minimal error achievable by such means is

$$\operatorname{err}(S, N, T) \ := \inf_{\varphi \in \mathcal{M}_n(Y)} \operatorname{err}(\varphi \circ N, T) \ ,$$

where $\mathcal{M}_n(Y)$ denotes the set of *all* mappings $\varphi : \mathbb{R}^n \to Y$. Since no restrictions upon φ are placed, a nearly optimal φ could be very complicated. Thus, it is interesting to compare this minimal error with the error achievable by using only linear mappings. Let $\mathcal{L}_n(Y)$ be the set of all linear mappings $\varphi : \mathbb{R}^n \to Y$, and set

$$\operatorname{err}^{\mathrm{lin}}(S, N, T) \ := \inf_{\varphi \in \mathcal{L}_n(Y)} \operatorname{err}(\varphi \circ N, T) \ .$$

In general, no assumptions upon boundedness or measurability are imposed on N or S, and neither do we assume a priori the completeness of X or Z.

It is well–known that sometimes $\operatorname{err}^{\mathrm{lin}}$ and err coincide. Namely, results of Smolyak and Packel ([12], [17]) guarantee:

Theorem 1. *Let X, Y, S, T, N describe a linear problem. If $Y = \mathbb{R}$, or more generally, $Y = \mathcal{B}(K)$ is the space of all bounded functions over some set K, then* $\operatorname{err}^{\mathrm{lin}}(S, N, T) = \operatorname{err}(S, N, T)$.

On the other hand, there are examples of linear problems where linear algorithms are not optimal. The first such example was found by Micchelli (see [15, NR 4.5.5:5]), and the most spectacular known example was provided by Werschulz and Woźniakowski ([18], see also [15], p. 81–84). The result of [18] can be stated like following.

Theorem 2. *There is a linear problem X, Y, S, T, with X being a pre–Hilbert space such that, for any finite-dimensional N we have* $\operatorname{err}^{\mathrm{lin}}(S, N, T) = \infty$ *while* $\operatorname{err}(S, N, T)$ *can be arbitrary small.*

On the other hand, upper bounds are known for nice cases. We shall say that the linear problem is *bounded* iff

- X and Y are Banach spaces,
- S and N are bounded linear maps,
- F is the closed unit ball B_X of X.

P. Mathé used an argument (originally invented by Pietsch for comparison of s–numbers), to show the following ([7], compare also [11, p. 18]):

Theorem 3. *For any bounded linear problem, we have*

$$\inf_N \mathrm{err}^{\mathrm{lin}}(S,N,T) \;\le\; (1+n^{1/2}) \cdot \inf_N \mathrm{err}(S,N,T) \;,$$

where the infimum is taken over all bounded linear information $N : X \to \mathbb{R}^n$.

2 Extension properties of normed spaces

Let us now forget for the moment about linear problems, and consider two normed spaces E, G with unit balls B_E, B_G. We are interested in the question how well we can extend a bounded linear operator $U : M \to G$, defined on a (large) subspace $M \subseteq E$, to a bounded linear operator $\tilde{U} : E \to G$. We will simply measure the quality of this extension in the size of the operator norm $\|\tilde{U}\| := \sup_{x\in B_E} \|\tilde{U}(x)\|$, relative to the size of $\|U\|$. For a subspace $M \subseteq E$ and a linear $U : M \to G$, let

$$p(U,M,E,G) \;:=\; \inf\Big\{ \sup_{x\in B_E} \|\tilde{U}(x)\| \;:\; \tilde{U} : E \to G \text{ linear}, \tilde{U}_{|M} = U \Big\}$$

be the infimum of norms of possible extensions of U to the whole space. Define

$$p(M,E,G) \;:=\; \sup\big\{ p(U,M,E,G) \;:\; U : M \to G, \sup_{x\in B_E\cap M} \|U(x)\| \le 1 \big\}$$

as the largest such increment of norms for norm 1 operators $U : M \to G$, and introduce

$$p_n(E,G) \;:=\; \sup\big\{ p(M,E,G) \;:\; M \subseteq E,\ \operatorname{codim} M \le n \big\};.$$

Thus, $p_n(E,G)$ is the smallest number such that, for any $\eta > 1$, any subspace $M \subseteq E$ with $\operatorname{codim} M \le n$ and any linear $U : M \to G$, we may find a linear $\tilde{U} : E \to G$ with $\tilde{U}_{|M} = U$ and such that

$$\sup_{x\in B_E} \|\tilde{U}(x)\| \le \eta \cdot p_n(E,G) \cdot \sup_{x\in B_E\cap M} \|U(x)\| \;.$$

The question of when and how good one may extend linear operators was extensively studied in Banach space theory, giving us some powerful estimates of p_n ready to use for our purposes. In the following, we will collect a number of such results. Although we cite the original references for these results below, we recommend the reader to the recent survey article [20], which provides an excellent introduction and overview over the topic.

The most general extension result applies the following bound of the minimal norm of projections onto finite–dimensional and –codimensional subspaces, a famous result which was found independently by Garling and Gordon ([2]) and Kadec and Snobar ([3]).

Theorem 4. *Let M be a closed subspace of the normed space E with* $\operatorname{codim} M \leq n$. *Then, for any $\varepsilon > 0$ there exists a projection $P : E \to E$ onto M (i.e., $P(E) = M$) such that $\|P\| \leq 1 + n^{1/2} + \varepsilon$. If instead of the codimension, we have* $\dim M \leq n$, *then we may find a projection P onto M with $\|P\| \leq n^{1/2}$.*

Actually, the statement of these results in the original articles [3], [2] is only made for the case of a complete (Banach) space E. However, for the finite–codimensional case, the proof e.g. of [2] verbally carries over to the case of normed spaces, while the generalization of the above result in the case of a finite–dimensional M is trivial.

From this result, we may draw a useful

Corollary 5. *If G is complete and E normed, $p_n(E,G) \leq 1 + n^{1/2}$.*

Proof. Let $M \subseteq E$ be given and $U : M \to G$ bounded with norm 1. Since G is complete, we may extend U to the closure $\bar{M}$ with the same norm, so let $\bar{U} : \bar{M} \to G$ be the extension of U, still with norm 1. For arbitrary $\varepsilon > 0$ and P as in Theorem 4, $\tilde{U} := \bar{U} \circ P$ is an extension of U to the whole space E, and $\|\tilde{U}\| \leq \|P\| = 1 + n^{1/2} + \varepsilon$. □

Remark: We note in passing that this result is semi–constructive in the following sense: If one knows a priori how to construct a good projection P, then one also knows how to construct good extensions. In some more specific cases it is easy to give optimal projections, sometimes of much smaller norm. A canonical example is given by $E = C([0,1])$ and $M = \{f(t_1) = \cdots = f(t_n) = 0\}$, with $\sigma(f)$ the linear interpolating spline with knots $t_1, \ldots, t_n$. Here, the mapping $P(f) := f - \sigma(f)$ is obviously a projection with $\|P\| = 2$. Hence, $p(M, C[0,1], G) \leq 2$.

What happens if G is not complete? In general, we can say nothing, but if $M \subseteq E$ is closed, the above argument works.

Proposition 6.

(i) *If E, G are normed and $M \subseteq E$ is closed, then $p(M,E,G) \leq 1 + n^{1/2}$.*

(ii) *If G is not complete, then there are M, E with* $\dim E/M = 1$ *and such that $p(M,E,G) = \infty$.*

Proof. The proof of the first part runs in analogy to the proof of Corollary 5. For the second part, let E_0 be the completion of G and choose $E \subseteq E_0$ such that $E = G \oplus B_1$, where B_1 is some 1–dimensional subspace. Considering now the identity $R = \mathrm{Id} : G \to G$, we obviously can't extend it continuously to a mapping $\tilde{R} : E \to G$, and hence, $p(G,E,G) = \infty$. □

Let us return to the case when G is complete. Here, one can improve the factor $1 + n^{1/2}$ if one knows more about E and G. The first idea is to exploit properties of the target space. The following theorem is an easy consequence of classical results due to Kantorovič (Part (i), [4]) and Lindenstrauss (Part (ii), [5]). Here and in the following, $C(K)$ shall be the Banach space of continuous functions over some compact Hausdorff space K, and $L_p(\mu)$ the classical Lebesgue space over some measure space $(\Omega, \mathcal{A}, \mu)$.

Theorem 7. *Let E be a normed space.*

(i) *For* $G \in \{\mathbb{R}, L_\infty(\mu), \mathcal{B}(K)\}$*, we have* $p_n(E,G) = 1$ *for all* $n \in \mathbb{N}$*.*

(ii) *If* $M \subseteq E$ *and* $R : M \to C(K)$ *is compact (i.e., the set* $R(B_E) \subseteq C(K)$ *is precompact), then* $p(R, M, E, C(K)) = 1$*.*

The second idea is to exploit properties of E, or of both spaces. We mention two results, due to Lewis (Part (i), [6]) and Maurey (Part (ii), [8]).

Theorem 8. *Assume that* G *is a Banach space.*

(i) *For any* $p \in (1,\infty)$*, we have* $p_n(L_p(\mu), G) \leq 1 + \ n^{|1/2-1/p|}$*. (If* E *is pre–Hilbert, then* $p_n(E,G) = 1$*.)*

(ii) *For* $q \leq 2 \leq p$*, it holds* $p_n(L_p(\mu), L_q(\mu)) \leq C_{p,q}$ *for all* $n \in \mathbb{N}$*, with* $C_{p,q}$ *depending only on* p,q*.*

Remark: The first part of this result can be stated in more generality. In fact, the result [14, Prop.27.4] easily yields that if the Banach space E is of type p and of cotype q, it follows that $p_n(E,G) \leq T_p(E)C_p(E)(1 + n^{|1/p-1/q|})$, where $T_p(E)$ and $C_q(E)$ denote the type and cotype constants of E. Similarly, the second part is also valid (and in fact, formulated in [8]) in more generality: If E has type 2 and G cotype 2, then $p_n(E,G) \leq T_2(E)C_2(G)$. We shall not explain the notion of type and cotype here, but rather state that they provide a description of spaces similar in geometry to L_p spaces, and refer the reader to [19], [10] or [14] for a detailed introduction into this topic.

We may of course extend mappings into finite–dimensional spaces.

Proposition 9.

(i) *If* $\dim G = d$*, then for any* $M \subseteq E$ *we have* $p(M, E, G) \leq d^{1/2}$*.*

(ii) *Denote* ℓ_2^d *the* d*–dimensional Euclidean space. Then there are* $c > 0$ *and spaces* E_d *such that* $\sup_n p_n(E_d, \ell_2^d) \geq c \cdot d^{1/2}$ *for all* d*. Thus, the estimate from* (i) *is order optimal in* d*.*

Proof. Part (i): We may always embed G into some $L_\infty(\mu)$ space, i.e. find $\tilde{G}$ isometric to G and a subspace $Z \subseteq L_\infty(\mu)$ such that

$$L_\infty(\mu) \ = \ \tilde{G} \oplus Z$$

as a topological direct sum. Obviously, $p_n(E,G) = p_n(E,\tilde{G})$, hence we assume wlog that $G = \tilde{G}$. We may expand any bounded operator $U : M \to G$ to an operator $\tilde{U} : E \to L_\infty(\mu)$ with $\|\tilde{U}\| \leq (1+\varepsilon)\|U\|$, according to Part (ii) of Proposition 7. Theorem 4 applies for the d–dimensional subspace G, yielding a projection P onto G with $\|P\| \leq (1+\varepsilon)d^{1/2}$, so the operator $\hat{U} = \tilde{U}P : E \to G$ is a continuation of U with $\|\hat{U}\| \leq (1+\varepsilon)^2\|U\|$. The assertion follows.
Part (ii): Let $\tilde{d} \in \mathbb{N}$. In the following, $C_1, C_2, \ldots$ denote constants independent of $\tilde{d}, d$, etc.. Considering $\tilde{E}_{\tilde{d}} = \ell_\infty^{\tilde{d}}$, we may find by the Dvoretzky Theorem (see [10, p.24]) a subspace $\tilde{G}_{\tilde{d}}$ of dimension

$$d \ := \ \dim \tilde{G}_{\tilde{d}} \ \geq \ 1/C_1 \cdot \log \tilde{d}$$

which is 2–isomorphic to ℓ_2^d, i.e., there is an isomorphism $\tau : \tilde{G}_{\tilde{d}} \to \ell_2^d$ with $\|\tau\| \cdot \|\tau^{-1}\| \leq 2$. Moreover, the projection constant of ℓ_2^d in ℓ_∞ is of order

$d^{1/2}$ (see [14, Thm. 32.9]), meaning that any projection $P : \ell_\infty^{\tilde d} \to \tilde G_{\tilde d}$ has $\|P\| \geq d^{1/2}/C_2$. But since any continuation of the mapping $\mathrm{Id} : \tilde G_{\tilde d} \to \tilde G_{\tilde d}$ to $\tilde E_{\tilde d}$ has to be such a projection, we derive for $n := \operatorname{codim} \tilde G_{\tilde d} = \tilde d/d$ that $p_n(\tilde E_{\tilde d}, \tilde G_{\tilde d}) \geq d^{1/2}/C_2$. Now, by extending the isomorphism τ to $\tilde E_{\tilde d}$ we get a new space E_d with $p_n(E_d, \ell_2^d) \geq d^{1/2}/C_4$. □

3 How this helps us

We will now introduce a simple relation between the contents of the two preceding sections. (The underlying idea is essentially already apparent in Section 2.4 of the monograph [1], where the relation of approximation and Gel'fand numbers of operators was studied.) For our considerations, the factor space $X_T := X/\ker(T)$, equipped with the norm

$$\|x + \ker(T)\|_{X_T} := \|Tx\|_Z ,$$

will play an important role. (Note that for a normed space X with $T = \mathrm{Id}$, $X_T = X$.) We shall study the *diameter of information*,

$$\operatorname{diam}(N,T) \;:=\; \sup_{x\in F} \operatorname{diam}\; S\big(\{\, y \in F \;:\; N(y) = N(x) \,\}\big) \;.$$

Since $\operatorname{err}(S,N,T)$ equals the (similarly defined) radius of information, we have

$$\operatorname{err}(S,N,T) \;\leq\; \operatorname{diam}(N,T) \;\leq\; 2\operatorname{err}(S,N,T) \,, \tag{1}$$

see e.g. [15, Sect. 4.2, 4.5.4]. It is well–known ([15, Sect. 4, Lemma 5.2.1]) that

$$\operatorname{diam}(N,T) \;=\; 2 \cdot \sup_{x\in F\cap\ker(N)} \|S(x)\| \;. \tag{2}$$

We now return to our problem. For $N : X \to \mathbb{R}^n$, we introduce $N_T := \big(\ker(N) + \ker(T)\big)/\ker(T) \subseteq X_T$. This need not be a closed subspace, but satisfies $\operatorname{codim} N_T \leq n$. The following result generalizes Proposition 2.4.1 of [1].

Proposition 10. *Let X, Y, S, T, N describe a linear problem. Then*

$$\operatorname{err}^{\mathrm{lin}}(S,N,T) \;\leq\; \tfrac{1}{2}\, p\big(N_T, X_T, Y\big) \cdot \operatorname{diam}(N,T) \;.$$

In particular,

$$\operatorname{err}^{\mathrm{lin}}(S,N,T) \;\leq\; \tfrac{1}{2}\, p_n(X_T, Y) \cdot \operatorname{diam}(N,T) \;\leq\; p_n(X_T, Y) \cdot \operatorname{err}(S,N,T) \;.$$

Proof. Let $M = \ker N$, $K = \ker T$, so $N_T = (M+K)/K \subseteq X_T$. We assume wlog that $\operatorname{diam}(N,T) < \infty$, otherwise there is nothing to show. As can be straightforwardly checked, this implies that $S(M \cap K) = \{0\}$. Consequently, the mapping

$$S_T : N_T \to Y \;, \qquad S_T(x + K) \;:=\; S(x)$$

is well–defined and linear. It is also bounded, since

$$\sup_{(x+K)\in B_{X_T}\cap N_T} \|S_T(x+K)\| \;=\; \sup_{x\in F\cap M} \|S(x)\| \;=\; \operatorname{diam}(N,T)/2 \;<\; \infty \;.$$

Hence, S_T is subject to the extension properties elaborated above, i.e., given some $\eta > 1$, we may find a linear $U_T : X_T \to Y$ such that $(U_T)_{|N_T} = S_T$, and

$$\begin{aligned}\sup_{(x+K)\in B_{X_T}} \|U_T(x+K)\| &\leq \eta \cdot p(N_T, X_T, Y) \cdot \sup_{(x+K)\in B_{X_T}\cap N_T} \|S_T(x+K)\| \\ &= \eta \cdot p(N_T, X_T, Y) \cdot \operatorname{diam}(N,T)/2 \ .\end{aligned}$$

Let $Q_T : X \to X_T$ denote the algebraic quotient mapping, set $U = U_T \circ Q_T : X \to Y$, and finally $R := S - U$. Then we derive that

$$\sup_{x\in F} \|R - S\| \;=\; \sup_{x\in F} \|U(x+K)\| \;\leq\; \eta\; p(N_T, X_T, Y) \cdot \operatorname{diam}(N,T)/2 \ .$$

Observe that from the definition of S_T and U_T we have $\ker(N) \subseteq \ker(R)$. We write $N(x) = [\lambda_1(x), \ldots, \lambda_n(x)]$ for $\lambda_i : X \to \mathbb{R}$ linear. Since we may assume wlog that the λ_i are linearly independent, we may find $x_i \in X$ such that $\lambda_i(x_j) = \delta_{i,j}$. Hence, for the projection $P(x) := \sum_{i=1}^n \lambda_i(x) x_i$ it follows that $\lambda_i(x) = \lambda_i(P(x))$, i.e., $x - P(x) \in \ker(N)$, which implies $R(x - P(x)) = 0$. In other words, $R = RP$, or $R(x) = \varphi\big(N(x)\big)$ with

$$\varphi : \mathbb{R}^n \to Y \ , \qquad \varphi([t_1, \ldots, t_n]) \;:=\; \sum_{i=1}^n t_i \cdot R(x_i)$$

obviously being linear. We thus showed that

$$\operatorname{err}^{\mathrm{lin}}(S, N, T) \;\leq\; \eta\; p(N_T, X_T, Y) \cdot \operatorname{diam}(N,T)/2 \ ,$$

and since $\eta > 1$ was arbitrary, we are done. □

We may partially reverse the above statement, showing that we did not lose much in our reasoning.

Proposition 11. *Let X, Y, T, N be given and M, N_T be as above. For any $\eta < 1$ there is a linear problem $S : X \to Y$ such that*

$$\operatorname{err}^{\mathrm{lin}}(S, N, T) \geq \eta \cdot p(N_T, X_T, Y) \cdot \operatorname{err}(S, N, T)/2 \ .$$

In particular,

$$\sup_{S,N} \frac{\operatorname{err}^{\mathrm{lin}}(S, N, T)}{\operatorname{err}(S, N, T)} \in \big[p_n(X_T, Y)/2, p_n(X_T, Y)\big] \ ,$$

where the left hand supremum is taken over all linear problems S and all linear information $N : X \to \mathbb{R}^n$. (Here, we set $\infty/\infty := 0/0 := 0$.)

Proof. We may find a bounded linear map $R : N_T \to Y$ such that

$$\sup_{x_T\in B_{X_T}\cap N_T} \|R(x_T)\| = 1$$

while for any linear continuation $U : X_T \to Y$ of R (i.e., $U_{|N_T} = R$) we have

$$\sup_{x_T\in B_{X_T}} \|U(x_T)\| \geq \eta \cdot p(N_T, X_T, Y) \ .$$

Consider now an arbitrary linear continuation $\tilde{S} : X_T \to Y$ of R, and set

$$S := \tilde{S} \circ Q_T : X \to Y \ .$$

We observe first that

$$\sup_{x \in F \cap \ker(N)} \|S(x)\| = \sup_{x_T \in B_{X_T} \cap N_T} \|\tilde{S}(x_T)\| = \sup_{x_T \in B_{X_T} \cap N_T} \|R(x_T)\| = 1 \ ,$$

which by virtue of (1) implies $\mathrm{err}(S, N, T) \leq 2$. We are done if we can show that $\mathrm{err}^{\mathrm{lin}}(S, N, T) \geq \eta \cdot p(N_T, X_T, Y)$. We may assume that $\mathrm{err}^{\mathrm{lin}}(S, N, T) < \infty$, and consider an arbitrary linear approximation $\bar{S} = \varphi \circ N$ of S such that $\mathrm{err}(\bar{S}, S) < \infty$. Obviously, this implies $(\bar{S} - S)_{|\ker T} = 0$, and hence $\bar{S} - S$ may be factorized through X_T,

$$\bar{S} - S = S_T \circ Q_T \ .$$

Furthermore, S_T is a continuation of R, since $(S_T - \tilde{S})(N_T) = \bar{S}(\ker(N)) = (\varphi \circ N)(\ker(N)) = \{0\}$, and we derive

$$\sup_{x \in F} \|\bar{S}(x) - S(x)\| = \sup_{x_T \in B_{X_T}} \|S_T(x_T)\| \geq \eta \cdot p(N_T, X_T, Y) \ .$$

Thus, $\mathrm{err}^{\mathrm{lin}}(S, N, T) \geq \eta \cdot p(N_T, X_T, Y)$, and the claim follows. □

The above relation makes the results stated in Section 2 available for our purposes; by simple translation, we can draw a number of corollaries. Let us start with the most general result:

Proposition 12. *Let X, Y, S, T, N describe an arbitrary linear problem with $N : X \to \mathbb{R}^n$. If Y is complete, it follows that*

$$\mathrm{err}^{\mathrm{lin}}(S, N, T) \ \leq \ (1 + n^{1/2}) \cdot \mathrm{err}(S, N, T) \ .$$

Proof. This follows immediately from Proposition 10 and Corollary 5. □

This result generalizes Theorem 3. As a second consequence, we derive from Propositions 7, 10:

Proposition 13. *Let X, Y, S, T, N describe a linear problem.*

(i) *If $Y \in \{\mathbb{R}, L_\infty(\mu), \mathcal{B}(K)\}$. Then $\mathrm{err}^{\mathrm{lin}}(S, N, T) = \mathrm{err}(S, N, T)$.*

(ii) *If $Y = C(K)$, and $S(F) \subseteq C(K)$ is precompact, $\mathrm{err}^{\mathrm{lin}}(S, N, T) = \mathrm{err}(S, N, T)$.*

This includes Smolyak's and Packel's Theorem ([12], [17], [15, Section 5.5]).

Proof. Part (i) is a trivial consequence of the above results. Part (ii) is easily derived by noting that, if $S(F) \subseteq C(K)$ is a precompact subset, then the operator S_T introduced in the proof of 10 is also compact, and hence $p(S_T, N_T, X_T, C(K)) = 1$ by Part (ii) of Theorem 7, which entails the assertion. □

Next, we have a look at what Theorem 8 entails in conjunction with Proposition 10:

Proposition 14. *Let X, Y, S, T, N describe a linear problem.*

(i) *If X_T is isometric to some $L_p(\mu)$ space, then*

$$\mathrm{err}^{\mathrm{lin}}(S, N, T) \;\leq\; (1 + n^{|1/2 - 1/p|}) \cdot \mathrm{err}(S, N, T) \;.$$

If X is pre–Hilbert, this improves to

$$\mathrm{err}^{\mathrm{lin}}(S, N, T) = \mathrm{err}(S, N, T) \;.$$

(ii) *Assume that $Y = L_q(\mu)$, and X_T is isometric to $L_p(\mu)$ for some $q \leq 2 \leq p$. Then*

$$\mathrm{err}^{\mathrm{lin}}(S, N, T) \;\leq\; C_{p,q}\; \mathrm{err}(S, N, T)$$

with $C_{p,q}$ depending solely on p, q.

Remark: Following the remark after Theorem 8, one can also easily generalize the above statements to spaces of type and cotype.

The estimates are order optimal. To see this, consider the classical approximation problem for the Sobolev spaces $W_1^p[0,1]$ of absolutely continuous functions f such that $f' \in L_p$ and $f(0) = 0$, equipped with the norm $\|f\|_{W_1^p} = \|f'\|_{L_p}$. If we consider the approximation on the unit ball of W_1^p in L_2, i.e., $F = B_{W_1^p}$, $S := \mathrm{App}(W_1^p, L_2) := \mathrm{Id} : W_1^p \to L_2$, we can introduce[1] $T : W_1^p \to L_p, f \mapsto f'$, and are in the situation of Part (ii) of the proposition. For $p \in [1,2]$, it is well–known (see, e.g., [13, p. 232]) that for suitably chosen bounded $N_n : X \to \mathbb{R}^n$ and $c_p, C_p \in (0, \infty)$ it holds

$$\mathrm{err}(\mathrm{App}(\mathrm{W}_1^{\mathrm{p}}, \mathrm{L}_2), N_n) \;\leq\; C_p \cdot n^{-1} \;,$$

while for any bounded information $N_n' : X \to \mathbb{R}^n$ we have

$$\mathrm{err}^{\mathrm{lin}}(S, N_n') \;\geq\; c_p \cdot n^{-1 + (1/p - 1/2)} \;.$$

Part (ii) of the theorem entails in particular that $App(W_1^p, L_q)$ has order optimal linear solutions whenever $q \leq 2 \leq p$ (which is also well-known of course).

Let us turn to the case of an incomplete target space.

Proposition 15. *Let X, Y, S, T, N describe a linear problem.*

(i) *If X is a normed space with $T = \mathrm{Id_X}$ and the information $N : X \to \mathbb{R}^n$ is bounded, then regardless of Y we have*

$$\mathrm{err}^{\mathrm{lin}}(S, N, T) \;\leq\; (1 + n^{1/2}) \cdot \mathrm{err}(S, N, T) \;.$$

(ii) *If Y is not complete, then for any $\varepsilon \geq 0$ there is a normed space X, $T = \mathrm{Id_X}$, an operator $S : X \to Y$ and information $N : X \to \mathbb{R}$ such that $\mathrm{err}^{\mathrm{lin}}(S, N, T) = \infty$ while $\mathrm{err}(S, N, T) = \varepsilon$.*

Proof. We just combine Proposition 6 with Propositions 10 and 11. (Note that, by homogeneity, it suffices to consider $\varepsilon = 1$ in the last part.) ☐

[1] If one follows the above remark, one can simply use the fact that W_1^p is of type $\min\{2, p\}$ and cotype $\max\{2, p\}$, and thus skip the workaround of introducing T.

Concerning part (ii), this result is more general, but not as strong as the result of Theorem 2, since there the infimum over $\mathrm{err}(S,N,T)$ is actually 0, while we only provide that it is not greater than ε.

Finally, we give a (order optimal) generalization of Smolyak's Theorem to finite–dimensional target spaces.

Proposition 16. *Let X, Y, S, T, N describe a linear problem.*

(i) *If* $\dim Y = d$*, then*

$$\mathrm{err}^{\mathrm{lin}}(S,N,T) \;\le\; d^{1/2} \cdot \mathrm{err}(S,N,T)\;.$$

(ii) *For $Y_d = \ell_2^d$ the d–dimensional Euclidean space, there are linear problems X_d, Y_d, S_d, N_d with $T_d = \mathrm{Id}$ such that*

$$\frac{\mathrm{err}^{\mathrm{lin}}(S_d,N_d,T_d)}{\mathrm{err}(S_d,N_d,T_d)} \;\ge\; c \cdot d^{1/2}$$

for some $c > 0$, i.e. the above estimate is order optimal in d.

Proof. Here, we apply Proposition 9 together with our main tools, Propositions 10 and 11. □

4 Adaptive information and bounded noise

The estimates found in the previous section easily generalize to adaptive information. To be more precise, let N^{ad} be an adaptive information operator,

$$N^{\mathrm{ad}}(x) = [\lambda_1(x), \lambda_2(x,y_1), \ldots, \lambda_{n(x)}(x,y_1,\ldots,y_{n(x)-1})]\;,$$

where $\lambda_i : X \times \mathbb{R}^{i-1} \to \mathbb{R}$ are such that, for fixed $t_j \in \mathbb{R}$, $\lambda_i(\cdot, t_1, \ldots, t_{i-1})$ are linear, $y_i := \lambda_i(x, y_1, \ldots, y_{i-1})$, and $n : X \to \mathbb{N}$. Since we are dealing with linear problems, adaption does not help much. Namely, the non–adaptive information operator

$$N^{\mathrm{non}}(x) \;:=\; [\lambda_1(x), \lambda_2(x,0), \ldots, \lambda_{n(0)}(x,0,\ldots,0)]\;.$$

is almost as good as N^{ad}: If we denote $\mathrm{err}(S,N^{\mathrm{ad}},T)$ the minimal error achievable by using N^{ad} and arbitrary algorithms, then Theorem 5.2.1 of [15] and (1) immediately yield that

$$\mathrm{diam}\,(N^{\mathrm{non}},T) \;\le\; 2\;\mathrm{err}(S,N^{\mathrm{ad}},T)\;. \tag{3}$$

Using this estimate as a substitute for (1), we easily find:

Proposition 17. *Let $X, Y, S, T, N^{\mathrm{ad}}$ and N^{non} be as above, with $n(x)$ the cardinality of information $N^{\mathrm{ad}}(x)$. All statements of Propositions 12–16 also hold if we consider the adaptive information N^{ad} and replace in the statements*

- $\mathrm{err}(S,N,T)$ *by* $\mathrm{err}(S,N^{\mathrm{ad}},T)$,
- $\mathrm{err}^{\mathrm{lin}}(S,N,T)$ *by* $\mathrm{err}^{\mathrm{lin}}(S,N^{\mathrm{non}},T)$,
- n *by* $n(0)$.

Finally we have a short look at noisy information. We restrict ourselves to the case of uniformly bounded noise. For simplicity, we consider only non–adaptive information. Assume that instead of exact information $N(x)$ we observe some noisy information $N(x)+y$, where $y=[y_1,\dots,y_n]$ is noise coming from some (unspecified) source. In the setting of uniformly bounded noise one assumes that the noise vector is guaranteed to be not too large, i.e. for some norm $\|\cdot\|$ on $\mathbb{R}^n$ and some $\eta \geq 0$ we have $\|y\| \leq \eta$. In the case $\eta = 0$ we are back to the 'exact information' case studied above, hence we assume that $\eta > 0$. Denote by E_n the Banach space induced by $\|\cdot\|$ with unit ball F_n. If we consider a linear problem $S : X \to Y$ with T, F as above, the task is to find a mapping $\varphi : \mathbb{R}^n \to Y$ which uses the noisy information to produce a good approximation of S. If we do not know more about the noise than boundedness, it is reasonable to use

$$\mathrm{err}(\varphi, N, T, F_n) \;:=\; \sup_{x\in F, \|y\|\leq\eta} \|S(x) - \varphi(N(x)+y)\|$$

as error measure. We denote by $\mathrm{e}(S,N,T,F_n,\eta)$ the infimum over all such errors with arbitrary mappings φ, and by $\mathrm{e}^{\mathrm{lin}}(S,N,T,F_n,\eta)$ the minimal error achieved by using linear mappings φ. Although this seems to be a generalization of the 'exact information' setting above, it can in turn be rewritten as a special 'exact information' problem, and by this we can generalize the above results. Set $\bar{X} := X \times \mathbb{R}^n$, $\bar{Z} := Z \times \mathbb{R}^n$ with $\|(z,y)\| := \max\{\, \|z\|_Z, \|y\|_{E_n}/\eta \,\}$, and $\bar{T} := T \times Id : \bar{X} \to \bar{Z}$. Further, let

$$\bar{N} : \bar{X} \to \mathbb{R}^n, \qquad \bar{N}(x,y) := N(x) + y\ .$$

Then for any $\varphi : \mathbb{R}^n \to Y$ it is well–known ([9], see also [15, Section 12.2]) that

$$\mathrm{err}(\varphi, N, T, F_n, \eta) \;=\; \mathrm{err}(\varphi \circ \bar{N}, \bar{T})\ ,$$

and consequently we have $\mathrm{e}(S,N,T,F_n,\eta) = \mathrm{e}(\bar{S},\bar{N},\bar{T})$ and $\mathrm{e}^{\mathrm{lin}}(S,N,T,F_n,\eta) = \mathrm{e}^{\mathrm{lin}}(\bar{S},\bar{N},\bar{T})$. Hence, Proposition 12 allows the upper bound

$$\mathrm{e}^{\mathrm{lin}}(S,N,T,F_n,\eta) \;\leq\; p(\bar{N}_{\bar{T}}, \bar{X}_{\bar{T}}, Y)\cdot \mathrm{e}(S,N,T,F_n,\eta)\ . \tag{4}$$

A closer look at the couple $(\bar{N}_{\bar{T}}, \bar{X}_{\bar{T}})$ is needed now. The space E_n' shall be E_n with the dilated norm $\|\cdot\|_{E_n'} := \|\cdot\|_{E_n}/\eta$. First, note that $\ker \bar{T} = \ker T \times \{0\}$, and this implies that $\bar{X}_{\bar{T}}$ is isometrically isomorphic to $X_T \times E_n'$, with the isometry given by $\Phi\big((x,y) + \ker \bar{T}\big) := (x + \ker T, y)$. The image of $\bar{N}_{\bar{T}}$ under this isometry is

$$N_T' \;=\; \Big\{ (x + \ker T, -N(x)) : x \in X \Big\}\ ,$$

and thus,

$$p(\bar{N}_{\bar{T}}, \bar{X}_{\bar{T}}, Y) \;=\; p(N_T', X_T \times E_n', Y)\ . \tag{5}$$

Furthermore, since $\bar{N}$ maps into $\mathbb{R}^n$, the isomorphic theorem tells us that $\dim(\bar{X}/\ker(\bar{N})) \leq n$, which easily implies that $\dim(\bar{X}_{\bar{T}}/\bar{N}_{\bar{T}}) \leq n$, and due to the isomorphism, $\dim(X_T/N_T') \leq n$. Consequently, (5) may be continued to

$$p(\bar{N}_{\bar{T}}, \bar{X}_{\bar{T}}, Y) \;\leq\; p_n(X_T \times E_n', Y)\ . \tag{6}$$

Summing up (4) and (6), we have found

$$\mathrm{e}^{\mathrm{lin}}(S,N,T,F_n,\eta) \;\le\; p_n(X_T \times E_n', Y)\cdot \mathrm{e}(S,N,T,F_n,\eta)\;. \tag{7}$$

This estimate is not completely satisfying, since the space E_n' is appearing as an additional factor, compared to Proposition 10. In particular, this forbids the use of specific properties of X_T[2]. However, (7), together with Proposition 8 and Corollary 5 allows to generalize Proposition 12 and 13, and also the first part of Proposition 16.

Proposition 18. *Let S,X,Y,T,N,F_n,η describe a linear problem with uniformly bounded noise as above, and assume that Y is complete.*

(i) *We have*

$$\mathrm{e}^{\mathrm{lin}}(S,N,T,F_n,\eta) \;\le\; (1+n^{1/2})\cdot \mathrm{e}(S,N,T,F_n,\eta)\;.$$

(ii) *If $Y \in \{\,\mathbb{R},\mathcal{B}(K),L_\infty(\mu)\,\}$, then*

$$\mathrm{e}^{\mathrm{lin}}(S,N,T,F_n,\eta) \;=\; \mathrm{e}(S,N,T,F_n,\eta)\;.$$

The same holds if $Y = C(K)$ and $S(F)\subseteq C(K)$ is precompact.

(iii) *If $\dim Y \le d$, we have*

$$\mathrm{e}^{\mathrm{lin}}(S,N,T,F_n,\eta) \;\le\; d^{1/2}\cdot \mathrm{e}(S,N,T,F_n,\eta)\;.$$

The second part of the proposition is well–known at least for $Y \in \{\mathbb{R},\mathcal{B}(K)\}$, see e.g. [15, p. 437]. The rest, however, seems to be new. The results are not as tight as the ones for exact information; in particular the lower bound (Proposition 11) is not applicable. This indicates that further study and/or different methods could lead to a better understanding of our questions in this case, e.g. for the question whether Proposition 14 can be generalized to the noisy setting.

Acknowledgment: The authors are most indebted to Henryk Woźniakowski for his continued interest and his encouragement to write this paper as well as many helpful and important remarks, and are grateful to Nicole Tomczak–Jaegermann for pointing us towards Proposition 27.4 of [14].

References

[1] Carl, B.; Stephani, I.: *Entropy, compactness and the approximation of operators.* Cambridge University Press, 1989.

[2] Garling, D.J.H.; Gordon, Y.: *Relations between some constants associated with finite dimensional Banach spaces.* Israel J. Math. **9** (1971), 346–361.

[2] In the case of Hilbert noise, i.e., E_n being a Hilbert space, one can fix this at the cost of an additional constant, as is easily seen.

[3] KADEC, M.I.; SNOBAR, M.G.: *Certain functionals on the Minkowksi compactum.* Mat. Zametski **10** (1971), 453–458 (Russian).

[4] KANTOROVIČ, L.V.: *On semi–ordered linear spaces and their applications to the theory of linear operators.* Doklady AN SSSR **4** (1935), 11–14 (Russian).

[5] LINDENSTRAUSS, J.: *On the extension property for compact operators.* Bull. AMS **68** (1962), 484–487.

[6] LEWIS, D.R.: *Finite dimensional subspaces of L_p.* Studia Math. **63** (1978), 207–212.

[7] MATHÉ, P.: *s–Numbers in Information-Based Complexity.* J. Compl. **6** (1990), 41–66.

[8] MAUREY, B.: *Un théorème de prolongement.* C.R. Acad. Sci. Paris **279** (1974), 329–332.

[9] MICCHELLI, C.A.; RIVLIN, T.J.: *A survey of optimal recovery.* In "Optimal Estimation in Approximation Theory", (eds.: C.A. Micchelli/T.J. Rivlin), Plenum, New York, 1977.

[10] MILMAN, V.D.; SCHECHTMAN, G.: *Asymptotic Theory of Finite Dimensional Normed Spaces.* LNM 1200, Springer 1986 (2nd printing 2001).

[11] NOVAK, E.: *Deterministic and Stochastic Error Bounds in Numerical Analysis* Lecture Notes Math. 1349, Springer–Verlag, 1988.

[12] PACKEL, E. W.: *Linear problems (with extended range) have linear optimal algorithms.* Aequationes Math. **30** (1986), 18–25.

[13] PINKUS, A.: *n–Widths in Approximation Theory.* Springer, 1985.

[14] TOMCZAK–JAEGERMANN, N: *Banach–Mazur distances and finite dimensional Operator Ideals.* Pitman Monographs and Surveys **38**, 1989.

[15] TRAUB, J. F.; WASILKOWSKI, G. W.; WOŹNIAKOWSKI, H.: *Information–based Complexity.* Academic Press, 1988.

[16] TRAUB, J. F.; WOŹNIAKOWSKI, H.: *A General Theory of Optimal Algorithms.* Academic Press, 1980.

[17] SMOLYAK, S.A.: *On Optimal Restoration of functions and functionals of them (Russian).* Candidate Dissertation, Moscow State University, 1965.

[18] WERSCHULZ, A. G.; WOŹNIAKOWSKI, H.: *Are linear algorithms always good for linear problems?* Aequationes Math. **31** (1986), 202–212.

[19] WOJTASZCZYK, P.: *Banach spaces for analysts.* Cambridge University Press, 1991.

[20] ZIPPIN, M.: *Extension of bounded linear operators.* Ch. 40 in the *Handbook of the Geometetry of Banach Spaces, Part II.* (Eds.:W.B. Johnson, J. Lindenstrauss), North–Holland 2003.