N-WIDTHS AND OPTIMAL RECOVERY

A. PINKUS

ABSTRACT. These lecture notes are intended as a short introduction
to the theory of n-widths, and to the theory of optimal recovery.
Some simple examples are studied in detail in an attempt to explain
and motivate the main ideas.

1. GENERAL INTRODUCTION. In this lecture I hope to whet the reader's interest in both the theory of n-widths and the theory of optimal recovery. As such, these notes are intended as an introduction to the subject matter, and not as an overview or survey.

The twinning of the two topics of n-widths and optimal recovery in one lecture is somewhat artificial. Nonetheless a relationship does exist in the type of problems considered. Both subjects differ from the more classical problems of approximation theory in that they are concerned with determining optimal subspaces, operators, algorithms, or whatever, with which to approximate elements of an a priori given set. However because we are dealing with two topics we have divided these notes into two distinct parts. In Section A we discuss n-widths, and in Section B optimal recovery. In each of the sections we present a simple example and try, using the example, to motivate some of the main concepts and ideas of the theory. Readers interested in more comprehensive surveys are urged to consult references [3], [4], [6], [7] and [9].

A. N-WIDTHS

2. INTRODUCTION. Perhaps "the" classic problem of approximation theory is the following. Given an element x of a normed linear space X, and an n-dimensional subspace X_n of X, find a best approximation to x from X_n, and determine the value of the error, i.e. the measure of the distance of x from a best approximant. Thus, for example, if X = H is a Hilbert space over $\mathbb{C}$ with inner product $(\cdot,\cdot)$, and X_n is spanned by $v_1,\ldots,v_n$ which, for convenience, we

---

1980 Mathematics Subject Classification. 41A46, 41A65.

assume to be an orthonormal basis for X_n, then $\sum_{i=1}^{n} (x,v_i)\, v_i$ is the unique

best approximant to x from X_n. The "error", generally denoted as $E(x;X_n)$, may

be expressed by

$$E(x;X_n) = [\, ||x||^2 - \sum_{i=1}^{n} |(x,v_i)|^2 \,]^{1/2}.$$

Very often one is not so much interested in approximating a given element
of X, but in approximating a subset A of X. By this we mean determining

$$E(A;X_n) = \sup_{x \in A} \inf_{y \in X_n} ||x - y||.$$

There are many reasons for considering such a quantity. One is often not
interested in a best approximant or in the specific error obtained, but in
measuring the error in terms of some other criteria such as, for example,
smoothness. We consider an example of such a problem.

We denote by $\widetilde{W}_2^{(r)}[0,2\pi]$ the (Sobolev) space of real-valued 2π-periodic
functions on $\mathbb{R}$, for which $f^{(r-1)}$ is absolutely continuous, and whose rth
derivative on $[0,2\pi]$ exists as a function of $L^2[0,2\pi]$. Set $X = L^2[0,2\pi]$, and

$$A \;(= \widetilde{B}_2^{(r)}) = \{f : f \in \widetilde{W}_2^{(r)}[0,2\pi],\; ||f^{(r)}||_2 \leq 1 \}.$$

Let T_n denote the $(2n + 1)$-dimensional subspace of trigonometric polynomials
of degree n, i.e.

$$T_n = \text{span}\{\, 1, \sin x, \cos x,\ldots, \sin nx, \cos nx \,\}.$$

It is not at all difficult to prove that $E(A;T_n) = (n + 1)^{-r}$ for all
non-negative integers n and r.

A more common, but totally equivalent way of stating this result is the
following. For every $f \in \widetilde{W}_2^{(r)}[0,2\pi]$,

$$E(f;T_n) \leq (n + 1)^{-r} ||f^{(r)}||_2,$$

and $(n + 1)^{-r}$ is the best constant in the above inequality. It is often to
obtain inequalities of this form, with best constants, that we study the
quantities $E(A;X_n)$.

As above, we assume that X is a normed linear space and A a subset of X.
For each n-dimensional subspace X_n, we have the associated quantity $E(A;X_n)$. In
1936, Kolmogorov [1] proposed the following idea. Instead of considering $E(A;X_n)$
for different but specific X_n, let us vary $E(A;X_n)$ over all n-dimensional
subspaces X_n of X. We then search for n-dimensional subspaces (if they exist)
which best approximate A, and also for the associated minimum value of $E(A;X_n)$.
To state this in more precise mathematical terms, we have

DEFINITION 1. X is a normed linear space and A a subset of X. The n-width
of Kolmogorov of A in X is given by

$$d_n(A;X) = \inf_{X_n} E(A;X_n)$$

$$= \inf_{X_n} \sup_{x \in A} \inf_{y \in X_n} ||x - y||,$$

where the left-most infimum is taken over all n-dimensional subspaces X_n of X.

We would like, if possible, to identify n-dimensional subspaces X_n^* of X for which $d_n(A;X) = E(A;X_n^*)$. Such subspaces are quite naturally said to be optimal for $d_n(A;X)$.

The quantity $d_n(A;X)$ measures the extent to which A may be approximated by n-dimensional subspaces of X. It is not only that $d_n(A;X)$ is an interesting theoretical quantity (and it is), but also that knowledge of it can help us in other problems. For example, suppose that while we may or may not know $d_n(A;X)$ precisely, we do know something about its asymptotic behaviour as n ↑ ∞. $d_n(A;X)$ is a lower bound on the extent to which A is approximable by n-dimensional subspaces. As such, if we have a given sequence $\{X_n\}$ of n-dimensional subspaces and estimates for $E(A;X_n)$, then it is possible to judge whether it is worthwhile spending energy, time, and money, in using better but more complicated subspaces in our approximation process.

In the above general framework, very little of interest can be said. But let us consider a specific example in detail. Before doing so we remark that many other n-width concepts now abound in the literature. We will touch upon some of these in the next few pages.

3. EXAMPLE. Set $X = L^\infty[0,1]$. The (Sobolev) space $W_\infty^{(1)}[0,1]$ is the set of absolutely continuous real-valued functions defined on [0,1] for which f' exists a.e. as an element of $L^\infty[0,1]$. We define

$$(A =) \ B_\infty^{(1)} = \{ f : f \in W_\infty^{(1)}[0,1], \ ||f'||_\infty \le 1 \},$$

and we are interested in the quantity $d_n(B_\infty^{(1)};L^\infty)$.

A natural approximating subspace to consider is $X_n = \pi_{n-1}$, where π_{n-1} is the set of algebraic polynomials of degree n-1 (dimension n). The quantity $E(B_\infty^{(1)};\pi_{n-1})$ has been much studied and although an exact formula is not known to the best of my knowledge, there does exist from Jackson's Theorem (see e.g. [8,p.22]) the upper bound $E(B_\infty^{(1)};\pi_{n-1}) \le 3/(n-1)$.

We now consider an even more elementary subspace. Let S_n denote the subspace (of dimension n) of left-continuous step functions with jumps at i/n, i=1,..., n-1. Thus $s \in S_n$ if $s(x) = c_i$ on $((i-1)/n,i/n]$, i=1,...,n, for some choice of real constants $\{c_i\}_1^n$. (Modify the first interval to include the point zero.) We claim

PROPOSITION 1. $E(B_\infty^{(1)};S_n) = 1/(2n)$, n = 1,2,... .

PROOF. For $f \in B_\infty^{(1)}$, let $s_f \in S_n$ be uniquely defined by $s_f((2i-1)/2n) = f((2i-1)/2n)$, $i = 1,\ldots,n$. We claim that $||f - s_f||_\infty \leq 1/2n$. For $x \in [0,1]$, we have $x \in ((j-1)/n, j/n]$ for some $j = 1,\ldots,n$ (recall that $x = 0$ is in the first interval). Thus $f(x) - s_f(x) = f(x) - f((2j-1)/2n)$. Since $||f'||_\infty \leq 1$, it follows that

$$|f(x) - f((2j-1)/2n)| \leq |x - ((2j-1)/2n)| \leq 1/2n.$$

Thus $||f - s_f||_\infty \leq 1/2n$. To prove the desired equality, it remains to find an $f^* \in B_\infty^{(1)}$ for which $E(f^*; S_n) = 1/2n$. Set $f^*(x) = x$. Then it is easily seen that $E(f^*; S_n) = 1/2n$, and s_{f^*} is in fact the unique best approximation to f^* from S_n. This proves the proposition. □

We claim that $d_n(B_\infty^{(1)}; L^\infty) = 1/2n$, for $n = 1,2,\ldots$. ($d_0(B_\infty^{(1)}; L^\infty) = \infty$ since every constant function is in $B_\infty^{(1)}$.) From Proposition 1 we have $d_n(B_\infty^{(1)}; L^\infty) \leq 1/2n$. It is therefore necessary to prove the lower bound, i.e. $E(B_\infty^{(1)}; X_n) \geq 1/2n$ for every n-dimensional subspace X_n of $L^\infty[0,1]$. This problem is non-linear in nature and is generally the more difficult. In the proof of this result we use the following general theorem which we will not prove.

THEOREM 2 [2]. Let X be a normed linear space. Let X_{n+1} and X_n be (n+1)- and n-dimensional subspaces of X, respectively. There then exists an $x \in X_{n+1} \backslash 0$ for which

$$E(x; X_n) = ||x||,$$

i.e. the zero element is a best approximation to x from X_n.

This theorem is often used in obtaining lower bounds for n-widths. Let us see how.

Let L_{n+1} denote the (n+1)-dimensional subspace of continuous functions on $[0,1]$ which are linear on $[(i-1)/n, i/n]$, $i = 1,\ldots,n$. The key to the lower bound is Theorem 2 and this next result.

PROPOSITION 3. If $f \in L_{n+1}$ and $||f||_\infty \leq 1/2n$, then $f \in B_\infty^{(1)}$.

PROOF. Obviously $L_{n+1} \subseteq W_\infty^{(1)}$. We must therefore prove that if $f \in L_{n+1}$ and $||f||_\infty \leq 1/2n$, then $||f'||_\infty \leq 1$. Assume that $f \in L_{n+1}$ and $||f||_\infty \leq 1/2n$. Then $|f(i/n)| \leq 1/2n$, $i=0,1,\ldots,n$. Since f is linear on $[(i-1)/n, i/n]$, $i = 1,\ldots,n$, then for $x \in ((i-1)/n, i/n)$,

$$|f'(x)| = n |f(i/n) - f((i-1)/n)| \leq 1. □$$

PROPOSITION 4. If X_n is any n-dimensional subspace of $L^\infty[0,1]$, then $E(B_\infty^{(1)}; X_n) \geq 1/2n$.

PROOF. Let L_{n+1} be as above, and X_n be any n-dimensional subspace of $L^\infty[0,1]$. Then from Theorem 2 there exists a non-zero $f \in L_{n+1}$, which we normalize

so that $||f||_\infty = 1/2n$, satisfying

$$E(f;X_n) = ||f||_\infty = 1/2n.$$

From Proposition 3, $f \in B_\infty^{(1)}$. Thus

$$E(B_\infty^{(1)};X_n) \geq 1/2n. \quad \square$$

To summarize, we have proved the following result.

THEOREM 5. Let $B_\infty^{(1)}$ be as previously defined. Then $d_n(B_\infty^{(1)};L^\infty) = 1/2n$, $n = 1,2,\ldots$. Furthermore S_n, the n-dimensional subspace of step functions with jumps at i/n, $i = 1,\ldots,n-1$, is an optimal subspace for $d_n(B_\infty^{(1)};L^\infty)$.

REMARK. Before proving Theorem 5 we noted that $E(B_\infty^{(1)};\pi_{n-1}) \leq 3/(n-1)$. While we do not know if algebraic polynomials of degree n-1 are optimal for $d_n(B_\infty^{(1)};L^\infty)$, it does follow from the above inequality that they are at least asymptotically optimal in the sense that both quantities decrease to zero at the same rate.

4. OTHER N-WIDTHS. Theorem 5 is a special case of a more general result. We defer the statement of the general result to the next section. We will now use the above example and its proof to motivate additional problems and definitions.

4.A. LINEAR N-WIDTH. In the proof of Proposition 1 we obtained the upper bound constant 1/2n by the simple linear process of interpolating from S_n to each $f \in B_\infty^{(1)}$ at $(2i-1)/2n$, $i = 1,\ldots,n$. That is, we did not calculate the best approximation to each $f \in B_\infty^{(1)}$ from S_{2n}. We calculated instead a linear approximation. This suffices because the quantity $E(A;X_n)$ is a "worst case" measure. It is not necessary when calculating $E(A;X_n)$ that we actually determine $E(x;X_n)$ for each $x \in A$.

In a Hilbert space setting the best approximation is an orthogonal projection and is therefore a linear approximation. This is no longer the case in a non-Hilbert space setting, for there the best approximation operator is a nonlinear operator which is generally exceedingly difficult to exactly determine. Linear approximations are easier to calculate, and are of interest in and of themselves. Let us therefore consider linear approximations rather than best approximations. In other words, we replace $E(A;X_n)$ by

$$E(A;P_n;X_n) = \sup_{x \in A} ||x - P_n x||$$

where P_n is a continuous linear operator from X to X_n. (P_n is said to be of rank n if its range space is of dimension n.)

Analogous to the Kolmogorov n-width we now define what is termed the linear n-width of A in X.

DEFINITION 2. X is a normed linear space and A a subset of X. The <u>linear</u>
n-<u>width of</u> A <u>in</u> X is defined by

$$\delta_n(A;X) = \inf_{P_n} \sup_{x \in A} ||x - P_n x||,$$

where the infimum is taken over all continuous linear operators of rank at
most n.

If $\delta_n(A;X) = E(A;P_n^*;X_n^*)$ where P_n^* is a continuous linear operator of rank
$\leq$ n, then P_n^* is said to be <u>optimal</u> for δ_n.

From our definitions, it follows that $d_n \leq \delta_n$. If X is a Hilbert space,
then $d_n = \delta_n$. In general δ_n is an easier quantity to determine. When d_n and δ_n
are unequal, then serious problems generally arise in the computation of the
former. Both d_n and δ_n depend on "worst case" situations. As such they may be
equal even in a non-Hilbert space setting. This is true of the example of the
previous section.

THEOREM 6. <u>Let</u> $B_\infty^{(1)}$ <u>and</u> S_n <u>be as previously defined. Then</u> $\delta_n(B_\infty^{(1)};L^\infty)$
= 1/2n, n = 1,2,... . <u>Furthermore</u> P_n^*, <u>the rank</u> n <u>linear operator defined by</u>
<u>interpolating from</u> S_n <u>to each</u> $f \in B_\infty^{(1)}$ <u>at</u> (2i-1)/2n, i = 1,...,n, <u>is optimal</u>
<u>for</u> $\delta_n(B_\infty^{(1)};L^\infty)$.

4.B. BERNSTEIN N-WIDTH. Let us examine, in our example, the proof of the lower
bound using Theorem 2. The following idea was used. Assume X_{n+1} is an (n+1)-
dimensional subspace of X, and let $S(X_{n+1})$ denote the unit ball of X_{n+1}. If
$\lambda S(X_{n+1}) \subseteq A$, then from Theorem 2 (see Proposition 4), $d_n(A;X) \geq \lambda$. This
technique for obtaining lower bounds for d_n has been so often used that it has
been codified.

DEFINITION 3. Let X be a normed linear space and A a closed, convex,
centrally symmetric ($x \in A$ implies $-x \in A$) subset of X. The <u>Bernstein</u> n-<u>width</u>
<u>of</u> A <u>in</u> X is defined by

$$b_n(A;X) = \sup_{X_{n+1}} \sup \{ \lambda : \lambda S(X_{n+1}) \subseteq A \},$$

where the X_{n+1} range over all subspaces of X of dimension n + 1.

Thus $d_n(A;X) \geq b_n(A;X)$. The quantity b_n is often of interest, other than
simply as a lower bound for d_n. In our example we proved that $b_n(B_\infty^{(1)};L^\infty)$ = 1/2n.
Restating this we may write

$$\sup_{\substack{f \in X_{n+1} \\ f \neq 0}} ||f'||_\infty / ||f||_\infty \geq 2n, \quad n = 1,2,...,$$

for any $(n+1)$-dimensional subspace X_{n+1} of $W_\infty^{(1)}$. From Proposition 3, equality holds with $X_{n+1} = L_{n+1}$.

This constant $2n$ is, in some sense, a smallest "smoothness constant" relating the size of f' with that of f. This fact should be considered in the light of Markov's inequality. The well-known Markov's inequality for algebraic polynomials of degree n states that if $p\in\pi_n$, then on $[0,1]$

$$||p'||_\infty \le 2n^2 ||p||_\infty,$$

and equality is attained (by the Chebyshev polynomial of the first kind). Equivalently we may write

$$\sup_{\substack{p\in\pi_n \\ p\neq 0}} ||p'||_\infty/||p||_\infty = 2n^2.$$

Algebraic polynomials are thus far from being optimal in the above sense.

4.C. GEL'FAND N-WIDTH. There is one other n-width concept which we wish to introduce and it is the <u>Gel'fand n-width of</u> A <u>in</u> X. It is related to the Kolmogorov n-width via a duality relationship which we will not discuss. It is considered again in Section 6 and will reappear when dealing with optimal recovery problems. The Gel'fand n-width is defined as follows.

DEFINITION 4.   X is a normed linear space and A is a closed, convex, centrally symmetric subset of X. The <u>Gel'fand n-width of</u> A <u>in</u> X is given by

$$d^n(A;X) = \inf_{L^n} \sup_{x\in A\cap L^n} ||x||,$$

where L^n varies over all subspaces of X of codimension n.

A subspace L^n is said to be of codimension n if there exist n linearly independent linear functionals $x_i^*\in X^*$ (the continuous dual of X) $i = 1,\ldots,n$, such that

$$L^n = \{ x : x_i^*(x) = 0, i = 1,\ldots,n \}.$$

There is no a priori relationship between $d_n(A;X)$ and $d^n(A;X)$. Either may be larger. Perhaps surprisingly they are often equal. There is a simple reason for this and it is that the inequalities $\delta_n(A;X) \ge d^n(A;X) \ge b_n(A;X)$ always hold. The inequality $\delta_n(A;X) \ge d^n(A;X)$ follows from the fact that every rank n continuous linear operator P_n may be written in the form

$$P_n x = \sum_{i=1}^n x_i^*(x)x_i$$

where $x_i^*\in X^*$, $i = 1,\ldots,n$, and the $\{x_i\}_1^n$ span the range of P_n. Thus

$$\sup_{\substack{x \in A}} ||x - P_n x|| \geq \sup_{\substack{x \in A \\ x_i^*(x)=0, \ i=1,\ldots,n}} ||x||$$

from which follows the desired inequality. The inequality $d^n(A;X) \geq b_n(A;X)$ is even simpler to prove. For every (n+1)-dimensional subspace X_{n+1} of X, and for every subspace L^n of codimension n of X, $X_{n+1} \cap L^n \neq \{0\}$ Thus if $\lambda S(X_{n+1}) \subseteq A$, then there exists an $x \in A \cap L^n$ for which $||x|| = \lambda$. Hence $d^n(A;X) \geq \lambda$, and the inequality follows. Thus in our previously considered example we necessarily have $d^n(B_\infty^{(1)};L^\infty) = 1/2n$.

5. N-WIDTHS OF SOBOLEV SPACES.

As stated earlier, the results of the previous two sections may be considered as particular cases of a more general theorem. We present this generalization and the remarks thereafter in an attempt to give the reader a taste for the type of research done in this subject.

The Sobolev space $W_p^{(r)}[0,1]$ for $p \in [1,\infty]$, r a positive integer, is defined as follows:

$$W_p^{(r)}[0,1] = \{f : f^{(r-1)} \text{ abs. cont.}, \ f^{(r)} \in L^p[0,1] \}.$$

Let $B_p^{(r)} = \{f : f \in W_p^{(r)}[0,1], \ ||f^{(r)}||_p \leq 1 \}$, and for a nonnegative integer m, set

$$x_+^m = \begin{cases} 0, & x < 0 \\ x^m, & x \geq 0. \end{cases}$$

THEOREM 7 [7]. Fix $p \in [1,\infty]$ and r a positive integer. Then for $n \geq r$,

$$d_n(B_p^{(r)};L^p) = d^n(B_p^{(r)};L^p) = \delta_n(B_p^{(r)};L^p) = b_n(B_p^{(r)};L^p)$$

Furthermore,

1) $X_n^* = \text{span} \{ 1, x, \ldots, x^{r-1}, (x-\xi_1)_+^{r-1}, \ldots, (x-\xi_{n-r})_+^{r-1} \}$ is an optimal subspace for $d_n(B_p^{(r)};L^p)$ for some choice of $0 < \xi_1 < \ldots < \xi_{n-r} < 1$.

2) $L^n = \{ f : f(\eta_i) = 0, \ i = 1,\ldots,n \}$ is an optimal subspace for $d^n(B_p^{(r)};L^p)$ for some choice of $0 < \eta_1 < \ldots < \eta_n < 1$.

3) The rank n linear operator P_n defined by interpolation from X_n^* to $f \in B_p^{(r)}$ at the $\{\eta_i\}_1^n$ is optimal for $\delta_n(B_p^{(r)};L^p)$.

It is also possible to identify an optimal subspace for $b_n(B_p^{(r)};L^p)$.

REMARK. The proof of Theorem 7 is far beyond the scope of this lecture. However to give an inkling as to how X_n^* arises, recall Taylor's formula with remainder in integral form. $B_p^{(r)}$ is the set of functions

$$f(x) = \sum_{i=0}^{r-1} a_i x^i + (1/(r-1)!) \int_0^1 (x-y)_+^{r-1} h(y) \, dy,$$

where $||h||_p \leq 1$ ($h = f^{(r)}$ and $a_i = f^{(i)}(0)/i!$, $i = 0,1,\ldots,r-1$). The subspace X_n^* is the span of the first r monomial terms (which must appear since there is no restriction on their coefficients) and the kernel $(x-y)_+^{r-1}$ evaluated at n-r distinct points.

REMARK. Theorem 7 is also valid in certain mixed norm cases. Consider the n-widths of $B_p^{(r)}$ in L^q, where p and q are arbitrary numbers in $[1,\infty]$. If $p = \infty$ or $q = 1$, then Theorem 7 holds (except that $b_n(B_p^{(r)};L^q)$ is unknown). It is conjectured that Theorem 7 is valid (except for b_n) for all $p \geq q$. For $p < q$, the situation is considerably more involved. No exact results are known and it was only some years ago that the asymptotic behaviour of each of the n-widths was determined. They do not all behave asymptotically in the same manner.

6. N-WIDTHS AS GENERALIZATIONS OF S-NUMBERS. Let T be a compact linear operator mapping X into itself. Set

$$A = \{ Tx : ||x|| \leq 1 \} .$$

The choice of A as the image of the unit ball under a linear map is a common choice in the theory of n-widths. $B_p^{(r)}$ is of this form, aside from the free r-dimensional polynomial subspace.

Assume for the moment that X = H is a Hilbert space with inner product $(\cdot,\cdot)$. Associated with T is its adjoint T^*. The compact maps T^*T and TT^* are self-adjoint and non-negative. They possess the same eigenvalues $\{\lambda_n(T)\}$, $n = 0,1,\ldots,$ (given in non-increasing order of magnitude) which are all non-negative numbers. The values $s_n(T) = [\lambda_n(T)]^{1/2}$ are called the s-numbers, or singular values, of T. Functional analysts who study n-widths generally regard them as generalizations of s-numbers, see e.g. Pietsch [5]. Let us explain why.

There exist many well-known characterizations for the s-numbers of T. Thus the "max-min" characterization is given by

$$s_n(T) = \left[\sup_{X_{n+1}} \inf_{x \in X_{n+1}} \frac{(T^*Tx , x)}{(x , x)} \right]^{1/2}$$

$$= \sup_{X_{n+1}} \inf_{x \in X_{n+1}} \frac{||Tx||}{||x||}$$

where X_{n+1} varies over all subspaces of H of dimension n+1. But this last quantity is simply a restatement of the definition of the Bernstein n-width $b_n(A;H)$ for this choice of A. Thus $b_n(A;H) = s_n(T)$. In a totally analogous manner, it may be seen that the classical "min-max" characterization of $s_n(T)$ given by

$$s_n(T) = \left[\inf_{L^n} \sup_{x \in L^n} \frac{(T^*Tx, x)}{(x, x)} \right]^{1/2}$$

where the L^n vary over all subspaces of H of codimension n, is essentially the Gel'fand n-width $d^n(A;H)$. In this same vein $\delta_n(A;H) = s_n(T)$, since the definition of $\delta_n(A;H)$ corresponds to the classical singular value decomposition of T (and $d_n(A;H) = \delta_n(A;H)$ since we are in a Hilbert space). Thus $d_n(A;H) = d^n(A;H) = \delta_n(A;H) = b_n(A;H) = s_n(T)$.

B. OPTIMAL RECOVERY

7.   INTRODUCTION. Let X be a normed linear space and A a subset of X. In a very general sense optimal recovery is concerned with the problem of estimating, in as efficient a manner as possible, some specific information about elements of A based on a number of given pieces of information.

Many problems fall into this wide setting. Before presenting a general framework, let us consider some specific examples.

8.   EXAMPLES.

8.A. RECOVERY OF A FUNCTIONAL. Let $X = L^\infty[0,1]$ and $A = B_\infty^{(1)}$ (see Section 3). Assume that for each $f \in B_\infty^{(1)}$ we are given the values $f(x_i)$, $i = 1,\ldots,n$, for some fixed $0 \le x_1 < \ldots < x_n \le 1$. For convenience we set $I(f) = (f(x_1),\ldots,f(x_n)) \in R^n$, and call $I(f)$ the _information vector_. Let $y \in [0,1]$, fixed. The problem we consider is that of optimally reconstructing $f(y)$, for $f \in B_\infty^{(1)}$, based only on the data $I(f)$.

Any function T which maps $I(f)$ to R is called an _algorithm_. The error of the algorithm T is defined by

$$E(T) = \sup \{ |f(y) - T(I(f))| : f \in B_\infty^{(1)} \}.$$

The value

$$E^* = \inf \{ E(T) : T \}$$

is the _intrinsic error_ in our problem. If $E^* = E(T^*)$ for some algorithm T^*, then we say that T^* is an _optimal algorithm_ or provides for an optimal recovery of $f(y)$. The problem is to find E^* and an optimal algorithm T^*.

An important tool in the solution of this problem is the following simple lower bound for E^*.

PROPOSITION 8. $E^* \ge \sup \{ |f(y)| : f \in B_\infty^{(1)}, I(f) = \underline{0} \}$.

PROOF.   Let $f \in B_\infty^{(1)}$ with $I(f) = \underline{0}$. Since $-f \in B_\infty^{(1)}$ and $I(-f) = \underline{0}$, it follows that for every algorithm T,

$$E(T) \geq \max \{ |f(y) - T(\underline{0})|, |-f(y) - T(\underline{0})| \}$$

$$\geq (|f(y) - T(\underline{0})| + |f(y) + T(\underline{0})|)/2$$

$$\geq |f(y)|.$$

The claim now follows. □

We will prove that equality holds, calculate E^*, and identify an optimal algorithm.

PROPOSITION 9. $E^* = \min \{ |y - x_i|: i = 1,\ldots,n \}$. Furthermore, if $|y - x_j| = \min \{ |y - x_i|: i = 1,\ldots,n \}$, then the algorithm T_y defined by $T_y(I(f)) = f(x_j)$ is optimal.

PROOF. The function $f^*(x) = \min \{ |x - x_i|: i = 1,\ldots,n \}$ is in $B_\infty^{(1)}$ and satisfies $f^*(x_i) = 0$, $i = 1,\ldots,n$. Thus from Proposition 8, $E^* \geq f^*(y) = \min \{ |y - x_i|: i = 1,\ldots,n \}$. For T_y as above,

$$E^* \leq E(T_y) = \sup \{ |f(y) - f(x_j)|: f \in B_\infty^{(1)} \}$$

$$= |y - x_j|$$

$$\leq E^*. □$$

8.B. RECOVERY OF A FUNCTION. Within this same framework, we change our example somewhat. Assume that based on the same information vector $I(f)$, we are now interested in recovering not $f(y)$, but the full function f on $[0,1]$. Our algorithms T are therefore functions from R^n to $L^\infty[0,1]$. As previously

$$E(T) = \sup \{ ||f - T(I(f))||_\infty : f \in B_\infty^{(1)} \},$$

and $E^* = \inf \{ E(T) : T \}$.

Totally analogous to Proposition 8, we have

PROPOSITION 10. $E^* \geq \sup \{ ||f||_\infty : f \in B_\infty^{(1)}, I(f) = \underline{0} \}$.

Thus, in particular, $E^* \geq ||f^*||_\infty$ where f^* is as defined in the proof of Proposition 9. We will prove equality. To this end set $z_i = (x_i + x_{i+1})/2$, $i = 1,\ldots,n-1$, ($z_0 = 0$, $z_n = 1$). Let S_n denote the space of step functions with jumps at $\{z_i\}_1^{n-1}$. For each $f \in B_\infty^{(1)}$, define $s_f \in S_n$ by $s_f(x_i) = f(x_i)$, $i = 1,\ldots,n$.

PROPOSITION 11. $E^* = ||f^*||_\infty$. Furthermore, if T^* is the algorithm given by $T^*(I(f)) = s_f$, then T^* is an optimal algorithm.

PROOF. For T^* as above,

$$E^* \leq E(T^*) = \sup \{ ||f - s_f||_\infty: f \in B_\infty^{(1)} \}.$$

For $x \in (z_{i-1}, z_i]$, $|f(x) - s_f(x)| = |f(x) - f(x_i)| \leq |x - x_i| = f^*(x)$. Thus

$$\hat{E}^* \leq E(T^*) \leq ||f^*||_\infty \leq E^*. \quad \square$$

8.C. OPTIMAL INFORMATION. We again alter our problem. We assume that we wish, as in (8.B), to recover $f \in B_\infty^{(1)}$ based on $I(f)$. However, now we may choose, a priori, the n information functionals which constitute $I(f)$. For ease of discussion, let us assume that we may choose n points $x_1,\ldots,x_n$ in $[0,1]$, which appear in $I(f)$, at which to sample f.

We exhibit this dependence on I by letting $E(T,I)$ and $E^*(I)$ denote the $E(T)$ and E^* of (8.B). We are therefore concerned with the problem of evaluating

$$\tilde{E} = \inf_I E^*(I),$$

where I ranges over all information vectors of the form $I(f) = (f(x_1),\ldots,f(x_n))$ with $0 \leq x_1 < \ldots < x_n \leq 1$. Set $\underline{x} = (x_1,\ldots,x_n)$, and let $f^*(x,\underline{x})$ denote the $f^*(x)$ of (8.B). It is now a simple matter to prove

PROPOSITION 12. $\tilde{E} = \inf\{ ||f^*(\cdot,\underline{x})||_\infty : \underline{x}\} = 1/2n$. Furthermore, an optimal $\underline{x}$ is given by $\underline{x}^* = (1/2n, 3/2n, \ldots, (2n-1)/2n)$.

Before continuing note that the value obtained is exactly the value of the n-widths of $B_\infty^{(1)}$ in L^∞. This is not a coincidence. We will discuss the connection in the next section.

8.D. OPTIMAL RECOVERY WITH ERROR. We now return to the problem, considered in (8A), of recovering $f(y)$, for f $B_\infty^{(1)}$, based on $I(f) = (f(x_1),\ldots,f(x_n))$ for fixed $0 \leq x_1 < \ldots < x_n \leq 1$. However, let us assume that we do not know $I(f)$ exactly. Errors may occur in our calculation and rather than being given $I(f)$, we are given $\underline{w} = (w_1,\ldots,w_n)$ where $|w_i - f(x_i)| \leq \varepsilon_i$, $i = 1,\ldots,n$. The error bounds $\varepsilon_i \geq 0$ are given fixed values. We therefore define, for an algorithm T mapping R^n to R,

$$E(T;\underline{\varepsilon}) = \sup\{ |f(y) - T(\underline{w})| : f \in B_\infty^{(1)}, |w_i - f(x_i)| \leq \varepsilon_i, i = 1,\ldots,n \},$$

and $E^*(\underline{\varepsilon}) = \inf\{ E(T;\underline{\varepsilon}) : T \}$.

This next result is totally analogous to Proposition 8.

PROPOSITION 13. $E^*(\underline{\varepsilon}) \geq \sup\{ |f(y)| : f \in B_\infty^{(1)}, |f(x_i)| \leq \varepsilon_i, i = 1,\ldots,n \}$.

Set $f^*(x;\underline{\varepsilon}) = \min\{ \varepsilon_i + |x - x_i| : i = 1,\ldots,n \}$. Then $f^*(\cdot,\underline{\varepsilon}) \in B_\infty^{(1)}$ and $0 \leq f^*(x_i;\underline{\varepsilon}) \leq \varepsilon_i$, $i = 1,\ldots,n$. Let $\varepsilon_j + |y - x_j| = \min\{ \varepsilon_i + |y - x_i| : i=1,\ldots,n\}$.

PROPOSITION 14. $E^*(\underline{\varepsilon}) = f^*(y;\underline{\varepsilon})$, and the algorithm defined by $T_y(\underline{w}) = w_j$ is optimal.

PROOF. From Proposition 13, $E^*(\underline{\varepsilon}) \geq f^*(y;\underline{\varepsilon})$. Let $f \in B_\infty^{(1)}$ with $|w_i - f(x_i)| \leq \varepsilon_i$, $i = 1,\ldots,n$. Then

$$|f(y) - T_y(\underline{w})| = |f(y) - w_j|$$
$$\leq |f(y) - f(x_j)| + |f(x_j) - w_j|$$
$$\leq |y - x_j| + \varepsilon_j$$
$$= f^*(y;\underline{\varepsilon}).$$

Thus $E^*(\underline{\varepsilon}) \leq E(T_y;\underline{\varepsilon}) \leq f^*(y;\underline{\varepsilon}).$ □

Analogues, with error, of examples (8B) and (8C) are similarly constructed.

9. GENERAL THEORY. The above simple examples are prototypes of some of the problems considered in the theory of optimal recovery. In our general discussion we will somewhat restrict ourselves. Thus, for example, all our operators will be linear, and we will not touch upon problems of recovery with error as exemplified by example (8D).

Let X, Y and Z be normed linear spaces. A is a subset of X which we assume to be closed, convex, and centrally symmetric. By U we denote a linear operator from X to Z. U(x), for x∈A, will be the element which we wish to recover, and U is therefore termed the _object_ operator. I is a linear operator from X to Y, called the _information_ operator. Any function T from I(A) to Z is said to be an _algorithm_. Each algorithm gives rise to a recovery scheme with error

$$E(T) = \sup \{ ||U(x) - T(I(x))||: x \in A \}.$$

The value $E^* = \inf \{ E(T) : T \}$ where T ranges over all possible algorithms, is called the _intrinsic error_ of the process. If $E^* = E(T^*)$ for a specific algorithm T^*, then T^* is called an optimal algorithm and we have found an _optimal recovery_ for U on A.

There is no reason to suppose that optimal algorithms exist, or if they do, are linear. To obtain such results, additional assumptions are needed. However, certain very general properties do hold. As an analogue of Proposition 8 we have the following.

PROPOSITION 15. $E^* \geq \sup \{ ||U(x)||: x \in A, I(x) = 0 \}.$

PROOF. Let x∈A and I(x) = 0. By assumption -x∈A and I(-x) = 0. Thus for every algorithm T,

$$||U(x) - T(0)||, ||U(-x) - T(0)|| \leq E(T).$$

Since U is linear, it follows that $||U(x)|| \leq E(T)$, proving the proposition. □

For ease of exposition, set

$$e^* = \sup \{ ||U(x)||: x \in A, I(x) = 0 \}.$$

In all our examples we had $E^* = e*$. However, this is not generally valid as the following simple example shows.

EXAMPLE. Let X be R^3 endowed with the Euclidean norm $||\underline{x}||_2 = (x_1^2 + x_2^2 + x_3^2)^{1/2}$. Set $Z = X$, $Y = R$, $U(\underline{x}) = \underline{x}$,

$$A = \{ \underline{x} : ||\underline{x}||_1 = |x_1| + |x_2| + |x_3| \leq 1 \}$$

and $I(\underline{x}) = x_1 + x_2 + x_3$. A simple calculation shows that $e* = 1/\sqrt{2}$. Now

$$E^* = \inf_T \sup \{ ||\underline{x} - T(I(\underline{x}))||_2 : ||\underline{x}||_1 \leq 1 \}$$

$$\geq \inf_T \max \{ ||\underline{e}^i - T(1)||_2 : i = 1,2,3 \},$$

where $\underline{e}^i$ is the $i\underline{th}$ unit vector, $i = 1,2,3$. $T(1)$ is a vector in R^3. No vector in R^3 is of distance less that $\sqrt{2/3}$ from each of the $\underline{e}^i$, $i = 1,2,3$. Thus $E^* \geq \sqrt{2/3}$. (Equality in fact holds for the algorithm $T(a) = (a/3,a/3,a/3)$.)

An opposite inequality to $E^* \geq e*$ is the following.

THEOREM 16 [3,p.3]. $E^* \leq 2e*$.

PROOF. For each $y \in I(A)$, choose an $x'(y) \in A$ satisfying $I(x'(y)) = y$. We define an algorithm T' by

$$T'(I(x)) = U(x'(I(x))).$$

Then

$$E(T') = \sup \{ ||Ux - T'(I(x))|| : x \in A \}$$

$$= \sup \{ ||U(x - x'(I(x)))|| : x \in A \}.$$

Set $w = x - x'(I(x))$. Then $I(w) = I(x) - I(x) = 0$, and since A is convex and centrally symmetric, $w/2 \in A$. Thus

$$E^* \leq E(T') \leq 2 \sup \{ ||Uw|| : w \in A, I(w) = 0 \} = 2e*. \quad \square$$

REMARK. The T' constructed above is, in general, neither continuous nor linear. If we demand a linear, continuous algorithm, them no inequality of the above form is valid.

One set of assumptions which implies the equality $E^* = e*$ is the following. (Note that our examples do not quite satisfy these assumptions.)

THEOREM 17 [3,p.5]. Assume that there exists a function S from I(A) to X for which $x - S(I(x)) \in A$, and $I(x - S(I(x))) = 0$ for every $x \in A$. Then $E^* = e*$ and $T = US$ is an optimal algorithm.

The proof of this theorem is an immediate consequence of the definitions of T and e*.

A totally different set of restrictions gives us similar results.

Assume that X is a normed linear space over the reals and Z = R, i.e. U is a linear functional. In addition to the previous assumptions, we also suppose that I(A) is absorbing in Y, i.e. for any $y \in Y$ there exists $\lambda > 0$ such that $\lambda y \in I(A)$. Let Y^* denote the continuous dual of Y. Then

THEOREM 18 [3,p.16]. <u>Under the above assumptions,</u>

$$e^* = E^* = \inf_{T \in Y^*} \sup \{ |U(x) - T(I(x))| : x \in A \}.$$

<u>Furthermore, if</u> I(A) <u>is a neighborhood of the origin in</u> Y, <u>and if there exists a</u> $T \in Y^*$ <u>for which</u>

$$\sup \{ |U(x) - T(I(x))| : x \in A \} < \infty,$$

<u>then there exists an optimal algorithm in</u> Y^*, <u>i.e. one which is linear and continuous.</u>

We close these notes by examining a connection between certain n-widths and problems of optimal recovery with optimal information. For convenience, we now assume that X = Z, $Y = R^n$, U(x) = x, and $I(x) = (x_1^*(x),\ldots,x_n^*(x))$, where $x_i^* \in X^*$ (the continuous dual of X), $i = 1,\ldots,n$. Thus for each algorithm T,

$$E(T) \geq E^* \geq e^* = \sup \{ ||x|| : x \in A, I(x) = \underline{0} \}.$$

Set $L^n = \{ x : I(x) = \underline{0} \}$. Then L^n is a subspace of codimension at most n, and we may write

$$e^* = \sup \{ ||x|| : x \in A \cap L^n \}.$$

From the definition of the Gel'fand n-width $d^n(A;X)$, it follows that $e^* \geq d^n(A;X)$ for any choice of n continuous linear information functionals. In particular

$$\inf_{L^n} E^* \geq d^n(A;X).$$

Let us also recall the definition of the linear n-width $\delta_n(A;X)$.

$$\delta_n(A;X) = \inf_{P_n} \sup_{x \in A} ||x - P_n x||$$

where P_n is an operator of the form $P_n x = \sum_{i=1}^{n} x_i^*(x) \, x_i$. Taking the infimum over P_n is equivalent to taking the infimum over $\{x_i^*\}_1^n$ and $\{x_i\}_1^n$. If we fix the $\{x_i\}_1^n$ and take the infimum over the $\{x_i^*\}_1^n$, then we are searching for a best continuous linear approximation to A from span $\{ x_1,\ldots,x_n \}$. On reversing the process by fixing the $\{ x_i^* \}_1^n$ and taking the infimum over the $\{x_i\}_1^n$, we are searching for the optimal continuous linear algorithm based on the information $\{x_i^*\}_1^n$. As such,

$$\delta_n(A;X) \geq \inf \{ E^* : L^n \}.$$

Thus the two n-widths δ_n and d^n are upper and lower bounds, respectively, in the problem of optimal recovery of A with n optimal linear continuous pieces of information. If $\delta_n(A;X) = d^n(A;X)$, and P_n^* is optimal for $\delta_n(A;X)$, then P_n^* gives rise to a continuous linear optimal algorithm for this problem (see e.g. Theorem 7).

BIBLIOGRAPHY

1. Kolmogoroff, A., "Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse", Annals of Math., 37 (1936), 107-110.

2. Krein, M.G., Krasnosel'ski, M.A., Milman, D.P., "On deficiency numbers of linear operators in Banach spaces and on some geometric problems", Sb. Trudov Inst. Mat. Akad. Nauk SSSR, 11 (1948), 97-112.

3. 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 Press, New York, 1977, 1-54.

4. Micchelli, C.A., Rivlin, T.J., "Lectures on optimal recovery", preprint.

5. Pietsch, A., Nuclear Locally Convex Spaces, Springer-Verlag, Berlin, 1972.

6. Pinkus, A., n-Widths in Approximation Theory, Springer-Verlag, Berlin, 1985.

7. Pinkus, A., "n-Widths of Sobolev spaces in L^p", Constr. Approx., 1 (1985), 15-62.

8. Rivlin, T.J., An Introduction to the Approximation of Functions, Blaisdell, Waltham, Mass., 1969.

9. Traub, J.F., Wozniakowski, H., A General Theory of Optimal Algorithms, Academic Press, New York, 1980.

DEPARTMENT OF MATHEMATICS
TECHNION
HAIFA, ISRAEL