MATHEMATICAL PROBLEMS IN COMPUTATION THEORY
BANACH CENTER PUBLICATIONS, VOLUME 21
PWN – POLISH SCIENTIFIC PUBLISHERS
WARSAW 1988

THE EXTENT TO WHICH LINEAR PROBLEMS HAVE LINEAR OPTIMAL ALGORITHMS

EDWARD W. PACKEL

Department of Mathematics and Computer Studies, Lake Forest College, Lake Forest, U.S.A.

1. Introduction

The question raised in the title has both a crisp negative answer (there are counterexamples) and a surprising variety of positive ones (theorems). The general setting in which the question can be meaningfully posed involves a relatively recent research area on the interface between mathematics and theoretical computer science. This area, which we shall refer to as *information-based complexity*, is known also under aliases such as the general theory of optimal algorithms and the information-centered approach to algorithms. Information-based complexity seeks to create a general theory about problems with partial or approximate information and to apply the results to specific problems. Some of the disciplines where problems with such incomplete information have arisen include computer science, economics, control theory, signal processing, and geophysics.

The goal of this article is to present a selective introduction to the field of information-based complexity and to describe the seesaw nature of results in the last decade on the extent to which linear problems have linear optimal algorithms. The approach we take will certainly not do justice to the breadth of ideas, techniques, and applications that are coming into play in the general theory. Instead, we will take a direct route to the results on linear optimal algorithm, pausing occasionally to survey the pleasing way in which ideas from linear algebra and (functional) analysis are used. In the final section we mention some other theoretical and applied topics currently under investigation in the theory of optimal algorithms. A more general expository overview can be found in Traub and Woźniakowski [14] and a somewhat more technical survey is offered in a recent paper by Woźniakowski [20].

2. Information-based complexity for linear problems

A thorough development of the framework for information-based complexity may be found in Traub and Woźniakowski [13]. Here we present the standard approach for *linear* problems.

Let F_1 and F_2 be normed linear spaces over the scalar field K, where K is either the real or complex numbers. The *problem elements* for which we have only partial or approximate information are assumed to belong to a convex subset F_0 of F_1. We further assume that F_0 is *balanced* ($f \in F_0$ and $|\alpha| = 1 \Rightarrow \alpha f \in F_0$). An important way of generating such an F_0 (and in a sense the "only" way) is to let $F_0 = \{f \in F_1 \colon \|T(f)\| \leqslant 1\}$, where $T\colon F_1 \to F_4$ is a linear *restriction operator* into another normed linear space F_4.

A *linear problem* is defined by specifying F_0 along with two linear operators on F_1. The underlying problem solution (for which we seek approximating algorithms) is defined by a linear *solution operator* $S\colon F_1 \to F_2$. As was suggested in the introduction, an algorithm to approximate $S(f)$ may not have full information about f to work with. Accordingly, we specify a linear *information operator* $N\colon F_1 \to K^n$, so that $N(f)$ gives us n (a positive integer) scalar pieces of information about the problem element f. The requirement that N has finite dimensional range is a concession to the reality that practical algorithms process only a finite amount of information. We will often be concerned only with the behavior of S and N on the convex and balanced subset F_0 of F_1. In this case, we regard an operator defined on F_0 as linear if it is the restriction of a linear operator defined on F_1.

A concept central to the whole theory is the *radius of information*, which gives the inherent error in approximating the solution $S(f)$ given only the information provided by $N(f)$. We concentrate throughout this paper on the "worst case error" setting, where the radius of information, $r(S, N)$, can be defined as follows:

Given $f \in F_0$, set $V(f) = \{g \in F_0 \colon N(g) = N(f)\}$. When we are concentrating on an information value y with $y = N(f)$, we will use the fact that $V(f) = N^{-1}(y) \cap F_0$.

Now define $r(S, N, f)$ to be the radius of $S(V(f))$ as a subset of F_2, where the *radius* of a subset G of a normed space F is defined by $\inf\limits_{c \in F} \sup\limits_{g \in G} \|c - g\|$. A point c (if it exists) for which this infimum is attained is called a *center* of G (it need not be unique).

Finally, define $r(S, N) = \sup\{r(S, N, f) \colon f \in F_0\}$.

The key idea here is that we do not know the specific f to which the solution operator should be applied. The best we can say is that it belongs to the set $V(f)$ of problem elements which share the same information as f. Accordingly, a bound on the possible error in approximating $S(f)$ is given by the radius of the set $S(V(f))$. The radius of information for the whole problem is the "largest" such radius for f in F_0.

Closely related to this radius is the *diameter of information*, $d(S, N)$, defined by $\sup_{f \in F_0} d(S, N, f)$ where $d(S, N, f)$ is the diameter of the set $S(V(f))$ defined above. Many results in the theory are more easily developed in terms of the diameter of information, which is not generally twice the radius, but can easily be shown to satisfy $r(S, N) \leqslant d(S, N) \leqslant 2r(S, N)$.

We now investigate algorithms to approximate $S(f)$. Since we only have the information $N(f)$ on f, such algorithms can only be defined on $N(F_0)$. Accordingly, we define an *algorithm* φ for the problem determined by F_0, S and N as an operator φ: $N(F_0) \to F_2$. Of obvious importance are the *optimal algorithms*, which we now define.

If we define the *error* of an algorithm φ by $e(\varphi) = \sup\{\|\varphi(N(f)) - S(f)\|: f \in F_0\}$, it follows directly that $e(\varphi) \geqslant r(S, N)$. An algorithm φ^* is defined to be *optimal* if $e(\varphi^*) = \inf\{e(\varphi)$: φ an algorithm for S and $N\}$. It follows directly that φ^* is optimal if and only if $e(\varphi^*) = r(N, S)$. Thus $r(S, N)$ provides a tight lower bound on the worst case error and can be viewed as the intrinsic uncertainty inherent in the problem with solution S and information N. If an algorithm φ satisfies the stronger condition that $\|\varphi(N(f)) - S(f)\| \leqslant r(S, N, f) \, \forall f \in F_0$, then φ is said to be *strongly optimal* or *central*. Generally it is asking too much to expect realizable algorithms which are strongly optimal, but we shall encounter one important case where this occurs. In Fig. 1 we summarize the situation as we have described it to this point. The fact that the problem domain F_0 may arise as the inverse image of a unit ball under a restriction operator T is incorporated in the diagram since we shall need this formulation at one important point.

Our restriction to problems which are linear is natural in view of the goals of this paper. While this does rule out nonlinear problems (root finding, for example), many of the important problems for which algorithms are sought turn out to be linear. In particular, classical problems of integration, linear differential and integral equations, approximation and interpolation are linear, as are many problems stemming from linear models in computer science and the other fields.

To help fix some of the above ideas, we now review two basic results for the linear case.

THEOREM 1. *Given linear operators* S: $F_1 \to F_2$, N: $F_1 \to K^n$, *and* T: $F_1 \to F_4$ *with* $F_0 = \{f \in F_1: \|T(f)\| \leqslant 1\}$, then

(a) $r(S, N) < \infty \Rightarrow \ker(N) \cap \ker(T) \subseteq \ker(S)$, *and*

(b) $d(S, N) = 2 \sup\{\|S(h)\|: h \in \ker(N) \cap F_0\}$.

Proof. (a) Suppose, on the contrary, $\exists\, h \in \ker(N) \cap \ker(T)$ with $S(h) \neq 0$. Then $\alpha h \in V(0) = \ker(N) \cap F_0 \, \forall \alpha \in K$. It follows that $S(V(0))$ is unbounded in

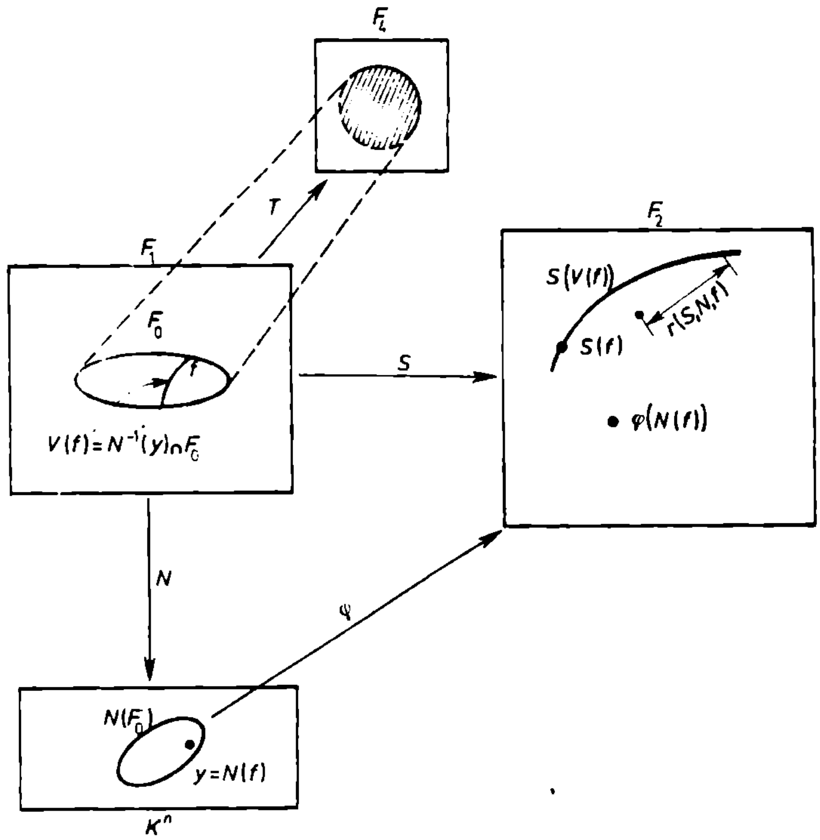

Fig. 1. Schematic view of the general theory

F_2 and hence has infinite radius, so $r(S, T) = \infty$. This contradiction establihes the result.

(b) Given any $f \in F_0$, and any $g_1, g_2 \in V(f)$, set $h = (g_1 - g_2)/2$. Then $\|T(h)\| \leqslant 1$ and $h \in \ker(N)$, so $h \in V(0)$. Thus $\|S(g_1) - S(g_2)\| = 2\|S(h)\| \leqslant 2\sup\{\|S(h)\|: h \in \ker(N) \cap F_0\}$. Taking *sups* first over g_1 and g_2 in $V(f)$ and then over $f \in F_0$, we get $d(S, N) \leqslant 2\sup\{\|S(h)\|: h \in \ker(N) \cap F_0\}$. The reverse inequality follows by noting that for any $h \in V(0)$ we also have $-h \in V(0)$ (F_0 is balanced). Thus $2\|S(h)\| = \|S(h) - S(-h)\| \leqslant d(S, N)$ and, since $h \in V(0)$ was arbitrary, the proof is complete.

Since problems with an infinite radius of information are of no great interest to us (there will always be infinite error yet every algorithm is optimal!), we henceforth assume $r(S, N) < \infty$. As a consequence of the above theorem, we will then always have $\ker(N) \cap \ker(T) \subseteq \ker(S)$.

For many of the same theoretical reasons that we concentrate on linear problems, we now restrict ourselves to the study of algorithms which are also linear. Such algorithms are desirable for a variety of practical reasons as well, which we briefly summarize. Linear problems appear to be natural for problems in a linear setting. Indeed, many of the standard algorithms for classical numerical problems (integration and interpolation for example) are linear. Linear algorithms tend to be simpler and easier to implement. Most importantly, linear algorithms have small combinatorial complexity and optimal linear algorithms can be formally shown to have *nearly optimal combinatorial complexity*. In addition to this valuable efficiency in time, linear algorithms also have small space complexity (if we ignore precomputations). Details can be found in Traub and Woźniakowski [13], Chapter 5. Finally, we note that, in some important classical works in numerical analysis, linearity has been an assumption imposed on a class of algorithms being studied. In the light of results we will present on the existence of optimal linear algorithms, it turns out that this assumption is not necessary. Indeed, as far as optimality is concerned, nonlinear algorithms do not help to reduce the error in these and many other important cases.

3. An example

There are many standard problems that fit neatly into the framework we have described. We mention a few of them briefly, saving our detailed development for a more obscure but potentially enlightening example. This example will also serve as a starting point for a subsequent counterexample.

Often F_1 will be a space of functions. Then F_0 might be its unit ball or perhaps the convex balanced set generated by differentiation as a restriction operator. Thus, with $F_1 = C([0, 1])$ (continuous, real-valued functions on $[0, 1]$) and $F_2 = \mathbf{R}$, $S(f)$ could be the definite integral of f, $S(f) = \int_0^1 f(t)\,dt$. If the information $N(f)$ consists of n function evaluations $f(t_1), \ldots, f(t_n)$, the natural algorithms to consider include the various integration schemes studied in a numerical analysis course. Which of these algorithms might be optimal depends upon F_0 and the way the evaluation points t_i are chosen. Based upon a result to be presented subsequently, we can be assured that there will be an optimal algorithm which depends linearly upon the function evaluations $f(t_i)$.

To capture the idea of root finding, we might choose F_0 to be the set of real functions $f \in C([0, 1])$ with $f(0)f(1) \leqslant 0$. The nonlinear solution operator S simply maps f into its set of zeros. The observant reader will notice that the codomain F_2 in this nonlinear case is not a linear space, but rather the power set of $[0, 1]$. Indeed, the whole theory can be generalized (with some inevitable loss of structure and power) to a surprising degree (see [15]). Having made our apologies for going beyond our established framework, we forge ahead by putting a pseudo-metric on F_2 by defining the distance between two nonempty subsets $X, Y \subseteq [0, 1]$ by $d(X, Y) = \inf\{\|x - y\|\colon x \in X, y \in Y\}$. If we now restrict ourselves to algorithms that are single-valued, we have captured the idea of seeking a real number that is "close" to *some* zero of f. Definitions of radius of information and optimality now go through with norm replaced by pseudo-metric.

Assume that information for this problem is provided by fixing n points $t_1, \ldots, t_n$ on [0, 1] and evaluating each problem element f at these points. It is easy to see that by judiciously using only the signs of the $f(t_i)$ and the intermediate value theorem, we can approximate a root for any f to within the radius of the maximum subinterval determined by the t_i. By choosing the n points equally spaced, a minimal radius of information results. Since we know that $f(0)f(1) \leqslant 0$, only one of the endpoints needs evaluation and this minimal radius turns out to be $1/(2n)$. Readers familiar with a bisection algorithm for root finding will not be overly impressed by this result, but recall that our evaluation points were assumed to be independent of the problem elements f. In the final section we briefly discuss the consequences of relaxing this restriction.

Having touched on two familiar types of problems, we now look at a more detailed example. Let both F_1 and F_2 be the space $C([0, 1])$ of continuous real-valued functions on [0, 1] endowed with the sup-norm. Let F_0 be the unit ball $\{f \in C([0, 1]): \|f\| \leqslant 1\}$. We take as our solution operator the indefinite integral $S: F_0 \to C[0, 1]$ defined by $S(f)(x) = \int_0^x f(t)\,dt$. The information for $f \in F_0$ consists of the real number $N(f) = \int_0^1 f(t)\,dt$. Thus our problem is, essentially, to approximate antiderivatives of certain functions defined on [0, 1] knowing only their definite integrals. We offer no testimonials to the relevance of this problem in any meaningful applied or theoretical setting, but it is of about the right subtlety to shed light on the concepts of radius of information and optimal algorithm. We now proceed to calculate these, introducing in the process some geometrical ideas which will be used again later.

For each fixed $x \in [0, 1]$, consider the simpler problem with the same information N and the real-valued solution operator S_x on F_0 defined by $S_x(f) = S(f)(x)$. The subset of the plane defined by $Y_x = \{(S_x(f), N(f)): f \in F_0\}$ is convex and balanced (this depends only on F_0 having these properties and on the linearity of the operators involved as suggested in Fig. 2a). There must then exist bounding parallel hyperplanes (in this case lines) tangent to Y_x at the symmetric points $(-r_x, 0)$ and $(r_x, 0)$. Such a pair of hyperplanes need not be unique (see Fig. 2b). This geometrical situation enables us to draw the following conclusions:

For each x, $r(S_x, N) = r_x$. Indeed, the subset $S_x(V(f))$ with maximum radius occurs when $N(f) = 0$. It is the image under S_x of $\ker(N) \cap F_0$.

The radius of information $r(S, N)$ for the indefinite integral S we are working with is $\frac{1}{2}$, realized as the maximum r_x value. This maximum occurs when $x = \frac{1}{2}$ (see Fig. 2c).

For each x, a *linear* optimal algorithm for the problem consisting of S_x and N is determined by the slope of a bounding hyperplane tangent to Y_x at $(r_x, 0)$. Specifically, if the parallel hyperplanes have equations $t = \mu y \pm r_x$, then any $(S_x(f), N(f)) \in Y_x$ must satisfy $S_x(f) \leqslant \mu N(f) + r_x$ and $S_x(f) \geqslant \mu N(f) - r_x$. Hence $|\mu N(f) - S_x(f)| \leqslant r_x$ and the linear algorithm $\varphi_x(N(f)) = \mu N(f)$ is optimal for S_x.

If we now define $\varphi(N(f))(x) = \varphi_x(N(f))$, we do get a function φ that is linear in $y = N(f)$; but $\varphi(y)$ is not (unless $y = 0$) a continuous function of x. As Fig. 2c indicates there must be a discontinuity at $x = \frac{1}{2}$. A linear optimal algorithm can be salvaged in this example, however, since we can find a single bounding tangent hyperplane at $(\frac{1}{2}, 0)$ (indicated by the dashed lines in Fig. 2c) for the *union* of all the sets Y_x, $x \in [0, 1]$. This generates a linear algorithm which gives functions which are continuous (even constant) in x. The optimal algorithm φ^* that results is $\varphi^*(y)(x) = \frac{1}{2} y \ \forall x$. In terms of earlier definitions, optimality says

$$\left|\tfrac{1}{2}\int_0^1 f - \int_0^x f\right| \leqslant \tfrac{1}{2} \qquad \forall f \in F_0,\ \forall x \in [0, 1].$$

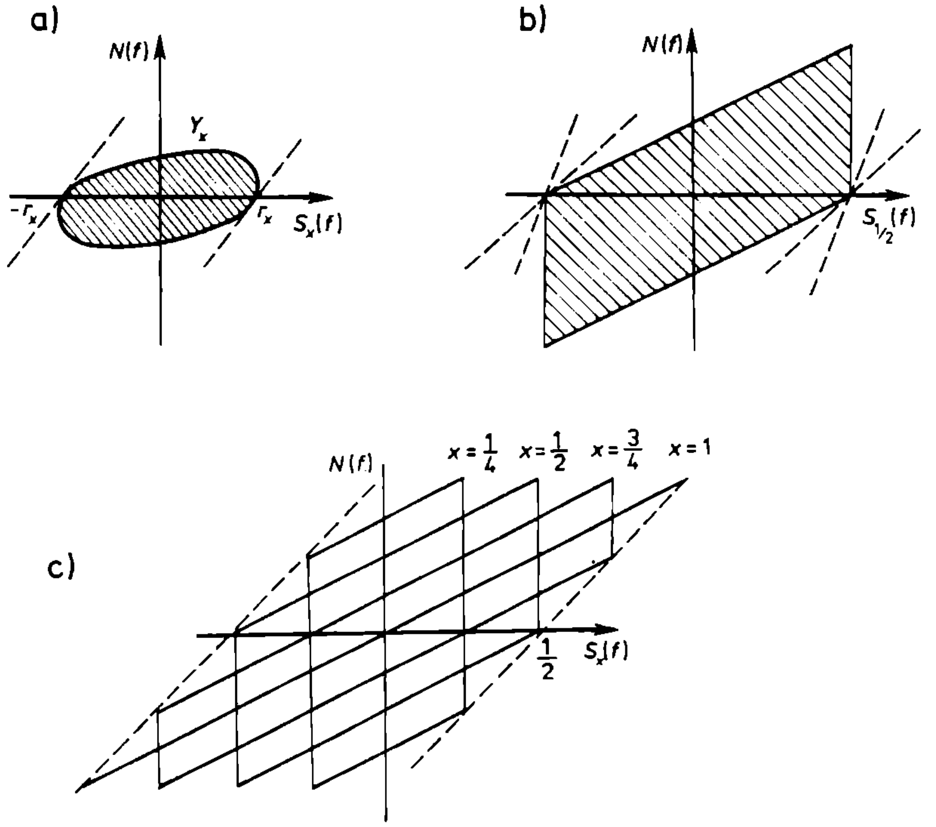

Fig. 2. Plots of $\{(S_x(f), N(f)) \colon f \in F_0\} = Y_x$
a) General picture
b) Nonuniqueness when $x = \frac{1}{2}$
c) Y_x for $x = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$

We conclude that the problem we have presented does have a linear optimal algorithm, but it was a close contest. This closeness will be exploited in a later section. We now survey the brief but interesting history of the general question of when such algorithms exist for linear problems.

4. Linear optimal algorithms for linear problems

Table 1 summarizes positive and negative results on existence of linear optimal algorithms. We will discuss the items in the table essentially in the chronological order listed. Before proceeding we mention a small bit of "prehistory". Unitl very recently it has been the belief of those working in the area that, despite the mounting evidence of counterexamples and the absence of a reasonably general theorem, realworld or naturally occurring linear problems *do* have linear optimal algorithms. Support for this "folk" result has been abundant, bolstered by classical numerical algorithms and some more recent applications of the work in information-based complexity. We will not supply specific examples here, but we shall see that recent developments have shaken this belief somewhat.

Results 1 and 2 cover the important special case of problems whose solution operator and approximating algorithms require a single scalar answer for each problem element. The discussion of the Example in the previous section gives a clear idea of the lovely proof of Result 1. Briefly, the same tangent hyperplane (separation theorem) arguments are applied to the balanced convex set $Y = \{(S(f), N(f)) : f \in F_0\}$ in $\boldsymbol{R}^{n+1}$. As in our Example, a linear optimal algorithm emerges neatly from the inequalities generated by the tangent hyperplanes. See [13], Chapter 3, for full details.

Result 3 extends Result 1 to the important case where there is error or uncertainty in the information. Since the general theory is motivated by the increased realism gained in admitting limited knowledge of a problem element f through its information $N(f)$, a greater degree of realism suggests that even $N(f)$ is only known to within a certain interval of error. Even with such *perturbed* information, linear optimal algorithms exist when $F_2 = \boldsymbol{R}$. The second part of Result 3 gives the first setting in which linear optimal algorithms must exist where the problem codomain is an arbitrary normed linear space. It is a forerunner and special case of Result 5.

Result 4 makes it clear, of course, that it is too much to hope that every linear problem will have a linear optimal algorithm. Micchelli's example (which can be found in [13], p. 60) is cleverly contrived, but rather far removed from any problem that might conceivably occur naturally. Result 6 gives a slightly simpler counterexample with problem codomain $\boldsymbol{R}^2$ under the Euclidean norm.

Result 5 is the culmination of a series of more specialized results which link the theory of optimal algorithms with the important idea of approximation by *splines*. Our plan here is to sketch a self-contained proof of the result without relying on the theory of splines. A careful statements of the Result 5 now follows.

Table 1. Results on Existence of Linear Optimal Algorithms for Linear Problems

	Year	Reference	Summary (LOA = Linear Optimal Algorithm)
1.	1965	Smolyak [11]	$\exists$ LOA when $F_2 = \boldsymbol{R}$
2.	1976	Osipenko [7]	$\exists$ LOA when $F_2 = C$
3.	1977	Micchelli–Rivlin [5]	$\exists$ LOA when $F_2 = \boldsymbol{R}$ and information is perturbed $\exists$ LOA when F_1 is a Hilbert space and F_0 is unit ball
4.	1978	Micchelli [3]	Counterexample with $F_1 = \boldsymbol{R}^3$ and $F_2 = \boldsymbol{R}^2$ with an l^4 norm
5.	1980	Traub–Woźniakowski [13]	$\exists$ LOA when $F_4 = T(F_1)$ is a Hilbert Space and $T(\ker(N))$ closed in F_4
6.	1985	Packel [9]	Counterexample with $F_1 = \boldsymbol{R}^3$ and $F_2 = \boldsymbol{R}^2$ with an l^2 norm
7.	1985	Packel [9]	$\exists$ LOA if allow extended range for S and φ
8.	1985	Werschulz–Woźniakowski [19]	Class of powerful counterexamples with real world overtones

THEOREM 2. *Let $F_0 = \{f \in F_1\colon \|T(f)\| \leqslant 1\}$ be generated by a restriction operator $T\colon F_1 \to F_4$ mapping F_1 into a Hilbert space $F_4 = \boldsymbol{H}$ and let $T(\ker(N))$ be closed in $\boldsymbol{H}$. Then a linear problem determined by S and N on F_0 has a linear and strongly optimal algorithm.*

Proof. We break the proof into bite-sized pieces, referring to Fig. 3 and relying on a few standard Hilbert space results as needed.

Given any $y = N(f) \in N(F_0)$, the subset $C = \{T(g) \in \boldsymbol{H}\colon g \in N^{-1}(y)\} = T(\ker(N)) + \{T(f)\}$ is a convex and closed subset in a Hilbert space and hence has a unique element of smallest norm. Denote this element by $\tau(y)$. It is also the case in Hilbert space (and geometrically evident in Fig. 3) that the inner products $\langle \tau(y), T(h)\rangle = 0 \,\forall\, h \in \ker(N)$ since C and $T(\ker(N))$ are "parallel" in $\boldsymbol{H}$. Furthermore, $\tau(y)$ is the unique member of C with this orthogonality property.

Let $g_1, g_2 \in N^{-1}(y)$ be such that $T(g_1) = T(g_2) = \tau(y)$. We show that $S(g_1) = S(g_2)$. Indeed, let $h = g_1 - g_2$ and note that $h \in \ker(N)$. Then

$$\|T(g_1)\|^2 = \|T(g_2 + h)\|^2 = \|T(g_2) + T(h)\|^2 = \|T(g_2)\|^2 + \|T(h)\|^2,$$

where the final equality comes from the orthogonality of $T(g_2)$ and $T(h)$ (the "Pythagorean theorem" generalized to Hilbert space). It follows directly that

$\|T(h)\| = 0$, so $h \in \ker(N) \cap \ker(T)$. From Theorem 1(a) proved earlier, we then have $h \in \ker(S)$, so $S(g_1) = S(g_2)$ as desired.

We can now define $\varphi^*(y) = S(g)$, where g is *any* representative of $N^{-1}(y)$ with $T(g) = \tau(y)$. It is easily checked that linearity of φ^* with respect to y results from the linearity of S and T and the previously stated orthogonality property characterizing $\tau(y)$.

Finally, the fact that φ^* is strongly optimal follows by showing that for each $y = N(f) \in F_0$, $\varphi^*(y)$ is a center of $S(V(f))$. To this end, choose $g \in N^{-1}(y)$ with $T(g) = \tau(y)$ and suppose $\varphi^*(y) + k \in S(V(f))$. We can restate this (with the help of linearity) as $S(g) + S(h) \in S(V(f))$ with $h \in \ker(N)$ and we must show that, symmetrically, $S(g) - S(h) \in S(V(f))$. We may assume that $g + h \in V(f)$ (so $\|T(g+h)\| \leqslant 1$) since otherwise $S(g+h) = S(g')$ for $g' \in V(f)$ and we can redefine h as $g' - g$. Orthogonality of $T(g)$ and $T(h)$ now gives

$$\|T(g-h)\|^2 = \|T(g+h)\|^2 \leqslant 1.$$

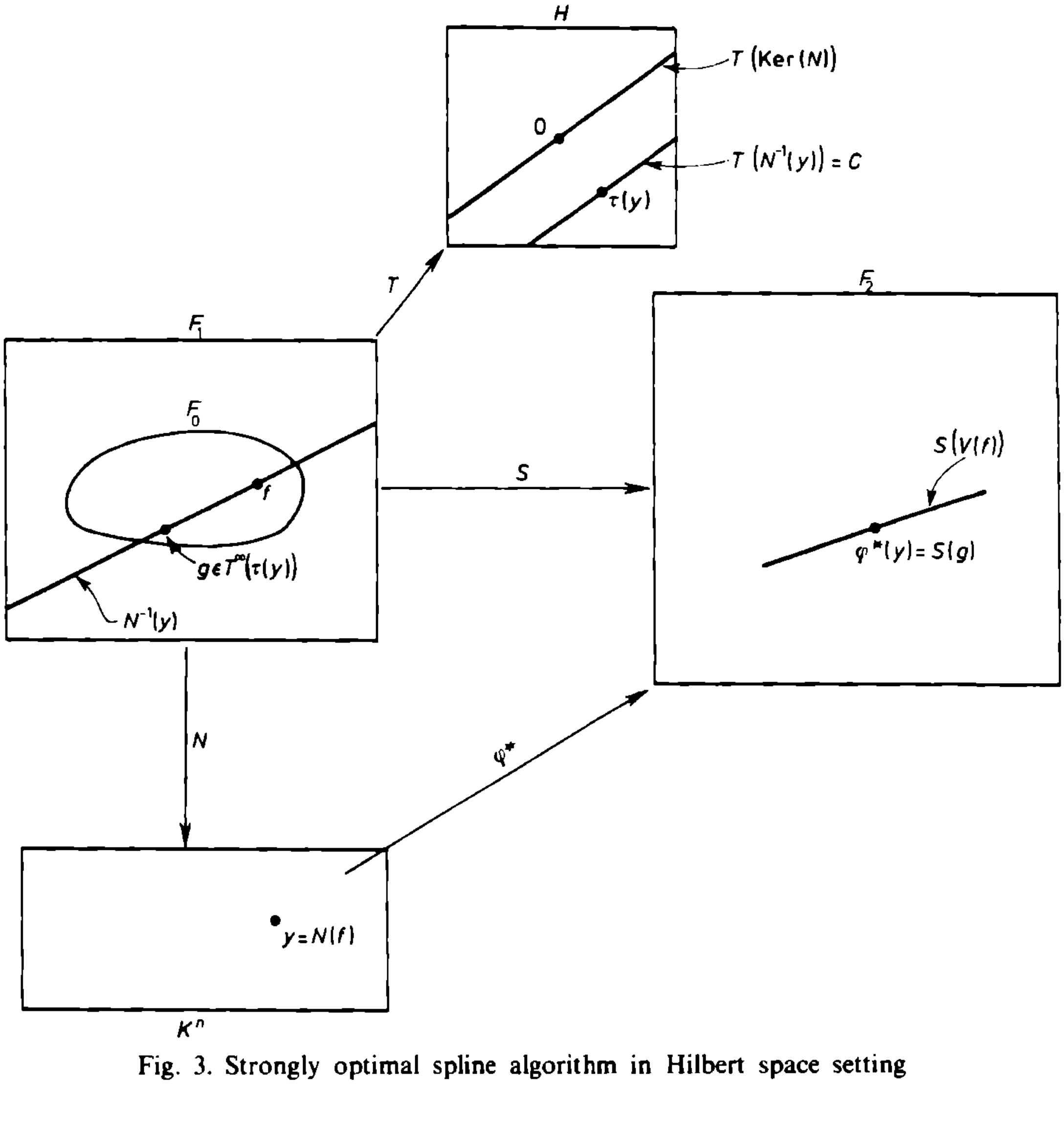

Fig. 3. Strongly optimal spline algorithm in Hilbert space setting

Hence $g-h \in F_0 \cap N^{-1}(y) = V(f)$ and $S(g)-S(h) \in S(V(f))$. This establishes strong optimality of φ^* and completes the proof of the theorem. We mention without proof (see [13], Chapter 4) that φ^* is the unique strongly optimal algorithm and the only optimal algorithm that is linear.

Result 7 resurrects, in an elegant but not totally satisfying way, the intuition that general linear problems *ought* to over linear optimal algorithms. To explain this we give a carefully stated theorem and a sketch of its proof. Fig. 4 gives a diagrammatic representation of the theorem.

THEOREM 3. *Given a linear problem defined by* $S\colon F_1 \to F_2$ *and* N *on a convex, balanced subset* F_0 *of* F_1, *there exists*:

(i) *A compact Hausdorff space* X *such that* F_2 *is isometrically isomorphic to a subspace* $\hat{F}_2$ *of* $B(X)$, *the bounded* $\mathbf{K}$*-valued functions on* X *with the sup-norm.*

(ii) *A linear optimal algorithm* $\varphi^*\colon N(F_0) \to \mathbf{B}(X)$ *satisfying* $\|\varphi^*(N(f)) - \hat{S}(f)\| \leqslant r(S, N) \,\forall f \in F_0$, *where* $\hat{S}(f)$ *denotes the image of* $S(f)$ *in* $\hat{F}_2$.

Proof. (i) It is a standard corollary to the Banach–Alaoglu theorem that any normed linear space is isometrically isomorphic to a subspace of $C(X)$, the continuous functions on some compact Hausdorff space X. Though we do not need more details for our purposes, it is worth noting that the space X is the unit ball of the conjugate space of X endowed with the weak* topology and the isometric action is provided by the *Gelfand map* which imbeds F_2 in its second conjugate F_2^{**}. See Packel [8] (or most other introductory functional analysis texts) for details. We shall need the larger space $B(X)$ to hold our embedding.

(ii) For each fixed $x \in X$, consider the linear problem defined by N and $S_x(f) = \hat{S}(f)(x)$. By Results 1 and 2, there exists a linear optimal algorithm $\varphi^*_x\colon N(F_0) \to \mathbf{K}$ such that

$$\|\varphi^*_x(N(f)) - S_x(f)\| \leqslant r(S_x, N) \leqslant r(S, N) \;\forall f \in F_0. \tag{1}$$

Letting x vary over X, we must now show that the linear operator φ^* thus defined on $N(F_0)$ has its range in $B(X)$. First note that

$$|S_x(f)| = |\hat{S}(f)(x)| \leqslant \|\hat{S}(f)\| = \|S(f)\|, \tag{2}$$

where the inequality results from $\|\hat{S}(f)\| = \sup\{|\hat{S}(f)(x)|\colon x \in X\}$. Using (1) and (2), we have for all $f \in F_0$,

$$\begin{aligned}|\varphi^*(N(f))(x)| &\leqslant |\varphi^*(N(f))(x) - S_x(f)| + |S_x(f)| \\ &\leqslant |\varphi^*_x(N(f)) - S_x(f)| + |S_x(f)| \leqslant r(S, N) + \|S(f)\|.\end{aligned}$$

Since the final expression is independent of x, $\varphi^*(N(f)) \in B(X)$ and the proof is complete.

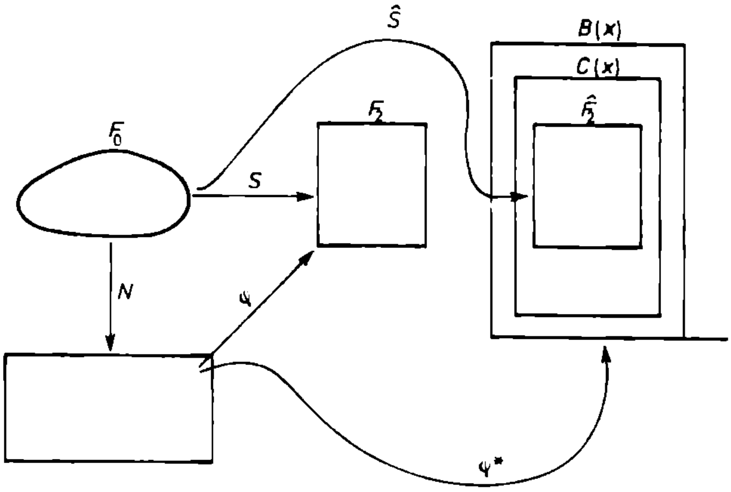

Fig. 4. Linear optimal algorithm with extended codomain

It is tempting to take the view, based on this result, that linear problems do indeed have linear optimal algorithms. We must simply be willing to give the solution operator (and its approximating algorithms) a codomain (namely $B(X)$) which extends beyond its range. There are problems with this view, however. The extended codomain is generally vastly larger and more complicated than the original. In addition, the members of $B(X)$ (other than isometric images from F_2) may have no real or abstract connection with the members of F_2. This leads to difficulties in relating a linear optimal algorithm (assuming we can constructively find one) to the original problem. However, the fact remains that a small but significant reformulation of the standard linear setting allows for linear optimal algorithms in a very general context. It is also possible that linear optimal algorithms might be obtained for less drastic extensions to the codomain of S. We take up this question briefly in the next section.

Finally, Result 8 is of considerable interest both in terms of its evolution and its content. The idea behind the basic counterexample emerged from an effort by Werschulz to model the problem of inverting a finite Laplace transform within the information-based framework. He then observed that the resulting linear problem could only have linear algorithms with infinite error. Woźniakowski showed how to construct nonlinear algorithms with finite error for such problems. By abstracting this example Werschulz and Woźniakowski [19] develop a class of linear problems with radii of information arbitrarily close to 0 for which all linear algorithms have infinite error.

Thus the positive thrust for the existence of linear optimal algorithms made by Result 7 is swiftly and decisively by Result 8. In addition, Result 8 includes the "naturally occurring" problem of inverting a finite Laplace transform (though it should be pointed out that one of the norms used to generate the counterexample is far from natural). Our faith in linear algorithms for linear problems is alternately strengthened and shaken, leaving us in a state of tantalizing mathematical ambiguity.

5. One more counterexample

We saw in the previous section that Result 7 (Theorem 3) has the drawback of requiring a vastly extended codomain for the solution operator in order to ensure a linear optimal algorithm. This section is motivated by the question of whether the extended codomain $B(X)$ might at least be shrunk to $C(X)$. We will exhibit an example of a linear problem in $C([0, 1])$ that has a linear optimal algorithm with codomain $B([0, 1])$, but not with codomain $C([0, 1])$. The example will require just minor modifications to the example developed in Section 3.

As in Section 3, let $F_1 = F_2 = C([0, 1])$ with the sup-norm and let F_0 be the unit ball of $C([0, 1])$. Then the information operator is again $N\colon F_0 \to \mathbf{R}$, $N(f) = \int_0^1 f(t)\,dt$. The solution operator is defined by $S\colon F_0 \to C([0, 1])$, $S(f)(x) = g(x)\int_0^x f(t)\,dt$, where

$$g(x) = \begin{cases} 1/x, & x \in [\frac{1}{4}, \frac{1}{2}], \\ 1/(1-x), & x \in (\frac{1}{2}, \frac{3}{4}], \\ 4, & \text{otherwise}. \end{cases}$$

Clearly the function g is carefully contrived to make something happen and a review of Fig. 2c should help to reveal the general intent. Indeed, for $x \in [\frac{1}{4}, \frac{3}{4}]$ the convex balanced sets Y_x will, thanks to the $g(x)$ factor, be "stretched" so that their boundary points on the horizontal axis all coincide, forcing a common radius of information $r_x = 1$ (see Fig. 5). Recall that the slopes of bounding tangent hyperplanes at $(\pm r_x, 0)$ for each fixed x determine linear optimal algorithms for the corresponding real-valued solution operator S_x. We conclude from the way such slopes must "jump" at $x = \frac{1}{2}$ that any linear optimal algorithm for the problem determined by S and N must be discontinuous at $\frac{1}{2}$ and must hence require $B([0, 1])$ rather than $C([0, 1])$ for its codomain. A little calculation shows, in fact, that any linear optimal algorithm φ^* for the problem must satisfy

$$\varphi^*(y)(x) = \begin{cases} 0, & x \in [\frac{1}{4}, \frac{1}{2}), \\ y/(1-x), & x \in (\frac{1}{2}, \frac{3}{4}]. \end{cases}$$

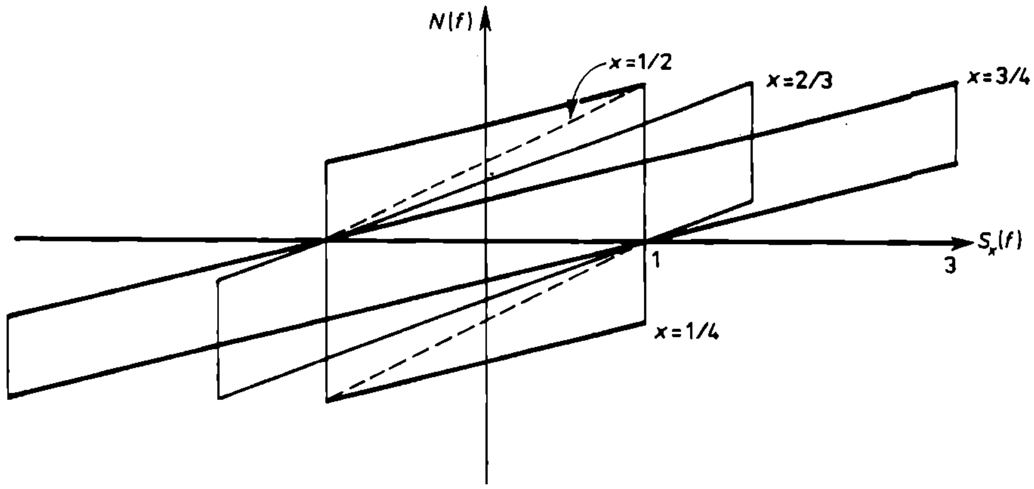

Fig. 5. Plots of $\{(S_x(f), N(f)): f \in F_0\} = Y_x$ with $g(x)$ factor

What have we gained for our efforts? It is tempting to conclude, based upon the above example, that the extended codomain of $B(X)$ in Theorem 3 cannot generally be shrunk to $C(X)$. A little thought shows that this conclusion would be premature. Even though the solution operator S has no linear optimal algorithm with values in $C([0, 1])$, applying Theorem 3 lands us in a larger and more mysterious underlying space X for which algorithm values might conceivably be restricted to $C(X)$ rather than just $B(X)$.

The example, while not serving its initially intended purpose, provides a pleasant surprise. With $C([0, 1])$ as the codomain of the solution operator (and of any approximating algorithm), the problem determined by S and N has no linear optimal algorithm. Thus we have constructed, from a problem which has at least some degree of naturality, perhaps the simplest counterexample to date for the nonexistence of linear optimal algorithms for linear problems.

6. Concluding remarks

The setting we have used for discussing error, as embodied in the definition of radius of information and algorithm optimality, has been the traditional "worst case" setting. Increasingly important in computer science are "average case" models of error and complexity. By assuming the existence of a measure on F_1 (the problem domain), average case models for information-based complexity can readily be formulated. Work in this area is in its early stages, but a surprising variety of results paralleling the worst case model have already been obtained (see [4], [16], and [17]). At this time there is no average case analogue for Theorem 3.

Our brief discussion in Section 3 of the nonlinear root finding problem restricted information on functions to evaluations at predefined points. If, instead, the ith evaluation point t_i can depend upon previous information values $f(t_j)$, $j < i$, an important example of *adaptive* information arises. In this case the well-known idea of root approximation by bisection naturally occurs. We choose $t_1 = 0$ (to get the sign of $f(0)$ and hence $f(1)$) and $t_2 = \frac{1}{2}$. We then use the sign of $f(\frac{1}{2})$ and the intermediate value theorem to decide adaptively upon what subinterval to iterate. With this choice of information it is readily checked that the radius of information for the problem is precisely $1/2^n$ and that the bisection algorithm which always takes the midpoint of the ith subinterval (if the process has not already terminated with an exact 0) is optimal.

It is no surprise that adaptively choosing the information should be a valuable strategy for this nonlinear problem. Indeed, we saw in Section 3 that restricting to information with n evaluation points chosen nonadaptively leads to a much larger radius of information. For linear problems with linear information of a given cardinality, it has been proved by Gal and Micchelli [1] and by Traub and Woźniakowski [13] (and this, at first glance, may be a surprise) that there always exists some nonadaptive information that will ensure the minimum possible uncertainty. Recent work by Wasilkowski and Woźniakowski [18] has extended this result to average case settings.

As the previous paragraphs suggest, information-based complexity has many open problems and research directions. We indicate a few of them here, referring the interested reader to [13] and [20] for a more substantial list. Given a specific solution operator S, a domain F_0, and a cardinality n for the amount of information specified by N, it is often a nontrivial problem to determine the *optimal information* operator N^* for minimizing $r(S, N)$. Even if $r(S, N^*) = \infty$, one can sometimes fix a problem element f and approximate $S(f)$ by a rapidly convergent sequence of algorithm values calculated for information operators with increasing cardinality. This forms the basis for an *asymptotic* setting for defining error, an area that is just beginning to be investigated.

We have focused on theoretical aspects of information-based complexity. There is also a considerable literature on applications of the theory to problems involving interpolation, approximation (take the solution operator S to be the identity), differentiation, integration, and differential equations. Work in these traditional areas has in some cases confirmed and in others called into question the optimality of classical algorithms, while introducing new results and approaches. An extensive list of references can be found in [13]. Given the generality of the theory and the ubiquitous nature of problems with partial information, potential for a wide variety of additional applications abounds. Some more recent areas of application include convergence of price mechanisms in mathematical economics (where a key result in the economics-directed paper of Saari and Simon [10] is closely related to a theorem in Traub and Woźniakowski [12]), computer vision (Lee [2]), and time series prediction (Milanese,. Tempo, and Vicino [6]).

Returning to the basic theoretical question considered in this paper, we have surveyed a variety of positive and negative results on the existence of linear optimal algorithms for linear problems. We hold out some hope for additional, more general positive results. Meanwhile, if pressed for an answer in three words or less as to whether linear optimal algorithms exist, we can only respond in our most authoritative voice with a definite "yes and no".

References

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

[2] D. Lee, *Optimal algorithms for image understanding: current status and future plans*, J. of Complexity 1 (1985).

[3] C. A. Micchelli, private communication, 1978.

[4] –, *Orthogonal projections are optimal algorithms*, J. Approx. Theory 40 (1984), 101–110.

[5] – and T. J. Rivlin, *A survey of optimal recovery*, in: *Optimal Estimation in Approximation Theory*, C. A. Micchelli and T. J. Rivlin, eds., Plenum Press, New York 1977, 1–54.

[6] M. Milanese, R. Tempo, and A. Vicino, *Robust time series prediction by optimal algorithms theory*, Report from Politecnico di Torino, Dipartimento di automatica e informatica, 1982.

[7] K. Yu. Osipenko, *Best approximation of analytic functions from information about their values at a finite number of points (English translation)* Math. Notes 19 (1976), 17–23.

[8] E. W. Packel, *Functional Analysis*: A Short Course, Krieger, Huntington, New York 1980.

[9] –, *Linear problems (with extended range) have linear optimal algorithms*, forthcoming in Aequationes Mathematicae, 1985.

[10] D. G. Saari and C. P. Simon, *Effective price mechanisms*, Econometrica 46 (1978) 1097–1125.

[11] S. A. Smolyak, *On optimal restoration of functions and functionals of them* (in Russian), Candidate Dissertation, Moscow State Univ., 1965.

[12] J. F. Traub and H. Woźniakowski, *Optimal linear information for the solution of nonlinear equations*, in *Algorithms andd Complexity: New Directions and Recent Results*, J. F. Traub, ed., Academic Press, New York 1976, 103–119.

[13] –, –, *A General Theory of Optimal Algorithms*, Academic Press, New York 1980.

[14] –, –, *Information and Computation*, in: *Advances in Computers*, 23, M. Yovits, ed., Academic Press, 1984, 35–92.

[15] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information, Uncertainty, Complexity*, Addison-Wesley, Reading, Mass, 1983.

[16] –, –, –, *Average case optimality for linear problems*, Journal of Theoretical Computer Science 29 (1984) 1–25.

[17] G. W. Wasilkowski and H. Woźniakowski, *Average case optimal algorithms in Hilbert space*, Report from Columbia University Department of Computer Science, 1982.

[18] –, –, *Can adaption help on the average?* Numer. Math. 44 (1984), 169–190.

[19] A. Werschulz and H. Woźniakowski, *Are linear algorithms always good for linear problems?* Report from Columbia University Department of Computer Science, 1985.
[20] H. Woźniakowski, *A survey of information-based complexity*, J. of Complexity 1 (1985).

Presented to the semester
Mathematical Problems in Computation Theory
September 16–December 14, 1985