An Information-Based Approach to Ill-Posed Problems

ARTHUR G. WERSCHULZ

Division of Science and Mathematics
Fordham University/College at Lincoln Center
New York, New York 10023

and

Department of Computer Science
Columbia University
New York, New York 10027

April, 1987

CUCS - 258-86

A problem is said to be ill-posed if the solution of the problem does not depend continuously on the input data. In this survey paper, we consider two different information-based settings for the optimal computation of approximate solution of ill-posed linear problems, namely the worst case and average case settings. These settings are studied for two different error criteria, namely, the absolute error and the residual error criteria. The main result for the absolute-error criterion is that algorithms having finite error exist for a given setting if and only if the solution operator is bounded in that setting. In the worst case setting with an absolute error criterion, this means that there is no algorithm for solving ill-posed problems having finite error. In the average case setting with an absolute error criterion, this means that algorithms having finite error exist if and only if the solution operator is "bounded on the average." Furthermore, when this holds, we exhibit optimal information of cardinality n, finding that the n^{th} minimal average error goes to zero as $n \to \infty$. The main result for the residual error criterion is that the problem may be formally reduced to the approximation problem. Hence, finite-error algorithms always exist. We exhibit optimal information of cardinality n. In the worst case setting, we give a necessary and sufficient condition for the n^{th} minimal error to go to zero as $n \to \infty$; in the average case setting, this always occurs. We use these results to determine the ε-complexity of ill-posed problems. The ε-complexity is infinite for *any* $\varepsilon > 0$ in the worst case setting with the absolute criterion. However, in the average case setting with either error criterion and the worst case setting with the residual error criterion, we determine necessary and sufficient conditions for the ε-complexity to be finite for all $\varepsilon > 0$; moreover, we find algorithms yielding ε-approximations with almost-minimal cost.

Presented at the Second Symposium on the Complexity of Approximately-Solved Problems held at Columbia University, New York, NY, April, 1987.

1. Introduction

The concept of a well-posed problem was introduced in Hadamard (1952). A problem is said to be well-posed if it its solution exists, is unique, and depends continuously on its data; a problem that is not well-posed is said to be ill-posed. Hadamard gives the impression that any well-formulated physical problem must be well-posed, and that ill-posed problems were merely problems that had not been formulated correctly. However, in the years since the appearance of Hadamard's treatise, many important practical problems have been found to be ill-posed. Examples of such problems include the following:

- inversion of the Laplace transform, whether the "usual" transform defined over $[0, \infty)$ or the finite transform whose inversion is discussed in Dunn (1967),
- Fujita's equation relating molecular weight distribution to the steady-state concentration or optical density in a centrifuged sample, see Gehatia (1970),
- problems in computational vision, such as edge detection, optical flow, surface reconstruction, and determining shape from shading, see Poggio (1985), and
- problems in remote sensing, see Twomey (1977).

These problems are examples of the *Fredholm problem of the first kind*, in which one is trying to solve a problem of the following form:

> Let $L: U \to V$ be a compact linear transformation of an infinite-dimensional normed linear space U into a normed or seminormed linear space V. Let D denote the range of L. For $f \in D$, find $u \in U$ such that
>
> $$Lu = f. \tag{1.1}$$

In this paper, we will assume that L is an injection, so that for any $f \in D$, there exists a unique $u \in U$ satisfying (1.1). This will allow us to talk about *the* solution of the problem (1.1).

To help fix our ideas, it will be useful to consider the following example, which we will follow throughout this paper.

Example 1.1. The problem of inverting the *finite Laplace transform* arises in the "measurement of the distribution of an absorbing gas (such as ozone in the earth's atmosphere) from the spectrum of scattered light;" see pp. 12–13 of Twomey (1977) for details.

For the sake of normalization, we let I denote the unit interval $[0, 1]$. We take $U = L_2(I)$, the usual space of (Lebesgue) square-integrable functions on I. Given a non-negative integer r, we choose $V = H^r(I)$, the *Sobolev space* consisting of those functions whose $(r-1)^{st}$ derivative is absolutely continuous and whose r^{th} derivative is in $L_2(I)$. The space $H^r(I)$ is a Hilbert space under the usual Sobolev norm $\|\cdot\|_{H^r(I)}$. (See Ciarlet (1978) for further discussion of Sobolev spaces and norms.)

We define an operator $L: L_2(I) \to H^r(I)$ by setting

$$(Lu)(s) = \int_0^1 e^{-st} u(t)\, dt \qquad (0 \le s \le 1)$$

for $u \in L_2(I)$. It is straightforward to verify that L is a compact operator. Moreover, Marti (1983) shows that the operator L is injective. □

Since the inverse of a compact linear transformation with infinite-dimensional domain is always unbounded, this means that the solution of a Fredholm problem of the first kind does not depend continuously on its data. Thus, the Fredholm problem of the first kind is ill-posed. Most of the important ill-posed problems arise as Fredholm problems of the first kind. Hence, we limit our discussion of ill-posed problems in this paper to such problems. We do this partially for expository purposes; many of the results described below hold for more general kinds of ill-posed problems.

There is a huge literature dealing with the calculation of approximations to solutions of ill-posed problems; see, e.g., Carasso and Stone (1975), Hämmerlin and Hoffmann (1983), Tikhonov (1963), and Tikhonov and Arsenin (1977), as well as the references contained therein. One successful technique is *regularization*, in which the ill-posed problem to be solved is replaced by a (well-posed) minimization problem:

> Let $f \in D$. Given $\lambda > 0$, find $u_\lambda \in U$ such that
>
> $$E_\lambda(u_\lambda) = \inf_{w \in U} E_\lambda(w),$$

where

$$E_\lambda(w) = \|Lw - f\|^2 + \lambda J(f),$$

J being a quadratic "penalty functional."

See, e.g., Tikhonov (1963), Tikhonov and Arsenin (1977) for further discussion.

Clearly, the success of a regularization methods depends on the choice of the regularization parameter λ and the penalty functional J. Note that a careful balance must be maintained in the choice of λ: if λ is too large, the error of the method will be large; however, as λ gets small, the minimization problem becomes less and less well-posed. This problem seems to be well-addressed in the literature. Perhaps one of the most successful techniques for choosing λ is generalized cross validation, as described in, e.g., Wahba (1985) and Wahba (1986). Unfortunately, good criteria for choosing the penalty functional are harder to find. A typical choice is to let $J(f)$ be the r^{th}-order Sobolev seminorm of f, for some value of r.

There is one additional problem with regularization methods. To get a bound on the error of a regularization method, it appears to be necessary to make additional a priori assumptions about the solution. Such assumptions are often difficult to verify in practice.

Why do we need such assumptions for regularization methods? The problem can be traced to one of two sources:

(1) The fault lies with using regularization methods for these problems. That is, we should look at a wider class of algorithms.
(2) There is something inherent in ill-posed problems that causes difficulties. That is, no matter *what* class of algorithms we use, we can expect ill-posed problems to be hard to solve.

This survey paper describes how information-based complexity has recently been used to analyze ill-posed problems. An expository account of information-based complexity may be found in Woźniakowski (1986); a more technical discussion may be found in Traub and Woźniakowski (1980). For more detailed information about the results in this paper (including more general statements of those results, as well as proofs of all the theorems), the interested reader is referred to Werschulz (1986) and to Werschulz (1987).

To use the information-based approach, we first specify the problem and the permissible information about the problem. Since this information is generally *incomplete*, in the sense that the information available about a particular right-hand side f does not completely determine f, no algorithm using this information can give the exact solution for all f. Hence, we must also specify our measure of the accuracy of an algorithm that (approximately) solves the problem using this information. Having done this, we may then seek the best algorithm for the problem that uses this information, where we say an algorithm is "best" if its error is minimal among all algorithms using the same information.

Our results will depend strongly on the *problem*, the available *information*, and the way in which we measure *error*. In this paper, our problem will be given by (1.1). We shall assume (mainly for expository purposes) that the spaces U and V are Hilbert spaces, although many of the results of this paper extend to more general normed linear spaces. In addition, we shall also assume that the range D of L is dense in V; of course, this is no loss of generality since we can always replace V by the closure $\overline{D}$ of D in V.

We assume that the available information about a problem element f consists of the values of a finite number of linear functionals at f. Since we are dealing with a Hilbert space setting, this means that we assume that there exist $f_1, \ldots, f_n$ such that for any $f \in V$, the only knowledge we have of f is the *information* Nf given by

$$Nf = \begin{bmatrix} (f, f_1) \\ \vdots \\ (f, f_n) \end{bmatrix}, \tag{1.2}$$

where $(\cdot, \cdot)$ denotes the inner product of V. The *cardinality* of information N defined by (1.2) is the number of linearly independent elements among $\{f_1, \ldots, f_n\}$. Thus, information N of cardinality n is a linear operator $N: V \to \mathbf{R}^n$.

EXAMPLE 1.1 (*continued*). For the problem of the finite Laplace transform, we assume that for any set

$\{f_1, \ldots, f_n\}$ of linearly independent functions from $H^{-r}(I)$, we can evaluate the information

$$Nf = \begin{bmatrix} \int_0^1 f(x) f_1(x)\, dx \\ \vdots \\ \int_0^1 f(x) f_n(x)\, dx \end{bmatrix} \qquad \forall f \in H^r(I).$$

Clearly, N is information of cardinality n. □

Once we have determined the available information, an algorithm is merely a function that combines the information values for a particular problem element f into an approximation of the solution u of the problem $Lu = f$. Hence, an *algorithm* using information N of cardinality n is a mapping $\varphi\colon \mathbf{R}^n \to U$, and $\varphi(Nf)$ is the approximation to $u = L^{-1}f$ produced by the algorithm φ.

Now, we must specify how to measure the error of an algorithm φ using information N. This may be conveniently divided into two parts. First, we must choose an *error criterion*; that is, we must decide how to measure the error at a particular problem element f. Then, we must choose a *setting* that tells us how to combine these error measurements at each f into a measurement over the full set of problem elements. In this paper, we will consider two error criteria. The *absolute error criterion* measures the error of φ at f by

$$e_{\text{abs}}(\varphi, N, f) = \|L^{-1}f - \varphi(Nf)\|_U.$$

The *residual error criterion* measures the error of φ at f by

$$e_{\text{res}}(\varphi, N, f) = \|f - L\varphi(Nf)\|_W,$$

Here, W is a Hilbert space containing V such that V is densely embedded in W; that is, V is a dense subspace of W, and the inclusion mapping from V into W is continuous. Thus, the norm of the space W measures the error in the residual at f, whereas the norm of the original space V measures the smoothness of the problem elements.

EXAMPLE 1.1 (*continued*). For the problem of inverting the finite Laplace transform, we let $W = L_2(I)$ when using the residual error criterion. That is, we use the $L_2(I)$-norm to measure the residual error and the $H^r(I)$-norm to measure smoothness of the problem elements. It is well-known that the inclusion mapping from $H^r(I)$ into $L_2(I)$ is a dense embedding. □

We also consider two *settings*. Let e denote either of e_{abs} or e_{res}. First, we give the worst case setting by defining the *worst case error* of φ by

$$e^{\text{wor}}(\varphi, N) = \sup_{\substack{f \in D \\ \|f\|_V \le 1}} e(\varphi, N, f);$$

the restriction to problem elements whose norm is at most one being a normalization. Next, we give an average case setting. Let μ be a probability measure on V such the range D of L is of full measure, i.e., $\mu(D) = 1$. Then the *average case error* of φ is given by

$$e^{\text{avg}}(\varphi, N) = \left(\int_D e(\varphi, N, f)^2\, \mu(df) \right)^{1/2}.$$

So, we have four possibilities to consider when measuring the error of the algorithm, since the choice of setting and the choice of how error is measured at a particular problem element are independent. We will denote these four possibilities by $e^{\text{wor}}_{\text{abs}}$, $e^{\text{wor}}_{\text{res}}$, $e^{\text{avg}}_{\text{abs}}$, and $e^{\text{avg}}_{\text{res}}$, which respectively denote the worst case setting with the absolute error criterion, the worst case setting with the residual error criterion, the average case setting with the absolute error criterion, and the average case setting with the residual error criterion.

The crucial idea underlying the results of this paper is the *radius of information*, which is defined to be the minimal error among all algorithms using given information. Note that the radius of information is a function only of the problem, the way that the error is measured, and the information. In particular, it

is independent of any particular algorithm. Stated differently, the radius is an invariant of the problem, depending only on the problem formulation and the resources that are available to solve the problem.

We now outline the remainder of this paper. In Section 2, we give results for the absolute error criterion. Our results for this setting may be summarized by saying that there exist algorithms having finite error for a given setting if and only if L^{-1} is bounded with respect to that setting. In the worst case setting, this means that there is no algorithm for solving an ill-posed problem whose error is finite. In the average case setting, this means that finite-error algorithms exist if and only if the operator L^{-1} is "bounded on the average." When this holds, we exhibit an *optimal error algorithm* using N, i.e., an algorithm whose average case error is minimal among all algorithms using N. Then, we exhibit optimal information N_n of cardinality n (i.e., information such that the average radius of N_n is minimal among all information of cardinality at most n) and show that the n^{th} minimal average radius (i.e., the radius of N_n) goes to zero as $n \to \infty$. Hence, we can get arbitrarily-accurate solutions of an ill-posed problem in the average case setting for the absolute error criterion if and only if L^{-1} is bounded on the average.

In Section 3, we give results for the residual error criterion. Our approach for both the worst case and average settings is to reduce the ill-posed problem to the standard "approximation problem" of approximating F in W. We find that finite-error algorithms always exist for either setting, and we exhibit optimal error algorithms. In particular, we show that for any information N, the same algorithm is optimal in the average case setting for both the absolute error criterion and the residual error criterion. We also exhibit optimal information of cardinality n for either setting. In the worst case setting, we show that the n^{th} minimal radius goes to zero as $n \to \infty$ if and only if the embedding of V in W is compact, whereas in the average case setting, the n^{th} minimal radius *always* goes to zero.

Finally, in Section 4, we use these results to determine the ε-complexity of ill-posed problems, i.e., the minimal cost of finding an ε-accurate approximation. In the worst case setting with the absolute error criterion, the ε-complexity is infinite for *any* $\varepsilon > 0$. In the worst case setting with the relative error criterion and the average case setting with the absolute error criterion, we find conditions that are necessary and sufficient for the ε-complexity to be finite for all $\varepsilon > 0$. In the average case setting with the relative error criterion, we find that the ε-complexity is always finite for all $\varepsilon > 0$. Moreover, we find algorithms yielding ε-accurate approximations with almost minimal cost in all these cases.

2. The absolute error criterion

In this section, we cite results about the existence of optimal error algorithms when the error at a given problem element is measured by the absolute error e_{abs}. In some instances, we sketch the proofs of results mentioned in this survey paper, while in others, we omit the proofs. The proofs of all results in this section may be found in Werschulz (1986).

We deal with the worst case and average case settings. In the worst case setting, we find that the error is always infinite. Moreover, we find that bad cases happen often, i.e., that the set of problem elements at which the error is large is "big." In the average case setting, we show that the error is finite iff L^{-1} is "bounded on the average." Furthermore, when L^{-1} is bounded on the average, we describe, for any information, a linear optimal error algorithm, i.e., an algorithm that is a linear combination of the information it uses whose error is minimal among *all* algorithms using that information. Finally, when L^{-1} is bounded on the average, we describe optimal information of cardinality n for any nonnegative integer n.

2.1. The worst case setting.

Recall that we are trying to solve the problem

$$Lu = f \qquad \forall f \text{ such that } \|f\|_V \le 1,$$

where $L: U \to V$ is a compact injection whose range D is dense in V. As before, our sole knowledge of a problem element f is the information Nf, where $N: V \to \mathbf{R}^n$ is a continuous linear transformation. Hence, there exist $f_1, \ldots, f_n$ such that

$$Nf = \begin{bmatrix} (f, f_1) \\ \vdots \\ (f, f_n) \end{bmatrix} \qquad \forall f \in V.$$

In this section, the error of an algorithm φ using information N is given by

$$e^{\text{wor}}_{\text{abs}}(\varphi, N) = \sup_{\substack{f \in D \\ \|f\|_V \le 1}} \|L^{-1}f - \varphi(Nf)\|_U .$$

Our main result is

THEOREM 2.1.1. *For any algorithm φ using information N,*

$$e^{\text{wor}}_{\text{abs}}(\varphi, N) = +\infty.$$

Sketch of proof. Let

$$r^{\text{wor}}_{\text{abs}}(N) = \inf_{\varphi} e^{\text{wor}}_{\text{abs}}(\varphi, N)$$

denote the *radius of information*, i.e., the minimal error among all algorithms using the information N. Then

$$\rho^{\text{wor}}_{\text{abs}}(N) \le r^{\text{wor}}_{\text{abs}}(N) \le 2\rho^{\text{wor}}_{\text{abs}}(N),$$

where

$$\rho^{\text{wor}}_{\text{abs}}(N) = \sup_{h \in D \cap \ker N} \frac{\|L^{-1}h\|_U}{\|h\|_V},$$

see, e.g., Traub and Woźniakowski (1980). Assume (without serious loss of generality) that $f_1, \ldots, f_n$ are orthonormal and belong to D. Let

$$h = f - \sum_{j=1}^{n} (f, f_j) f_j, .$$

Then $h \in D \cap \ker N$, from which we easily see that

$$\frac{\|L^{-1}f\|_U}{\|f\|_V} \le \sum_{j=1}^{n} \|f_j\|_V \|L^{-1}f_j\|_U + \rho^{\text{wor}}_{\text{abs}}(N) .$$

Since L^{-1} is unbounded, this inequality implies that $\rho^{\text{wor}}_{\text{abs}}(N)$ is infinite, which in turn implies that $r^{\text{wor}}_{\text{abs}}(N)$ is infinite. Hence, there exists no finite-error algorithm using the information N. □

So, no algorithm can give finite worst case error for our problem. But perhaps we can find an algorithm whose worst case behavior (which, by the previous theorem, must be bad) does not often occur. The next result quashes that hope. For any information N, define the *zero algorithm φ^0 using N* by

$$\varphi^0(Nf) = 0 \qquad \forall f \in D.$$

Admittedly, the zero algorithm is about as naive an algorithm as one could possibly invent. The next theorem tells us that no algorithm can be much better than the zero algorithm:

THEOREM 2.1.2. *Let $q \in [0, 1)$. For any information N and for any algorithm φ using N, let*

$$A_q = \left\{ f \in D : \frac{\|L^{-1}f - \varphi(Nf)\|_V}{\|L^{-1}f - \varphi^0(Nf)\|_V} \le q \right\} .$$

Then the relative interior of A_q in D is empty. □

What does this really tell us? For $0 \le q < 1$, the set A_q is the set of all problem elements f at which

$$e_{\text{abs}}(\varphi, N, f) \le q\, e_{\text{abs}}(\varphi^0, N, f).$$

That is, $f \in A_q$ if and only if the absolute error of φ at f is no worse than q times the absolute error of the zero algorithm at f. The theorem tells us that A_q has empty relative interior, i.e., that A_q is "small." Since q can be arbitrarily close to 1, this means that the set of problem elements at which the algorithm φ does appreciably better than the zero algorithm is small. Note that this result does not make use of the worst case setting; it tells us what happens at an arbitrary problem element. Hence, it tells us that there are many problem elements that are "bad cases."

So, there is no algorithm with finite worst case absolute error. Moreover, the bad cases occur often. One might think that the fault may be found in the class of permissible information, i.e., this class is too restrictive. In Werschulz (1986), we show that this is not the case. We can greatly generalize the class of permissible information, including such cases as adaptive information, non-continuous linear information, and mildly-smooth nonlinear information; in all these cases, there is no algorithm for the Fredholm problem of the first kind whose worst case absolute error is finite.

2.2. The average case setting.

We now turn to the average case setting for the absolute error criterion. To do this, we must first put a probability measure on the Hilbert space V, see, e.g., Kuo (1975) and Skorohod (1974). Let $A: V \to V$ be a compact, self-adjoint, positive definite injection whose *trace* (i.e., the sum of its eigenvalues) is finite. We then let μ be a *Gaussian measure* whose *mean element* is zero, i.e.,

$$\int_V (f, v)\, \mu(df) = 0 \qquad \forall\, v \in V\,,$$

and *covariance operator* is A, i.e.,

$$(Av, w) = \int_V (f, v)(f, w)\, \mu(df) \qquad \forall\, v, w \in V\,.$$

Expressed using the characteristic functional, this means that

$$\int_V e^{\sqrt{-1}(f,v)}\, \mu(df) = e^{-(Av,v)/2} \qquad \forall\, v \in V\,.$$

Alternatively, for any $v \in V$ and any $d \in \mathbf{R}$, we have

$$\mu\big(\{\, f \in V : (f, v) \le d \,\}\big) = \frac{1}{\sqrt{2\pi(Av, v)}} \int_{-\infty}^{d} e^{-t^2/(2(Av,v))}\, dt\,,$$

which gives a relation between Gaussian measures on Hilbert spaces and the familiar Gaussian distribution on the real numbers. In what follows, we will write S_μ for the covariance operator A of the measure μ, to emphasize the connection between the measure and the covariance operator.

Given such a measure μ on V, we make one further assumption relating L and μ, namely, that D is measurable and $\mu(D) = 1$. The reason for this assumption is that we are interested in average behavior over the domain D of L^{-1}.

EXAMPLE 2.2.1. Suppose that $\lambda_1 \ge \lambda_2 \ge \ldots > 0$ are the singular values of L, and that u_i is the eigenvector of L^*L corresponding to the eigenvalue λ_i^2. Define

$$e_i = \frac{1}{\lambda_i} L u_i \qquad i = 1, 2, \ldots\,.$$

Let μ be the Gaussian measure whose covariance operator S_μ is defined by

$$S_\mu e_i = \sigma_i e_i \qquad i = 1, 2, \ldots\,,$$

where

$$\sigma_1 \ge \sigma_2 \ge \ldots > 0 \qquad \text{with} \qquad \lim_{i \to \infty} \sigma_i = 0$$

satisfies

$$\sum_{i=1}^{\infty} \frac{\sigma_i}{\lambda_i^2} < \infty.$$

Then D is measurable and $\mu(D) = 1$. □

As before, information N of cardinality at most n is a continuous linear transformation

$$N: V \to \mathbf{R}^n.$$

However, an *algorithm* φ *using* N is now defined to be a mapping

$$\varphi: \mathbf{R}^n \to U$$

such that $\|\varphi(N\cdot)\|_U^2$ is measurable. The *average case error* of φ (under the absolute error criterion) is then given by

$$e_{\text{abs}}^{\text{avg}}(\varphi, N) = \left[\int_D \|L^{-1}f - \varphi(Nf)\|_U^2 \, \mu(df)\right]^{1/2}.$$

Since L is compact, it may be checked that L^{-1} is measurable, and so $e_{\text{abs}}^{\text{avg}}(\varphi, N)$ is well-defined. Note that the average case error $e_{\text{abs}}^{\text{avg}}(\varphi, N)$ can be either finite or infinite.

Guided by our experience in Section 2.1, we first ask whether there exist algorithms whose error is finite. Let

$$r_{\text{abs}}^{\text{avg}}(N) = \inf_{\varphi} e_{\text{abs}}^{\text{avg}}(\varphi, N)$$

denote the (average case) *radius of information*. We need to know when $r_{\text{abs}}^{\text{avg}}(N)$ is finite.

Let us say that L^{-1} is *bounded on the average* if

$$\int_D \|L^{-1}f\|_U^2 \, \mu(df) < \infty.$$

Note that there exist unbounded operators that are bounded on the average:

EXAMPLE 2.2.1 (*continued*). With the measure μ defined above, we find that

$$\int_D \|L^{-1}f\|_U^2 \, \mu(df) = \sum_{i=1}^{\infty} \frac{\sigma_i}{\lambda_i^2} < \infty.$$

Hence, L^{-1} is bounded on the average, even though L^{-1} is unbounded. □

We then have

THEOREM 2.2.1. *The following are equivalent:*

(1) L^{-1} *is bounded on the average.*

(2) *For any continuous linear information* N,

$$r_{\text{abs}}^{\text{avg}}(N) < \infty.$$ □

So, if L^{-1} is not bounded on the average, then for *any* information N, there exists no algorithm using N whose average case error is finite. If L^{-1} *is* bounded on the average, then for *every* information N, there exists an algorithm using N whose average case error is finite.

Without loss of generality, we assume that there exist $f_1, \ldots, f_n$ such that

$$Nf = \begin{bmatrix} (f, f_1) \\ \vdots \\ (f, f_n) \end{bmatrix} \qquad \forall f \in V,$$

where

$$(S_\mu f_i, f_j) = \delta_{ij} \qquad (1 \le i, j \le n).$$

Define $P: V \to V$ by

$$Pf = \sum_{i=1}^{n} (f, f_i) \, S_\mu f_i \qquad \forall f \in V;$$

we say that Pf is the *μ-spline interpolating* $f \in V$. Then the *μ-spline algorithm* φ^s *using* N is given by

$$\varphi^s(Nf) = L^{-1}Pf = \sum_{i=1}^{n} (f, f_i) \, L^{-1} S_\mu f_i \qquad \forall f \in V.$$

It may be shown that under the assumptions of this subsection, the mapping $L^{-1}S_\mu^{1/2}: V \to U$ is a bounded linear transformation. From this, it follows that the μ-spline algorithm is well-defined. The well-definedness of the μ-spline algorithm holds, regardless of whether L^{-1} is bounded on the average.

We then have

THEOREM 2.2.2. *For any information N, the μ-spline algorithm using N is an optimal error algorithm, i.e.,*

$$e^{\text{avg}}_{\text{abs}}(\varphi^s, N) = r^{\text{avg}}_{\text{abs}}(N) = \left[\int_D \|L^{-1}f\|_U^2 \, \mu(df) - \sum_{i=1}^{n} \|L^{-1}S_\mu f_i\|^2\right]^{1/2}. \qquad \square$$

Thus the algorithm φ^s is an optimal error algorithm using N in the average case. Note that this optimal error algorithm is *linear*, i.e., it is a linear combination of the information it uses. Hence, the structure of this optimal error algorithm is uncomplicated, which makes it easy to implement.

Now that we know the optimal algorithm using given information, it is only natural to seek optimal information of given cardinality n. Let

$$r^{\text{avg}}_{\text{abs}}(n) = \inf\{\, r^{\text{avg}}_{\text{abs}}(N) : N \text{ is information of cardinality at most } n \,\}$$

denote the n^{th} *minimal radius of information*. Information N_n of cardinality at most n is said to be n^{th} *optimal information* if

$$r^{\text{avg}}_{\text{abs}}(N_n) = r^{\text{avg}}_{\text{abs}}(n).$$

If L^{-1} is not bounded on the average, then

$$r^{\text{avg}}_{\text{abs}}(n) = \infty,$$

and so any information is (trivially) optimal. We now suppose that L^{-1} is bounded on the average. Let

$$K = \left(L^{-1}S_\mu^{1/2}\right)^* \left(L^{-1}S_\mu^{1/2}\right) : V \to V.$$

Since $L^{-1}S_\mu^{1/2}$ is bounded, its adjoint $\left(L^{-1}S_\mu^{1/2}\right)^*$ is defined; however, since L^{-1} is unbounded, it is not generally true that $\left(L^{-1}S_\mu^{1/2}\right)^* = S_\mu^{1/2}\left(L^{-1}\right)^*$. We find that K is compact and has finite trace. More precisely, if we let $\kappa_1 \geq \kappa_2 \geq \cdots > 0$ denote the eigenvalues of K, corresponding to the orthonormal eigenvectors $z_1, z_2, \ldots$, it may be shown that

$$\operatorname{tr} K = \sum_{i=1}^{\infty} \kappa_i = \int_D \|L^{-1}f\|_U^2 \, \mu(df) = \|L^{-1}S_\mu^{1/2}\|^{1/2}.$$

We now describe how to find nearly optimal information of cardinality n. Let $\delta > 0$. Choose

$$\theta \in \left(0, \min\left\{\kappa_n^{1/2}, \tfrac{1}{2}\delta \frac{\sum_{i=n+1}^{\infty} \kappa_i}{\sum_{i=1}^{n} \kappa_i^{1/2}}\right\}\right).$$

It may be shown that there exist $f_1, \ldots, f_n$ such that

$$(S_\mu f_i, f_j) = \delta_{ij} \qquad (1 \leq i, j \leq n)$$

and

$$\|z_i - S_\mu^{1/2} f_i\| < \frac{\theta}{\|L^{-1}S_\mu^{1/2}\|} \qquad (1 \leq i \leq n).$$

Define information $N_{n,\delta}$ of cardinality n by

$$N_{n,\delta} f = \begin{bmatrix} (f, f_1) \\ \vdots \\ (f, f_n) \end{bmatrix} \qquad \forall f \in V.$$

We then have

THEOREM 2.2.3. *Let L^{-1} be bounded on the average. For any non-negative integer n, the n^{th} minimal average radius is given by*

$$r_{\text{abs}}^{\text{avg}}(n) = \left[\int_D \|L^{-1}f\|^2\,\mu(df) - \sum_{i=1}^{n}\kappa_i\right]^{1/2} = \left[\sum_{i=n+1}^{\infty}\kappa_i\right]^{1/2}.$$

Moreover, for any $\delta > 0$,

$$r_{\text{abs}}^{\text{avg}}(N_{n,\delta}) < \sqrt{1+\delta}\, r_{\text{abs}}^{\text{avg}}(n),$$

and so the information $N_{n,\delta}$ is n^{th} optimal information, to within a factor of $\sqrt{1+\delta}$. □

Note also that this result implies that when L^{-1} is bounded on the average,

$$\lim_{n\to\infty} r_{\text{abs}}^{\text{avg}}(n) = 0.$$

This means that for any $\varepsilon > 0$, one can find information and an algorithm using that information, such that the average case absolute error of that algorithm is at most ε. In the terminology of Traub and Woźniakowski (1980), this means that the problem is *convergent.*

We apply these results to our

EXAMPLE 2.2.1 (*concluded*). We find that

$$Ke_i = \frac{\sigma_i}{\lambda_i^2}e_i.$$

Hence,

$$r_{\text{abs}}^{\text{avg}}(n) = \left[\sum_{i=n+1}^{\infty}\frac{\sigma_i}{\lambda_i^2}\right]^{1/2}.$$

Choosing

$$f_i = \sigma_i^{-1/2}e_i \qquad (1 \le i \le n),$$

we find that $e_i = S_\mu^{1/2}f_i$. So, we will be able to find optimal information, rather than information that is optimal only to within a factor of $\sqrt{1+\delta}$ for arbitrary $\delta > 0$. Using the construction of Theorem 2.2.3, we define information N_n by

$$N_n f = \begin{bmatrix}(f, f_1)\\ \vdots \\ (f, f_n)\end{bmatrix} \qquad \forall f \in V.$$

Then N_n is n^{th} optimal information, i.e.,

$$r_{\text{abs}}^{\text{avg}}(N_n) = r_{\text{abs}}^{\text{avg}}(n).$$

Finally, define an algorithm φ_n using N_n by

$$\varphi_n(N_n f) = \sum_{i=1}^{n}\frac{(f, e_i)}{\lambda_i}u_i.$$

Then φ_n is an optimal error algorithm using N_n, i.e.,

$$e_{\text{abs}}^{\text{avg}}(\varphi_n, N_n) = r_{\text{abs}}^{\text{avg}}(N_n) = r_{\text{abs}}^{\text{avg}}(n),$$

and so φ_n is an n^{th} minimal error algorithm for this problem, i.e.,

$$r_{\text{abs}}^{\text{avg}}(\varphi_n, N_n) \le r_{\text{abs}}^{\text{avg}}(\varphi, N)$$

for any algorithm φ using information N of cardinality at most n. □

3. The residual error criterion

In this section, we cite results about the existence of optimal error algorithms when the error at a given problem element is measured by the residual error e_{res}. Our main approach is to show that for the residual error criterion, the solution of the problem $Lu = f$ using information Nf about the problem element f can be formally reduced to the *approximation problem* of approximating f using information Nf. In particular, this means that optimal information for the two problems is the same.

As in the previous section, we deal with worst case and average case settings. The results of the worst case setting are taken from Werschulz (1987), whereas the results of the average case setting are new. For the worst case setting, we find that the problem is convergent if and only if the embedding of the space V (which measures the smoothness of the problem elements) into the space W (in which we are measuring the error) is compact. For the average case setting, we find that the problem is *always* convergent. In each of these settings, we characterize n^{th} optimal information.

3.1. The worst case setting.

In this section, the error of an algorithm φ using information N is given by

$$e_{\text{res}}^{\text{wor}}(\varphi, N) = \sup_{\substack{f \in D \\ \|f\|_V \le 1}} \|f - L\varphi(Nf)\|_W \,. \tag{3.1.1}$$

Here, W is a Hilbert space containing V such that V is densely embedded in W.

Our first goal is to find the minimal error among all algorithms using information N. This is given by the *radius of information*

$$r_{\text{res}}^{\text{wor}}(N) = \inf_{\varphi} e_{\text{res}}^{\text{wor}}(\varphi, N).$$

The following theorem gives a simple formula for the radius of information:

THEOREM 3.1.1. *For any information N, the radius of information is given by*

$$r_{\text{res}}^{\text{wor}}(N) = \sup_{\substack{h \in \ker N \\ \|h\|_V \le 1}} \|h\|_W \,, \tag{3.1.2}$$

where

$$\ker N = \{\, h \in V : Nh = 0 \,\}$$

denotes the nullspace of N in V.

Sketch of proof. Given information N, we show that the right-hand side of (3.1.2) is a lower bound for $r_{\text{res}}^{\text{wor}}(N)$. (That it is also an upper-bound for $r_{\text{res}}^{\text{wor}}(N)$ will follow from Theorem 3.1.2.) Let φ be an algorithm using N. Choose $h \in \ker N \cap D$ satisfying $\|h\|_V \le 1$. Since $Nh = N(-h) = 0$, we see that

$$\begin{aligned} 2e_{\text{res}}^{\text{wor}}(\varphi, N) &\ge \|h - L\varphi(Nh)\|_W + \|(-h) - L\varphi\big(N(-h)\big)\|_W \\ &= \|h - L\varphi(0)\|_W + \|h + L\varphi(0)\|_W \\ &\ge 2\|h\|_W \,. \end{aligned}$$

Since h is arbitrary, this implies that

$$e_{\text{res}}^{\text{wor}}(\varphi, N) \ge \sup_{\substack{h \in \ker N \cap D \\ \|h\|_V \le 1}} \|h\|_W \,. \tag{3.1.3}$$

Since φ is an arbitrary algorithm using N, it follows that $r_{\text{res}}^{\text{wor}}(N)$ is bounded from below by the right-hand side of (3.1.3). Finally, since D is dense in V, it may be shown that the right-hand sides of (3.1.2) and (3.1.3) are equal, which establishes that $r_{\text{res}}^{\text{wor}}(N)$ is bounded from below by the right-hand side of (3.1.2). □

Note that since V is embedded in W, the radius $r_{\text{res}}^{\text{wor}}(N)$ is finite for any information N. Hence, for any information, there always exists a finite-residual algorithm using that information. We now exhibit an algorithm with almost optimal error.

Let N be information of the form

$$Nf = \begin{bmatrix} (f, f_1) \\ \vdots \\ (f, f_n) \end{bmatrix} \qquad \forall f \in V,$$

where the inner product $(\cdot, \cdot)$ is the inner product in V, and not that of W. Without loss of generality, we may assume that $f_1, \ldots, f_n \in W$ have been chosen so that

$$(f_i, f_j) = \delta_{ij} \qquad (1 \le i, j \le n).$$

Choose $\delta > 0$. Since D is dense in U, there exist $u_1, \ldots, u_n \in U$ such that

$$\|Lu_j - f_j\|_W < \frac{\delta}{2n} \sup_{\substack{h \in \ker N \\ \|h\|_V \le 1}} \|h\|_W = \frac{\delta}{2n} r_{\text{res}}^{\text{wor}}(N) \qquad (1 \le j \le n).$$

Define the algorithm φ_δ using N by

$$\varphi_\delta(Nf) = \sum_{j=1}^{n} (f, f_j) u_j \qquad \forall f \in V.$$

We then have the following result:

THEOREM 3.1.2. *For any information N, and for any $\delta > 0$,*

$$e_{\text{res}}^{\text{wor}}(\varphi_\delta, N) < (1 + \delta) r_{\text{res}}^{\text{wor}}(N). \qquad \square$$

Thus the algorithm φ_δ is, to within a factor of $1 + \delta$, an optimal error algorithm using N. Note that this almost-optimal error algorithm is linear, which makes it easy to implement.

EXAMPLE 1.1 (*continued*). Let N be information of cardinality n for the problem of inverting the finite Laplace transform. Then there exist functions $f_1, \ldots, f_n$ such that

$$Nf = \begin{bmatrix} \int_0^1 f(s) f_1(s)\, ds \\ \vdots \\ \int_0^1 f(s) f_n(s)\, ds \end{bmatrix} \qquad \forall f \in H^r(I). \tag{3.1.4}$$

Without loss of generality, we may assume that $f_1, \ldots, f_n$ have been chosen so that

$$\int_0^1 f_i(s) f_j(s)\, ds = \delta_{ij} \qquad (1 \le i, j \le n). \tag{3.1.5}$$

Pick $\delta > 0$. Since $H^r(I)$ is dense in $L_2(I)$, there exist functions $u_1, \ldots, u_n \in L_2(I)$ such that

$$\|Lu_j - f_j\|_{L_2(I)} < \frac{\delta}{2n} \sup_{\substack{h \in \ker N \\ \|h\|_V \le 1}} \|h\|_W .$$

Then the algorithm φ_δ defined by

$$\varphi_\delta(Nf)(t) = \sum_{j=1}^{n} \left[\int_0^1 f(s) f_j(s)\, ds \right] u_j(t) \qquad (0 \le t \le 1) \quad \forall f \in H^r(I)$$

satisfies

$$e_{\text{res}}^{\text{wor}}(\varphi_\delta, N) < (1 + \delta) r_{\text{res}}^{\text{wor}}(N),$$

and so φ_δ is an almost-optimal algorithm using N.

Unfortunately, it may be difficult to explicitly construct such functions. However, this part of the algorithm is independent of any particular problem element $f \in H^r(I)$. So, finding such $u_1, \ldots, u_n$ may be considered as preprocessing that can be done before calculating $\varphi_\delta(Nf)$ for any problem element f.

Alternatively, one can start out with functions $u_1, \ldots, u_n \in L_2(I)$ and define $f_1, \ldots, f_n \in H^r(I)$ by

$$f_j(s) = (Lu_j)(s) = \int_0^1 e^{-st} u_i(t)\, dt \qquad (1 \le j \le n). \tag{3.1.6}$$

Assume without loss of generality that $u_1, \ldots, u_n$ have been chosen so that (3.1.5) holds. For information N of the form (3.1.4), with $f_1, \ldots, f_n$ given by (3.1.6), we define an algorithm φ^* using N by

$$\varphi^*(Nf)(t) = \sum_{j=1}^{n} \left[\int_0^1 f(s) f_j(s)\, ds \right] u_j(t) \qquad (0 \le t \le 1) \quad \forall f \in H^r(I).$$

Then φ^* is an optimal error algorithm using the information N. □

Knowing the optimal algorithm using given information, we seek optimal information of given cardinality n. Let

$$r_{\text{res}}^{\text{wor}}(n) = \inf\{\, r_{\text{res}}^{\text{wor}}(N) : N \text{ is information of cardinality at most } n \,\}$$

denote the n^{th} *minimal radius of information*. Information N_n of cardinality at most n is said to be n^{th} *optimal information* if

$$r_{\text{res}}^{\text{wor}}(N_n) = r_{\text{res}}^{\text{wor}}(n).$$

To do this, we recall the concept of the Gelfand n-width, see, e.g., Pinkus (1985). Let $\mathcal{A}^n$ denote the class of subspaces of W whose codimension is at most n. For a balanced convex subset X of W, the Gelfand *n-width $d^n(X, W)$ of X in W* is defined to be

$$d^n(X, W) = \inf_{A \in \mathcal{A}^n} \sup_{x \in A \cap X} \|x\|_W .$$

Furthermore, if there exists a subspace $A^n \in \mathcal{A}^n$ such that

$$d^n(X, W) = \sup_{x \in A^n \cap X} \|x\|_W ,$$

then A^n is said to be a Gelfand *extremal subspace* of codimension at most n.

Using Theorem 2.6.1 of Traub and Woźniakowski (1980), we immediately have

THEOREM 3.1.3.

(1) *The n^{th} minimal radius is given by*

$$r_{\text{res}}^{\text{wor}}(n) = d^n(BV, W),$$

where BV denotes the unit ball of V.

(2) *Suppose that A^n is a Gelfand extremal subspace of codimension at most n. Choose a basis $w_1, \ldots, w_k$ for the orthogonal complement of A^n in W, where $k = \operatorname{codim} A^n \le n$. Let $(\cdot, \cdot)_W$ denote the inner product of W. Then information N_n defined by*

$$N_n f = \begin{bmatrix} (f, w_1)_W \\ \vdots \\ (f, w_k)_W \end{bmatrix} \qquad \forall f \in V$$

is n^{th} optimal information. □

Let us say that the problem is *convergent* if $\lim_{n \to \infty} r_{\text{res}}^{\text{wor}}(n) = 0$. Using standard results on n-widths, such as Pinkus (1985), we find

COROLLARY 3.1.1. *The problem is convergent if and only if the embedding of V into W is compact.* □

We now consider convergent problems in more detail. Let E denote the embedding of V into W, which is now a *compact* dense injection. Denote the singular values of E by

$$\kappa_1 \geq \kappa_2 \geq \ldots > 0.$$

There is a complete orthonormal basis $\{v_j\}_{j=1}^{\infty}$ for V such that

$$E^*Ev_j = \kappa_j^2 v_j \qquad j = 1, 2, \ldots .$$

We then have the following

THEOREM 3.1.4. *For any non-negative integer n, define information N_n of cardinality n by*

$$N_n f = \begin{bmatrix} (f, v_1) \\ \vdots \\ (f, v_n) \end{bmatrix} \qquad \forall f \in V.$$

Then N_n is n^{th} optimal information, and

$$r_{\text{res}}^{\text{wor}}(n) = r_{\text{res}}^{\text{wor}}(N_n) = \kappa_{n+1}.$$ □

We illustrate the ideas of this subsection by their application to our example of inverting the finite Laplace transform.

EXAMPLE 1.1 (*continued*). Recall that the class of problem elements is the unit ball of $H^r(I)$, and that the (residual) error of an algorithm is being measured in the $L_2(I)$ norm. Suppose first that $r = 0$. Then the embedding of $H^r(I) = L_2(I)$ into $L_2(I)$ is not compact, and so the problem is not convergent. This means that for some positive threshold ε_0, there is no algorithm whose residual error is less than ε_0. Hence, we need only consider the case where r is a positive integer.

Since $r \geq 1$, the embedding $E\colon H^r(I) \to L_2(I)$ is compact. Integrating by parts, we find that the eigenfunctions v_j and eigenvalues κ_j^2 of E^*E are the nonzero solutions v and κ^2 of the eigenproblem

$$\sum_{i=0}^{r} (-1)^i v^{(2i)}(s) = \kappa^{-2} v(s) \qquad \forall s \in [0, 1],$$
$$v^{(i)}(0) = v^{(i)}(1) = 0 \qquad (0 \leq i \leq r-1).$$

So, we see that the information N_n defined by

$$N_n f = \begin{bmatrix} \int_0^1 f(s) v_1(s)\, ds \\ \vdots \\ \int_0^1 f(s) v_n(s)\, ds \end{bmatrix} \qquad \forall f \in H^r(I)$$

is n^{th} optimal information, and the n^{th} minimal radius is

$$r_{\text{res}}^{\text{wor}}(n) = \kappa_{n+1}.$$

Suppose first that $r = 1$. We find that

$$v_j(s) = \frac{1}{\sqrt{2\pi}} \sin j\pi s$$

and that

$$\kappa_j = \frac{1}{\sqrt{1 + j^2\pi^2}}.$$

Hence,

$$r_{\text{res}}^{\text{wor}}(n) = \frac{1}{\sqrt{1+\pi^2(n+1)^2}} \sim \frac{1}{\pi(n+1)} \qquad \text{as } n \to \infty.$$

Unfortunately, exact determination of the eigenvalues and eigenfunctions appears to be intractable for arbitrary r. However, an asymptotic result for arbitrary r may be found in, e.g., Pinkus (1985). Using the standard theta-notation of Knuth (1976), we have

$$r_{\text{res}}^{\text{wor}}(n) = \Theta(n^{-r}) \qquad \text{as } n \to \infty.$$

Moreover, it is possible to show that inner products with the basis functions of an n-dimensional spline space of piecewise polynomials of degree $r-1$ taken over a uniform discretization of I is (to within a constant factor, independent of n) n^{th} optimal information. □

3.2 The average case setting.

We now consider the average case setting for the residual error criterion. The results of this subsection are new, and so we prove them in somewhat more detail than the results of previous sections.

Since we are dealing with an average case setting, we first need a measure μ on the Hilbert space V. We suppose that μ has the properties described in Section 2.2. That is, we assume that μ is a Gaussian measure on V with zero mean and positive definite covariance operator S_μ. Moreover, we also assume that D is measurable and that $\mu(D) = 1$.

Once again, our only knowledge of a problem element f is the information Nf, where $N: V \to \mathbf{R}^n$ is a continuous linear transformation. An *algorithm* φ using N is then a mapping $\varphi: \mathbf{R}^n \to U$ such that $\|L\varphi(\cdot)\|_W^2$ is measurable. We then let

$$e_{\text{res}}^{\text{avg}}(\varphi, N) = \left[\int_D \|f - L(\varphi(Nf))\|_W^2 \, \mu(df)\right]^{1/2}.$$

denote the *error* of an algorithm φ using information N. Since the information N is continuous and V is a Hilbert space, there exist $f_1, \ldots, f_n \in V$ satisfying

$$(S_\mu f_i, f_j) = \delta_{ij} \qquad (1 \le i, j \le n)$$

such that

$$Nf = \begin{bmatrix} (f, f_1) \\ \vdots \\ (f, f_n) \end{bmatrix} \qquad \forall f \in V.$$

Our first goal is to find the minimal error among all algorithms using the information N. This is given by

$$r_{\text{res}}^{\text{avg}}(N) = \inf_{\varphi} e_{\text{res}}^{\text{avg}}(\varphi, N),$$

the *radius of information*. In addition, we wish to find an *optimal error algorithm* using N, which is an algorithm φ^* using N whose error is minimal, so that

$$e_{\text{res}}^{\text{avg}}(\varphi^*, N) = r_{\text{res}}^{\text{avg}}(N).$$

To do this, we will reduce our problem to that of approximating the identity injection E of V into W (which we call the *approximation problem* for short). For information N of the form (3.2.1), an *algorithm* for the approximation problem is a mapping $\psi: \mathbf{R}^n \to V$ such that $\|\psi(N\cdot)\|_W^2$ is measurable. The *error* of such an algorithm ψ for the approximation problem will be denoted by

$$e_{\text{app}}^{\text{avg}}(\psi, N) = \left[\int_V \|f - \psi(Nf)\|_W^2 \, \mu(df)\right]^{1/2}.$$

Let

$$r_{\text{app}}^{\text{avg}}(N) = \inf_{\psi} e_{\text{app}}^{\text{avg}}(\psi, N)$$

denote the *radius of information* for the approximation problem.

Recalling the definition of the μ-spline algorithm φ^s using information N from Section 2.2, we have

THEOREM 3.2.1. *For any information N, the following hold:*

(1) *The radii of information for the Fredholm problem and the approximation problem are the same, i.e.,*

$$r_{\text{res}}^{\text{avg}}(N) = r_{\text{app}}^{\text{avg}}(N).$$

(2) *The μ-spline algorithm φ^s is an optimal error algorithm for the Fredholm problem, i.e.,*

$$e_{\text{res}}^{\text{avg}}(\varphi^s, N) = r_{\text{res}}^{\text{avg}}(N) = r_{\text{app}}^{\text{avg}}(N).$$

Proof. Let N be information of cardinality n. We first show that

$$r_{\text{res}}^{\text{avg}}(N) \geq r_{\text{app}}^{\text{avg}}(N).$$

Let φ be an algorithm for the Fredholm problem. Then $L\varphi$ is an algorithm for the approximation problem. Since D is of full measure, we find

$$e_{\text{res}}^{\text{avg}}(\varphi, N)^2 = \int_D \|L\varphi(Nf) - f\|_W^2 \, \mu(df) \geq \inf_{\psi} \int_V \|\psi(Nf) - f\|_W^2 \, \mu(df) = r_{\text{abs}}^{\text{avg}}(N)^2.$$

Since φ is an arbitrary algorithm using N for the Fredholm problem, we may take the infimum over all such algorithms to establish the desired lower bound.

To complete the proof of the theorem, it suffices to show that

$$e_{\text{res}}^{\text{avg}}(\varphi^s, N) = r_{\text{app}}^{\text{avg}}(N).$$

We have

$$\varphi^s(Nf) = L^{-1}Pf \qquad \forall f \in V,$$

where Pf is the μ-spline interpolating f. As we pointed out in Section 2.2, Pf is in the range of L, and so the μ-spline algorithm φ^s using N is well-defined. Now Pf depends on f only through Nf. Hence we may write

$$Pf = \psi^*(Nf).$$

From the results of Wasilkowski and Woźniakowski (1986), ψ^* is an optimal error algorithm for the approximation problem, i.e.,

$$e_{\text{app}}^{\text{avg}}(\psi^*, N) = r_{\text{app}}^{\text{avg}}(N).$$

Since

$$e_{\text{res}}^{\text{avg}}(\varphi^s, N)^2 = \int_V \|f - L(\varphi^s(Nf))\|_W^2 \, \mu(df) = \int_V \|f - \psi^*(Nf)\|_W^2 \, \mu(df)$$

the desired result follows immediately. □

Hence, for any information N, the μ-spline algorithm is an optimal error algorithm using N in the average case setting, regardless of which error criterion (absolute or relative) we are using.

Knowing the optimal algorithm using given information, it is natural to seek optimal information of given cardinality n. Let

$$r_{\text{res}}^{\text{avg}}(n) = \inf\{\, r_{\text{res}}^{\text{avg}}(N) : N \text{ is information of cardinality at most } n \,\}$$

denote the n^{th} *minimal radius of information.* Information N_n of cardinality at most n is said to be n^{th} *optimal information* if

$$r_{\text{res}}^{\text{avg}}(N_n) = r_{\text{res}}^{\text{avg}}(n).$$

Moreover, an optimal error algorithm using n^{th} optimal information is said to be an n^{th} *minimal error algorithm*, since the error of such an algorithm is minimal among all algorithms using information of cardinality at most n.

The main idea that we use in finding n^{th} optimal information is to reduce the problem to the approximation problem, since Theorem 3.2.1 implies that

$$r_{\text{res}}^{\text{avg}}(n) = r_{\text{app}}^{\text{avg}}(n).$$

That is, the n^{th} minimal radii for the problems are the same, and the same information is optimal information for both problems.

Let $\sigma_1 \geq \sigma_2 \geq \ldots > 0$ be the eigenvalues of the covariance operator S_μ. Let $e_1, e_2, \ldots$ denote the corresponding eigenvectors, which we assume to be orthonormal without loss of generality. Of course,

$$\operatorname{tr} S_\mu = \sum_{i=1}^{\infty} \sigma_i < \infty.$$

Let

$$f_i = \sigma_i^{-1/2} e_i \qquad i = 1, 2, \ldots .$$

Then

$$(S_\mu f_i, f_j) = \delta_{ij} \qquad i, j = 1, 2, \ldots .$$

Define the information N_n to be

$$N_n f = \begin{bmatrix} (f, f_1) \\ \vdots \\ (f, f_n) \end{bmatrix} \qquad \forall f \in V.$$

We also define an algorithm φ_n using N_n by

$$\varphi_n(N_n f) = \sum_{i=1}^{n} (f, e_i) L^{-1} e_i .$$

Note that the algorithm φ_n is well-defined, since the eigenvectors of the covariance operator lie in the range of S_μ, which is a subset of the range of L^{-1}.

Using the results of Wasilkowski and Woźniakowski (1986), along with Theorem 3.2.1, we immediately find

THEOREM 3.2.2. *For any non-negative integer n, the information N_n is n^{th} optimal information and the algorithm φ_n is an n^{th} minimal error algorithm. That is,*

$$e_{\text{res}}^{\text{avg}}(\varphi_n, N_n) = r_{\text{res}}^{\text{avg}}(N_n) = r_{\text{res}}^{\text{avg}}(n) = \left[\sum_{i=n+1}^{\infty} \sigma_i \right]^{1/2}. \qquad \square$$

Since the covariance operator has finite trace, Theorem 3.2.3 implies that

$$\lim_{n \to \infty} r_{\text{res}}^{\text{avg}}(n) = 0.$$

That is, for any positive error tolerance ε, there exists an algorithm whose error is at most ε. Hence the problem is always convergent when using the residual error criterion in the average case setting.

4. COMPLEXITY ANALYSIS

The previous sections of this paper dealt with optimal error algorithms for the Fredholm problem of the first kind. In this section, we use these results to find the ε-complexity of such problems, i.e., the minimal cost of finding an approximation whose error is at most ε. Of course, we expect the complexity to depend on the error criterion and the setting. Moreover, it will also depend on the model of computation used, which defines the cost of any algorithm.

Our first step in defining our model of computation is to decide how much to charge for basic operations. As in Traub and Woźniakowski (1980), we assume that

(1) the cost of evaluating any linear functional is $c > 0$, and
(2) any arithmetic operation has unit cost.

Next, we determine the cost of evaluating an algorithm φ using information N at a particular problem element f as follows. Let $\mathrm{cost}(Nf)$ denote the cost of evaluating the information Nf, and let $\mathrm{cost}(\varphi, Nf)$ denote the cost of combining this information to find $\varphi(Nf)$. Then

$$\mathrm{cost}(\varphi, N, f) = \mathrm{cost}(Nf) + \mathrm{cost}(\varphi, Nf).$$

That is, the cost of evaluating an algorithm using given information at a particular problem element is the sum of its informational cost and its combinatory cost. Finally, we define the cost of an algorithm φ using information N to be

$$\mathrm{cost}(\varphi, N) = \sup_{\substack{f \in D \\ \|f\|_F \le 1}} \mathrm{cost}(\varphi, N, f).$$

We can now define the complexity of the problem. Fix a particular error criterion (absolute or residual) and setting (worst or average case). This, of course, determines the error $e(\varphi, N)$ of an algorithm φ using information N. For $\varepsilon > 0$, we define the *ε-complexity* of the problem to be

$$\mathrm{comp}(\varepsilon) = \inf\{\, \mathrm{cost}(\varphi, N) : \varphi \text{ and } N \text{ such that } e(\varphi, N) \le \varepsilon \,\}.$$

Note that there are four ways of defining $\mathrm{comp}(\varepsilon)$, depending on the error criterion and the setting.

REMARK 4.1. Note that we are using a worst case model of computation, i.e., $\mathrm{cost}(\varphi, N)$ is given by a worst case of $\mathrm{cost}(\varphi(Nf))$ over all problem elements f. One can also consider an average case model of computation, in which we take

$$\mathrm{cost}(\varphi, N) = \int_D \mathrm{cost}(\varphi, N, f)\, \mu(df).$$

Since we are dealing with information of *fixed* cardinality (i.e., the cardinality of the information is independent of the problem element), it is easy to see that the results of this section are the same for either model of computation. From the results of Wasilkowski (1986), we find that this will also be true for information of *varying* cardinality (in which the cardinality can vary with the problem element). □

Our main goal is to find conditions that are necessary and sufficient for $\mathrm{comp}(\varepsilon)$ to be finite for all $\varepsilon > 0$. To do this, we define the *ε-cardinality number* to be

$$m(\varepsilon) = \inf\{\, n : r(n) \le \varepsilon \,\}.$$

Then

$$\mathrm{comp}(\varepsilon) \ge c\, m(\varepsilon).$$

Moreover, if there exists a *linear* $m(\varepsilon)^{\text{th}}$ minimal error algorithm, then

$$\mathrm{comp}(\varepsilon) \le (c+2)m(\varepsilon) - 1.$$

Since $c \gg 1$ in practice, we see that the ε-complexity is roughly equal to $c\, m(\varepsilon)$. Of course, $m(\varepsilon)$ will depend on the setting and the error criterion, and so there will be four different ε-cardinality numbers.

We first look at the absolute error criterion. For the worst case setting, it is easy to see that the results of Section 2.1 imply that $m^{\mathrm{wor}}_{\mathrm{abs}}(\varepsilon) = +\infty$ for any $\varepsilon > 0$. This immediately gives us

THEOREM 4.1. *For the worst case setting in the absolute error criterion,*

$$\mathrm{comp}^{\mathrm{wor}}_{\mathrm{abs}}(\varepsilon) = +\infty \qquad \forall\, \varepsilon > 0.$$ □

Hence, it is impossible to find an ε-approximation using the absolute error criterion in the worst case setting, no matter how large ε is.

EXAMPLE 1.1 (*continued*). For the problem of inverting the finite Laplace transform with problem elements in the unit ball of $H^r(I)$, we have $\mathrm{comp}^{\mathrm{wor}}_{\mathrm{abs}}(\varepsilon) = +\infty$ for any $\varepsilon > 0$. This means that it is impossible to find a finite-error approximation to this problem with finite cost. Note that this result is true for all r, no matter how large (where, as before, r denotes the smoothness of the class of problem elements). □

Next, we turn to the average case setting:

THEOREM 4.2. *For the average case setting in the absolute error criterion, the following alternatives hold:*
(1) *If L^{-1} is not bounded on the average, then for any $\varepsilon > 0$,*

$$\mathrm{comp}_{\mathrm{abs}}^{\mathrm{avg}}(\varepsilon) = +\infty.$$

(2) *If L^{-1} is bounded on the average, then*

$$m_{\mathrm{abs}}^{\mathrm{avg}}(\varepsilon) = \inf\left\{ n : \sum_{i=n+1}^{\infty} \kappa_i \le \varepsilon^2 \right\},$$

where κ_i is the i^{th}-largest eigenvalue of the Hilbert-Schmidt operator $\left(L^{-1}S_\mu^{1/2}\right)^* \left(L^{-1}S_\mu^{1/2}\right)$. Hence,

$$c\, m_{\mathrm{abs}}^{\mathrm{avg}}(\varepsilon) \le \mathrm{comp}_{\mathrm{abs}}^{\mathrm{avg}}(\varepsilon) \le (c+2) m_{\mathrm{abs}}^{\mathrm{avg}}(\varepsilon) - 1 \qquad \forall \varepsilon > 0. \qquad \square$$

Thus the ε-complexity is finite for all $\varepsilon > 0$ if and only if L^{-1} is bounded on the average. When this holds, we can find an almost-optimal complexity algorithm as follows. Let $n = m_{\mathrm{abs}}^{\mathrm{avg}}(\varepsilon)$, and let N_n be n^{th} optimal information. (Of course, if N_n does not exist, we can use the approximation techniques of Section 2.2.) Then the μ-spline algorithm using N_n is an almost-optimal complexity algorithm.

We now turn to the residual error criterion. Recall that E denotes the embedding of V into W. In the worst case setting, we have

THEOREM 4.3. *The following alternatives hold for the worst case setting using the residual error criterion:*
(1) *If E is not compact, then there exists $\varepsilon_0 > 0$ such that*

$$\mathrm{comp}_{\mathrm{res}}^{\mathrm{wor}}(\varepsilon) = +\infty \qquad \text{if } 0 < \varepsilon \le \varepsilon_0.$$

(2) *If E is compact, then*

$$m_{\mathrm{res}}^{\mathrm{wor}}(\varepsilon) = \inf\{ n : \kappa_{n+1} \le \varepsilon \},$$

where κ_n is the n^{th}-largest singular value of E. Hence,

$$c\, m_{\mathrm{res}}^{\mathrm{wor}}(\varepsilon) \le \mathrm{comp}_{\mathrm{res}}^{\mathrm{wor}}(\varepsilon) \le (c+2) m_{\mathrm{res}}^{\mathrm{wor}}(\varepsilon) - 1 \qquad \forall \varepsilon > 0. \qquad \square$$

Thus the ε-complexity is finite for all $\varepsilon > 0$ if and only if V is compactly embedded in W. In that case, we can find an almost-optimal complexity algorithm as follows. Let $n = m_{\mathrm{res}}^{\mathrm{wor}}(\varepsilon)$, and let N_n be n^{th} optimal information. Then a linear optimal error algorithm using N_n will be an almost-optimal complexity algorithm.

EXAMPLE 1.1 (*continued*). We find that for our problem of inverting the finite Laplace transform with problem elements in the unit ball of $H^r(I)$,

$$\mathrm{comp}_{\mathrm{res}}^{\mathrm{wor}}(\varepsilon) = \Theta(\varepsilon^{-1/r}) \qquad \text{as } \varepsilon \to 0$$

if $r > 0$. Of course, when $r = 0$, then the problem is not convergent. Hence,

$$\mathrm{comp}_{\mathrm{res}}^{\mathrm{wor}}(\varepsilon) = +\infty \qquad \text{for sufficiently small } \varepsilon > 0$$

if $r = 0$. $\square$

Finally, we consider the average case setting:

THEOREM 4.4. *For the average case setting using the residual error criterion, the ε-cardinality number is given by*

$$m_{\mathrm{res}}^{\mathrm{avg}}(\varepsilon) = \inf\left\{ n : \sum_{i=n+1}^{\infty} \sigma_i \le \varepsilon^2 \right\},$$

where σ_i is the i^{th}-largest eigenvalue of the covariance operator S_μ. Hence, the ε-complexity is given by

$$c\, m_{\mathrm{res}}^{\mathrm{avg}}(\varepsilon) \le \mathrm{comp}_{\mathrm{res}}^{\mathrm{avg}}(\varepsilon) \le (c+2) m_{\mathrm{res}}^{\mathrm{avg}}(\varepsilon) - 1 \qquad \forall \varepsilon > 0. \qquad \square$$

In particular, this means that the ε-complexity is finite for *all* $\varepsilon > 0$. We find an almost-optimal complexity algorithm as follows. Let $n = m_{\mathrm{res}}^{\mathrm{avg}}(\varepsilon)$, and let N_n be n^{th} optimal information. Then the μ-spline algorithm using N_n will be an almost-optimal complexity algorithm.

Acknowledgements

I am grateful to H. Woźniakowski for his valuable comments on this paper.

References

Carasso, A. and Stone, A. (1975), "Improperly Posed Boundary Value Problems," Pitnam, London.

Ciarlet, P. G. (1978), "The Finite Element Method for Elliptic Problems," North-Holland, Amsterdam.

Dunn, H. S. (1967), A generalization of the Laplace transform, *Proc. Camb. Phil. Soc.* 63 (1967), 155–160.

Gehatia, M. and Wiff, D. R. (1970), Solution of Fujita's equation for equilibrium sedimentation by applying Tikhonov's regularizing functions, *J. Polymer Science (Part A-2)* 8, 2039–2049.

Hadamard, J. (1952), "Lectures on the Cauchy Problem in Linear Partial Differential Equations," Dover, New York.

Hämmerlin, C. and Hoffmann, K.-H. (1983), "Improperly Posed Problems and Their Numerical Treatment," International Series of Numerical Mathematics 63, Birkhäuser-Verlag, Basel.

Knuth, D. E. (1976), Big omicron and big omega and big theta, *SIGACT News* (April, 1976), 18–24.

Kuo, H.-H. (1975), "Gaussian Measures in Banach Spaces," Lecture Notes in Mathematics 463, Springer-Verlag, Berlin.

Marti, J. T. (1983), Numerical solution of Fujita's equation, in "Impropoerly Posed Problems and Their Numerical Treatment," (G. Hämmerlin and K.-H. Hoffmann, eds.), Birkhäuser Verlag, Basel.

Pinkus, A. (1985), "n-Widths in Approximation Theory," Springer-Verlag, Berlin.

Poggio, T., Torre, V., and Koch, C. (1985), Computational vision and regularization theory, *Nature* 317, 314–319.

Skorohod, A. V. (1974), "Integration in Hilbert Space," Springer-Verlag, New York.

Tikhonov, A. N. (1963), Solution of incorrectly formulated problems and the regularization method, *Sov. Math. Dokl.* 4, 1036–1038.

Tikhonov, A. N. and Arsenin, V. Y. (1977), "Solutions of Ill-Posed Problems," V. H. Winston and Sons, Washington.

Traub, J. F. and Woźniakowski, H. (1980), "A General Theory of Optimal Algorithms," Academic Press, New York.

Twomey, S. (1977), "Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements," Developments in Geomathematics 3, Elsevier Scientific Publ., Amsterdam.

Wahba, G. (1985), A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem, *Ann. Statist.* 13, 1378–1402.

Wahba, G. (1986), Three topics in ill-posed problems, Technical Report, Department of Statistics, University of Wisconsin, Madison.

Wasilkowski, G. W. (1986), Information of varying cardinality, *J. Complexity* 2, 204–228.

Wasilkowski, G. W. and Woźniakowski, H. (1986), Average case optimal algorithms in Hilbert spaces, *J. Approx. Th.* 47, 17–25.

Werschulz, A. G. (1986), What is the complexity of ill-posed problems?, Research Report, Department of Computer Science, Columbia University, New York.

Werschulz, A. G. (1987), Optimal residual algorithms for linear operator equations, (in preparation).

Woźniakowski, H. (1986), Information-based complexity, in "Annual Review of Computer Science," Vol. 1, pp. 319–380, Annual Reviews, Inc., Palo Alto, CA.