E. WARD CHENEY
University of Texas at Austin

Multivariate Approximation Theory: Selected Topics

SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS
PHILADELPHIA, PENNSYLVANIA 1986

Library of Congress Catalog Card Number 86-61533.

ISBN 0-89871-207-6.

Printed for the Society for Industrial and Applied Mathematics by Capital City Press, Montpelier, Vermont.

CHAPTER 6
Algorithms

In this subject, we mean by an **algorithm** any procedure or "recipe" for producing approximations. In this broad sense, Lagrange interpolation, Shepard interpolation, and orthogonal expansions are algorithms. Indeed, any projection operator can be regarded as an algorithm. Projections, of course, do not generally produce *best* approximations; algorithms for the latter are usually more difficult to invent, and are less efficient computationally.

At the practical level, an algorithm must be judged by such criteria as ease of programming, speed of computation, stability and accuracy of computation. Two algorithms that are equally attractive in theory may differ markedly in their numerical properties. For an example of this, consider two projections into Π_n, one given by Lagrange interpolation and another given by an orthogonal expansion:

$$Px = \sum_{i=0}^{n} x(s_i)\ell_i, \qquad \ell_i(s_j) = \delta_{ij}, \tag{1}$$

$$Qx = \sum_{i=0}^{n} \langle x, u_i \rangle u_i, \qquad \langle u_i, u_j \rangle = \delta_{ij}. \tag{2}$$

Theoretically, these are quite similar: each involves a basis for Π_n and an accompanying system of "biorthogonal" functionals. Notice, however, that a functional of the type $x \mapsto \langle x, u_i \rangle$ is much more difficult to compute than one of the type $x \mapsto x(s_i)$, since the former involves an integration. Numerically such an integration will be effected with a quadrature formula, which, in turn, will be based upon functionals of the form $x \mapsto x(s_i)$.

Algorithms for best approximation will usually be iterative in nature, and each step in the procedure will often involve roughly as much work as a single linear projection process. Thus, the Remez Second Algorithm can be interpreted as a sequence of Lagrange interpolation processes, or as a sequence of linear projections similar to ordinary interpolation. The major exception to these general observations occurs in the case of quadratic norms – the Hilbert-space case. Here, the best approximation operator onto a linear subspace is a linear projection. In the finite-dimensional case, it is of the type illustrated by the operator Q in Eq. (2).

Let us proceed now to subspaces in tensor product spaces. We have already seen that, for producing best approximations in a subspace $C(S) \otimes H$,

when H is a subspace of $C(T)$, there will be no theoretical difficulties if H is a finite-dimensional *Haar* subspace (see Chapter 5). Thus, if A is the proximity map of $C(T)$ onto H, then one proximity map of $C(S \times T)$ onto $C(S) \otimes H$ is defined by

$$(\bar{A}f)(s,t) = (Af_s)(t).$$

In terms of a basis $\{h_1, \cdots, h_m\}$ for H, $\bar{A}$ is given by

$$(\bar{A}f)(s,t) = \sum_{j=1}^{m} x_j(s)\, h_j(t)$$

where, for each s, $\sum_{j=1}^{m} x_j(s)h_j$ is the unique best approximation of f_s in H. The theory assures us that the functions x_j determined point-by-point in this manner are continuous.

In a practical realization of this algorithm, one could select a discrete subset $\{s_1, \cdots, s_N\}$ in S, and determine the functions x_j only on this discrete set. Then, by a spline-interpolation procedure, the x_j-functions could be extended to all of S. The resulting function will not be a best approximation, but a near substitute for it. If this strategy has been adopted, then the resulting approximation will lie in a subspace $G \otimes H$, where G is now a finite-dimensional subspace of splines having knots $s_1, \cdots, s_N$. In this roundabout way, then, we are led again to consider the general question of best approximation in $C(S \times T)$ by elements of a prescribed subspace $G \otimes H$, both G and H being finite-dimensional. Since $G \otimes H$ is also finite-dimensional, the exchange algorithm or the Remez First Algorithm can be used.

It should be noted that the operator $\bar{A}$ defined in the previous paragraphs is not the only proximity map of $C(S \times T)$ onto $C(S) \otimes H$. The subspace $C(S) \otimes H$ never has the Chebyshev property (except in degenerate cases), and so there is much latitude in the choice of best approximations. The map $\bar{A}$ selects the optimal one of these, since it minimizes the norm of each section $f_s - (Af)_s$. It is not necessary to do this. The theorem governing this situation is as follows (Franchetti and Cheney [1981-a]).

THEOREM. *Let H be a subspace of a Banach space Y, f an element of $C(S,Y)$, and g an element of $C(S,H)$. In order that g be a best approximation of f in $C(S,H)$ it is necessary and sufficient that there exist a point s_0 such that*

$$\|f - g\| = \operatorname{dist}\big(f(s_0), H\big).$$

Proof. A lemma of Buck [1974] asserts that

$$\operatorname{dist}\big(f, C(S,H)\big) = \sup_s \operatorname{dist}\big(f(s), H\big). \tag{3}$$

Let s_0 be a point of S such that

$$\operatorname{dist}\big(f(s_0), H\big) = \operatorname{dist}\big(f, C(S,H)\big).$$

If g is a best approximation of f then

$$\|f - g\| = \operatorname{dist}\big(f, C(S,H)\big) = \operatorname{dist}\big(f(s_0), H\big).$$

For the converse, let σ be a point such that $\|f - g\| = \operatorname{dist}\big(f(\sigma), H\big)$. By (3), $\|f - g\| \leq \operatorname{dist}\big(f, C(S,H)\big)$. □

It is clear that if g and s_0 have the property

$$\|f - g\| = \operatorname{dist}\big(f(s_0), H\big),$$

then $g(s_0)$ is a best approximation of $f(s_0)$ in H. Thus, in principle, there need be only one s_0 where $g(s_0)$ is a solution to an extremal problem. For all the other points it is merely necessary to secure the weaker requirement

$$\|f(s) - g(s)\| \leq \|f(s_0) - g(s_0)\|.$$

For the more interesting and versatile subspaces

$$C(S) \otimes H + G \otimes C(T),$$

there exist no general algorithms for producing minimizing sequences. The sole exception to this is the Diliberto-Straus algorithm for producing best approximations in

$$W_0 = C(S) + C(T).$$

Their algorithm proceeds as follows. Let f be an element of $C(S \times T)$ whose best approximation in $C(S) + C(T)$ is sought. A process of "alternating proximity maps" is used. Observe that we can construct best approximations of f in $C(S)$ and $C(T)$ by these processes:

$$(Af)(s,t) = \tfrac{1}{2} \max_{\sigma \in S} f(\sigma, t) + \tfrac{1}{2} \min_{\sigma \in S} f(\sigma, t), \tag{4}$$

$$(Bf)(s,t) = \tfrac{1}{2} \max_{\tau \in T} f(s, \tau) + \tfrac{1}{2} \min_{\tau \in T} f(s, \tau). \tag{5}$$

It is clear that Af is really a function of t alone, or, more precisely, is constant in s. Thus $Af \in C(T)$. Similarly, $Bf \in C(S)$. The alternating process constructs the sequence

$$f_0 = f, \quad f_{2n+1} = f_{2n} - Af_{2n}, \quad f_{2n+2} = f_{2n+1} - Bf_{2n+1}. \tag{6}$$

By successively subtracting from f certain elements of $C(S)$ and $C(T)$ we can expect to reduce the norm of the residual functions f_n to an absolute minimum. The first significant result in the theory asserts that, indeed,

$$\lim_{n \to \infty} \|f_n\| = \operatorname{dist}(f, W_0).$$

This was proved already by Diliberto and Straus. The second significant result was proved by Aumann [1958], who was apparently unaware of Diliberto and Straus' work. He established the convergence of the sequence $[f_n]$. Since the definition of the algorithm ensures that $f - f_n \in W_0$ for all n, and since W_0 is closed, it follows that $f - \lim f_n$ is a best approximation of f in W_0. The algorithm therefore provides a constructive proof that W_0 is proximinal.

Golomb [1959] noticed that the theory of Diliberto and Straus could be applied to any pair of proximity maps A and B in any normed space provided that a certain property called "centrality" was possessed by A and B. That property, for A, is expressed by the equation

$$\|x - Ax + Ay\| = \|x - Ax - Ay\|,$$

assumed valid for all x and y. The particular maps A and B in Eq. (4) and (5) are indeed central proximity maps. Also, any orthogonal projection on an inner-product space is a central proximity map. If a central proximity map $B : Y \rightarrow\!\!\!\!\rightarrow H$ is given, then a central proximity map $\bar{B} : C(S,Y) \rightarrow\!\!\!\!\rightarrow C(S,H)$ is constructed by the definition

$$\bar{B}f = B \circ f.$$

This construction is, in fact, used in Eqs. (4) and (5).

The main theorem in Golomb's theory states that if $A : X \rightarrow\!\!\!\!\rightarrow U$ and $B : X \rightarrow\!\!\!\!\rightarrow V$ are central proximity maps and if $U + V$ is closed, then the Diliberto-Straus algorithm given in Eq. (6) produces a sequence $[f_n]$ such that

$$\|f_n\| \downarrow \operatorname{dist}(f, U + V).$$

This is valid in any Banach space X. In order to draw the conclusion that the sequence $[f_n]$ converges, one can add uniform convexity of X as a hypothesis. These matters are expounded in Chapter 4 of Light and Cheney [1985].

A paper by Light [1980] reveals that central proximity maps are rare. For example, a Haar subspace of dimension 2 or more in $C[a,b]$ does not have a central proximity map. The constant functions in $L_p[a,b]$ have a central proximity map only if $p = 2$.

Further negative results were established by Dyn [1980] and by von Golitschek and Cheney [1983-b]. The work of these three authors shows that if G and H are Haar subspaces in $C(I)$ with $I = [a,b]$, $\dim G \geq 1$, and $\dim H \geq 2$, then the Diliberto-Straus algorithm fails when applied to the subspaces $G \otimes C(I)$ and $H \otimes C(I)$. This means that the sequence $[f_n]$ obtained by alternately subtracting best approximations may have the property

$$\lim \|f_n\| > \operatorname{dist}(f, W)$$

where W is the vector sum of $G \otimes C(I)$ and $C(I) \otimes H$.

If the topological space T is a disjoint union of m closed sets, then the functions which are constant on these closed sets form an m-dimensional subspace H in $C(T)$. One can visualize this as a subspace of 0-degree spline functions. Such a subspace possesses a central proximity map. Figure 2 shows how this map is defined.

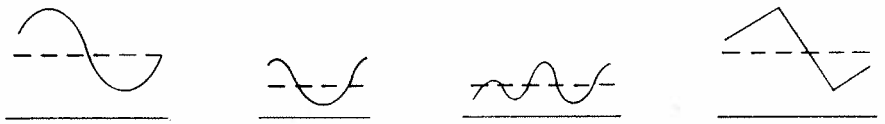

Fig. 2.

If S is also disconnected and has, say, n components, then a similar subspace G of dimension n exists in $C(S)$. The generalized Diliberto-Straus algorithm now is completely successful. With each $f \in C(S \times T)$ we associate a convergent sequence $[f_k]$ such that $f - \lim f_k$ is a best approximation of f in

$$Z = C(S) \otimes H + G \otimes C(T).$$

Therefore the algorithm gives a constructive proof of proximinality of Z. These matters have been dealt with by Respess and Cheney [1982-b].

Because of its natural and elegant character, the algorithm we have called by the names of Diliberto and Straus has been used in many diverse situations. Its use in Hilbert space predates the work of Diliberto and Straus by almost 20 years, having been exploited by von Neumann to obtain the orthogonal projection onto the closure of the sum of two subspaces. Von Neumann proved that if P and Q are orthogonal projections of a Hilbert space onto subspaces U and V, respectively, then the projection of an element f onto $\overline{U+V}$ is given by $f - \lim f_n$, where $[f_n]$ comes from the algorithm. In this situation, the algorithm has been known as the **alternating algorithm.** The formulae are

$$f_0 = f, \quad f_{2n+1} = f_{2n} - Pf_{2n}, \quad f_{2n+2} = f_{2n+1} - Qf_{2n+1}.$$

These are, of course, the same as those used in the Diliberto-Straus algorithm. In Hilbert space, the algorithm is effective for any pair of closed subspaces, while in $C(S)$ or $C(S \times T)$ it is effective only for very special pairs of subspaces. More recent work by Sullivan [1975], Atlestam and Sullivan [1976], Deutsch [1984], and Franchetti and Light [1984] has led to various generalizations of von Neumann's theorem, such as this one:

THEOREM. *Let X and its conjugate be uniformly convex Banach spaces. If U, V and $U + V$ are closed subspaces in X, then the alternating algorithm produces best approximations in $U + V$.*

A new algorithm having many favorable features has recently been developed by von Golitschek [1984]. His algorithm finds best approximations to elements of $C(D)$, $D \subset S \times T$, by functions of the form

$$z(s,t) = \phi[x(s)h(t) + y(t)g(s)] \tag{7}$$

in which ϕ, g, and h are prescribed continuous functions, and the functions x and y are at our disposal. It is assumed that

$$g \in C(S), \quad h \in C(T), \quad \phi \in C(\mathbf{R}), \quad g > 0, \quad h > 0, \quad \text{and} \quad \phi^{-1} \in C(\mathbf{R}).$$

We will outline von Golitschek's results as they apply to the simpler problem

$$f(s,t) \approx x(s) + y(t),$$

for in this case the analogy to the Diliberto-Straus algorithm will be more apparent. This procedure is iterative and uses the following formulae, in which α is a parameter lying in the interval $0 \leq \alpha \leq \|f\|$. Ideally, we would set $\alpha = \operatorname{dist}(f, W_0)$, where $W_0 = C(S) + C(T)$.

$$\begin{cases} x_0(s) = 0 & y_0(t) = \inf\limits_s \left[f(s,t) + \alpha\right] \\ x_n(s) = x_{n+1}(s) \vee \sup\limits_t \left[f(s,t) - \alpha - y_{n-1}(t)\right]. & \text{If } \|x_n\| > 4\|f\|, \text{ STOP.} \\ y_n(t) = y_{n-1}(t) \wedge \inf\limits_s \left[f(s,t) + \alpha - x_n(s)\right]. & \text{If } y_n = y_{n-1}, \text{ STOP.} \end{cases}$$

In these formulae $\vee$ and $\wedge$ denote the larger and the smaller of two numbers. It is already clear from these formulae that

$$0 = x_0 \leq x_1 \leq x_2 \leq \cdots \quad \text{and} \quad y_0 \geq y_1 \geq y_2 \geq \cdots.$$

Another favorable property of the functions x_n and y_n is that they have the same modulus of continuity as f; thus the sequences $[x_n]$ and $[y_n]$ are equicontinuous. The *uniform convergence* of these sequences will therefore depend solely upon whether they are bounded in $C(S)$ and $C(T)$. Here are the three principal results concerning the algorithm:

THEOREM 1. *If the inequality $\alpha > \operatorname{dist}(f, W_0)$ is true then the iteration terminates with $y_n = y_{n-1}$ for some n. When this occurs, $\|f - x_n - y_n\| \leq \alpha$.*

THEOREM 2. *The inequality $\alpha < \operatorname{dist}(f, W_0)$ is true if and only if the iteration terminates with $\|x_n\| > 4\|f\|$ for some n.*

THEOREM 3. *If $\alpha = \operatorname{dist}(f, W_0)$, then either the algorithm terminates as in Theorem 1 or else it produces a sequence $[x_n + y_n]$ that converges uniformly to a best approximation of f in W_0.*