JOURNAL OF APPROXIMATION THEORY 12, 61–69 (1974)

Uniform Piecewise Polynomial Approximation with Variable Joints*

T. PAVLIDIS AND A. P. MAIKA

Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08540

Communicated by John R. Rice

A procedure based on functional iteration is proposed for solving the problem of uniform piecewise polynomial approximation of a function, and a proof for its convergence is given. This method compares favorably with Lawson's algorithm, especially in cases in which the given function is not of smooth form.

1. INTRODUCTION

The problems of splines with variable knots and piecewise polynomial approximation with variable joints have received considerable attention in the literature [1–11]. The problem of optimization when the knots or joints are variable is a nonlinear one and no general solution to it is known. Nevertheless, the problem is of considerable practical importance since, to quote Rice, "the key to the successful use of splines is to have the location of the knots as variables" [1, Vol. 2, p. 123]. The same statement could be made about piecewise polynomial approximations. This paper presents a solution to the problem of finding the optimum piecewise approximating polynomial. It is based on functional iteration and it reduces the problem to that of finding the zero of a vector valued function.

Removing the continuity conditions at the break points not only simplifies the problem but it is also desirable for the following reasons: In many applications, and in particular feature selection for pattern recognition and picture processing, continuity conditions are not very meaningful since one is often faced with large changes in both amplitude and derivative of the input function. Furthermore, speed of computation can be very important in the same applications, and finding the approximation separately on each interval reduces significantly the computational complexity of the problem besides making it easily amenable to parallel computation.

* This work was supported by NSF Grant GK36180.

Let $f(x)$ be a given function defined on an interval $[\alpha, \beta]$ and let $\{x_i\}_{i=0}^{n}$ be dividing points such that:

$$\alpha = x_0 < x_1 < \cdots < x_n = \beta. \tag{1}$$

Then on the mth segment $(x_{m-1}, x_m]$, $f(x)$ is approximated by a polynomial $p_m(x)$. The approximation is called optimal if the coefficients of $p_m(x)$ $(m = 1, 2,..., n)$ and the points $x_m (m = 1, 2,..., n - 1)$ are chosen in such a way as to minimize

$$e = \max_m e_m , \tag{2}$$

where e_m is the error norm on $(x_{m-1}, x_m]$. We restrict ourselves to the case of uniform approximation, i.e., when

$$e_m = \max_x | f(x) - p_m(x)|. \tag{3}$$

Such a choice of a norm combined with the use of variable joints preserves local features. For example, consider the case in which

$$f(x) = \cos kx \text{ for } x \in [0, 0.5), (0.5 + \epsilon, 1]$$

and

$$f(x) = 2 + \cos kx \text{ for } x \in [0.5, 0.5 + \epsilon].$$

For a sufficiently high value of k in comparison to ϵ an approximation of $f(x)$ by a piecewise constant function with 3 break points will "miss" the pulse if an integral square error norm is used. This will not happen with uniform approximation.

In certain applications the optimization should be carried with respect to n, by finding its minimum value for a given error tolerance e. However, this problem requires the solution of the former and in the sequence we will assume that n is given.

2. Properties of e_m

We assume that $f(x)$ is continuous and that it satisfies a Lipschitz condition on $[\alpha, \beta]$. Some of the properties listed below hold even if these assumptions are relaxed.

It can be shown [8, 10] that:

(1) e_m is a continuous function of x_{m-1} and x_m .
(2) e_m is nonincreasing in x_{m-1} .
(3) e_m is nondecreasing in x_m .

We have:

THEOREM 1. *e_m satisfies a Lipschitz condition with respect to both x_{m-1} and x_m.*

Proof. Consider an approximation over $[x_0, x]$ and express the minimum error norm as a function of the right endpoint:

$$d(x) = \min_p \| f - p \|_{[x_0, x]}.$$

Let p^* be the optimizing polynomial on $[x_0, x_1]$. Then for $x_2 > x_1$

$$d(x_2) - d(x_1) = \min_p \| f - p \|_{[x_0, x_2]} - \| f - p^* \|_{[x_0, x_1]}$$

or

$$d(x_2) - d(x_1) \leqslant \| f - p^* \|_{[x_0, x_2]} - \| f - p^* \|_{[x_0, x_1]}.$$

The right-hand side is obviously not greater than $M(x_2 - x_1)$ for some M and therefore,

$$d(x_2) - d(x_1) \leqslant M(x_2 - x_1).$$

A similar argument can be made for the left endpoint. Let Δe_m denote the change in the error norm of the optimal uniform approximation when an endpoint is moved. Let Δx_m denote a change in the location of x_m. Then Theorem 1 together with properties (2) and (3) imply:

$$-M \leqslant \frac{\Delta e_m}{\Delta x_{m-1}} \leqslant 0 \qquad \text{(nonincreasing)} \tag{4a}$$

$$0 \leqslant \frac{\Delta e_m}{\Delta x_m} \leqslant M \qquad \text{(nondecreasing)} \tag{4b}$$

The above inqualities are critical in the subsequent development. As a matter of fact, most of the results are valid not only on uniform piecewise polynomial approximation but any other type as long as the dependence of the error norm on the endpoints conforms with Eq. (4) and the goal is to minimize the error norm *on each* interval.

3. BALANCED ERROR SOLUTION

It can be easily shown [8, 10] that if

$$e_1 = e_2 = \cdots = e_n,$$

then the solution is optimal, and that an optimal solution with this property always exists. Such a solution is called a balanced error solution.

Lawson [10] has proposed an iterative algorithm for finding such a solution; let the superscript k indicate the iteration. Then in the case of piecewise q order polynomial approximation, assume that the error is related to length of the interval by an equation of the form:

$$e_m{}^k = c_m{}^k(s_m{}^k)^{q+1}$$

where

$$s_m{}^k = x_m{}^k - x_{m-1}^k . \tag{5}$$

At each step the variables $s_m{}^k$ and $e_m{}^k$ are known, thus Eq. (5) can be solved with respect to $c_m{}^k$. A balanced error solution could be achieved if $c_m{}^k$ was indeed a constant and the lengths were chosen according to the following relations:

$$c_m{}^k(s_m^{k+1})^2 = E, \qquad m = 1,..., n, \tag{6}$$

where E is a constant obtained from the length invariance of the sum of intervals [10]

$$E = \left(\frac{\beta - \alpha}{\sum_m \frac{1}{\sqrt[q+1]{c_m{}^k}}} \right)^{q+1}. \tag{7}$$

Because Eq. (5) is only an approximation, it is necessary to proceed through successive iterations.

The disadvantage of this method is that if the error $e_m{}^k$ on some interval is zero (or even much smaller than on all the others) then Eq. (5) gives $c_m{}^k = 0$ (or close to zero) and then Eq. (6) either cannot be solved or s_m^{k+1} tends to $\beta - \alpha$. This is a rather serious defect since it is not unlikely that a function has an almost linear section. Practical experience with various types of data has shown that such failures can be quite common [12, 13]. Although the method could be modified to avoid the "catastrophes" when $e_m{}^k = 0$, it may still not converge.

4. A Functional Iteration Method

The following is a scheme based on functional iteration of the form [14]:

$$\mathbf{x}^{k+1} = \mathbf{h}(\mathbf{x}^k). \tag{8}$$

Specifically, for $k = 0, 1, 2,....$

$$x_m^{k+1} = x_m{}^k + c(e_{m+1}^k - e_m{}^k) \qquad m = 1,..., (n-1), \tag{9a}$$

$$x_n^{k+1} = x_n{}^k. \tag{9b}$$

Where x_m^0 $(m = 1,..., n)$ is an arbitrary initial segmentation and c is a positive number, it will be shown that a reasonable choice for the latter is $1/M$.

Obviously at a balanced error solution

$$\mathbf{x}^{k+1} = \mathbf{x}^k$$

Let e_{ij} denote the ration $\Delta e_i/\Delta x_m$ (as defined in Section 2). Define the $(n-1) \times (n-1)$ matrix $\mathbf{H}$ as follows:

$$h_{mm} = 1 - c\,(e_{mm} - e_{m+1,m}) \qquad m = 1, 2,..., n-1 \tag{10a}$$

$$h_{m,m-1} = -ce_{m,m-1} \qquad m = 2,..., n-1 \tag{10b}$$

$$h_{m,m+1} = ce_{m+1,m+1} \qquad m = 1, 2,..., n-1 \tag{10c}$$

$$h_{mj} = 0 \qquad \text{otherwise.} \tag{10d}$$

The various elements of $\mathbf{H}$ are evaluated at a solution. It is well known that a functional iteration scheme will converge at a neighborhood of a solution if the matrix of the first differences has eigenvalues lying within the unit circle. It is easy to verify that H is that matrix in this case and, therefore, we must investigate its eigenvalues. Note that because x_n is fixed the n-segment problem actually has only $n-1$ degrees of freedom.

THEOREM 2. *For sufficiently small c, the matrix* $\mathbf{H}$ *defined by Eq.* (10) *has all its eigenvalues within the unit circle provided that all the error differences are nonzero. If some of them are zero, then* $\mathbf{H}$ *may have some eigenvalues equal to* 1.

Proof. Let $P_n(\lambda)$ denote the characteristic polynomial of $\mathbf{H}$. Then it can be readily verified that

$$P_n(\lambda) = [\lambda - 1 - c(e_{n,n-1} - e_{n-1,n-1})]\,P_{n-1}(\lambda) + c^2 e_{n-1,n-1}e_{n-1,n-2}P_{n-2}(\lambda). \tag{11}$$

By grouping together terms multiplied by c as $O(c)$, Eq. (11) can be written as

$$P_n(\lambda) = (\lambda - 1)\,P_{n-1}(\lambda) + O(c). \tag{12}$$

For sufficiently small c the roots of $P_n(\lambda)$ will be arbitrarily close to those of $P_{n-1}(\lambda)$ because the roots of a polynomial are continuous functions of its coefficients. Thus, if all the roots of $P_{n-1}(\lambda)$ are inside the unit circle, the same will be true for the roots of $P_n(\lambda)$, except possibly for the value $\lambda = 1$. This suggests a proof by induction plus a separate proof that $P_n(1)$ is nonzero.

Equation (4) plus the assumption of nonzero error differences imply that for some M

$$M \geqslant e_{mm} > 0 \qquad m = 1, 2, \ldots, n-1 \tag{13a}$$

$$M \geqslant -e_{m,m-1} > 0 \qquad m = 2, \ldots, n-1. \tag{13b}$$

For $n = 2$, we have

$$P_2(\lambda) = \lambda - 1 + c(e_{11} - e_{21}), \tag{14}$$

which for $c < 1/M$ has a root less than 1 in absolute value. Then because of Eq. (12) this will be true for all n, except possibly for roots $\lambda = 1$. We now examine this case. Equation (11) yields

$$P_n(1) + ce_{n,n-1}P_{n-1}(1) = ce_{n-1,n-1}[P_{n-1}(1) + ce_{n-1,n-2}P_{n-2}(1)] \tag{15}$$

Equation (13) implies that the sign of the left-hand side of Eq. (15) will be the same for all values for n. It can be easily shown that

$$P_3(1) = c^2[e_{32}(e_{21} - e_{11}) + e_{22}e_{11}].$$

Therefore

$$P_3(1) + ce_{32}P_2(1) = c^2[2e_{32}(e_{21} - e_{11}) + e_{22}e_{11}].$$

It can be seen from Eq. (13) that the right-hand side of the above equation is always positive. Hence

$$P_n(1) + ce_{n,n-1}P_{n-1}(1) > 0 \tag{16}$$

or

$$P_n(1) > -ce_{n,n-1}P_{n-1}(1). \tag{17}$$

Equations (13) and (14) imply that indeed

$$P_n(1) > 0 \text{ for all } n.$$

This completes the proof of the theorem.

COROLLARY. *The iterative procedure of Eq.* (9) *converges locally to a balanced error solution if such a solution exists, provided that c is sufficiently small.*

Note that the only case when eigenvalues equal one is that when error differences are zero, i.e., the change in the boundaries does not change the error. If this is the case in the neighborhood of a balanced error solution, it can only mean that the error norms in those intervals are already balanced.

Although we have proven only local convergence, practical computational experience has shown that the method converges globally (see next section). However, a formal proof for this case is still missing. We may add here that the usual proof for convergence of functional iteration schemes [14] is not applicable here because H may have elements greater than $1/(n-1)$.

Equation (11) can be used to show that the coefficient of λ^n is one while that of λ^{n-1} is

$$-(n-1) + c \sum_{i=2}^{n} [e_{1-1,i-1} - e_{i,i-1}]. \tag{12}$$

The negative of the above quantity is the sum of the roots and this will be less than $(n-1)$ in absolute value if

$$c < 1/M.$$

This suggests a choice for c. Note that after each iteration, one can calculate estimates of M easily by comparing the new with the previous error norms.

5. Time of Computation and Robustness

It is known that the convergence of functional methods like that discussed in the previous section is of first order [14] and therefore rather slow, in the sense that it requires many steps of the algorithm. However, the following observation is pertinent:

Let N be the total number of sample points of the function to be approximated, q the order of the approximating polynomial on each segment, and s the number of steps till convergence. Usually linear programming is used for curve fitting on each segment [10, 16] and thus let p be the average number of pivots per iteration. Then the total time of computation will be proportional to the quantity

$$T = sp[N + n(q+2)^3]. \tag{13}$$

If no information is available about the location of the error maxima, p is usually of the order of $q+2$. Thus, if the changes in the endpoints are large, the mth interval after the kth iteration may have very little overlap with the mth interval before the kth iteration and the locations of the maxima will differ substantially. On the other hand, if the changes in the endpoints are small (because of a small value of c) the two intervals will overlap and if the information on the location of the maxima is used to obtain a starting solution for the curve fitting, then the number of pivots will be very small. Tests with

piecewise linear approximations of various types of experimental data have shown that p is usually equal to one in such cases, where it averages around four for "cold starts." Thus, a first-order functional iteration method could be competitive with another method even if it required four times as many steps. Unfortunately, in most tests Lawson's method failed to converge and thus we have no statistics for the value of s [13, 15].

In many practical applications the data are given as discrete sample points and in general no balanced error solutions exist. It is then possible to use some direct discrete optimization techniques but they are quite slow, if one starts far from a solution [13]. Thus, a method searching for a balanced error solution can be used as a starter to bring the location of the joints to the neighborhood of the solution. The scheme of Eq. (9) will indeed drive the system there because it will attempt to make the pairwise error differences as small as possible. Lawson's algorithm will fail if one or more intervals at the neighborhood of a solution have zero error norms.

6. Numerical Results

A number of tests were made using the method of Eq. (9) (*Scheme* 1) or a variant of it (*Scheme* 2). The latter was defined as following. Let

$$\mu_m{}^k = \frac{1}{m} \sum_{i=1}^{m} e_i{}^k \qquad m = 1, 2, \ldots, n. \tag{14}$$

Then adjust the endpoints according to the following equation:

$$x_m^{k+1} = x_m{}^k + mc(\mu_n{}^k - \mu_m{}^k). \tag{15}$$

A stability analysis of this scheme gives the same results as for the first.

A trivial example involved an $f(x)$ which was piecewise linear to start with. Lawson's algorithm is not applicable but a test with scheme 2 converged in six iterations [15].

For smooth functions, like sin (x) and polynomials in x, Lawson's method requires fewer iterations than any of the above functional iteration schemes but the rate of convergence of functional iteration schemes (1 and 2) is still comparable. As we mentioned in the previous section, functional iteration is usually superior to Lawson's algorithm on irregular data. Digitized geographical map data belong to this category. In one case a terrain profile was approximated by five linear segments. For the same balance criterion (15% deviation of errors about $\mu_n{}^k$) the results were: Scheme 1–17 iterations; Scheme 2–10 iterations; Lawson's algorithm–13 iterations. For a greater number of segments (20) Lawson's algorithm always failed because in such

cases there were always intervals with zero error. The same experience was found in tests with electrocardiograms and scanning electron microscope data [13].

ACKNOWLEDGMENTS

We would like to thank Professor J. R. Rice for helpful suggestions which resulted in an improvement of the paper.

REFERENCES

1. J. R. RICE, "The Approximation of Functions," Vol. I (1964), and Vol. II (1969), Addison-Wesley.
2. L. L. SCHUMAKER, Uniform approximation by Chebyshev spline functions. II: Free knots, *SIAM J. Numer. Anal.* **5** (1968), 647–656.
3. C. E. DE BOOR AND J. R. RICE, "Least Square Cubic Spline Approximation: Variable Knots," *Tech. Report No. 81* (April 1968), Computer Science Department, Purdue University.
4. R. E. ESCH AND W. L. EASTMAN, Computational methods for best spline approximation," *J. Approximation Theory* **2** (1969), 85–96.
5. D. D. BRAESS, Chebyshev approximation by spline functions with free knots, *Numer. Math.* **17** (1971), 357–366.
6. R. BELLMAN, On the approximation of curves by line segments, *Comm. ACM* **4** (1961), 284.
7. B. GLUSS, Further remarks on line segment curve-fitting using dynamic programming, *Comm. ACM* **5** (1962), 441–443.
8. C. L. LAWSON, Characteristic properties of the segmented rational minimax approximation problem, *Numer. Math.* **6** (1964), 293–301.
9. G. BIRKHOFF AND C. R. DE BOOR, "Piecewise Polynomial Interpolation and Approximation," *in* "Approximation of Functions" (H. L. Garabedian, Ed), pp. 164–190. Elsevier, Amsterdam, 1965.
10. R. E. ESCH AND W. L. EASTMAN, "Computational Methods for Best Approximation," Tech. Report *SEG-TR*-67-30, Dec. 1967, Sperry Rand Research Center, Sudbury, Mass.
11. M. G. COX, Curve fitting with piecewise polynomials, *J. Inst. Math. Appl.* **8** (1971), 36–52.
12. T. PAVLIDIS, "Linguistic Analysis of Waveforms", "Software Engineering," Vol. 2, pp. 203–225. Academic Press, New York, 1971.
13. T. PAVLIDIS, Waveform segmentation through functional approximation, *IEEE Transactions on Computers* **C-22** (1973), 689–697.
14. E. ISAACSON AND H. B. KELLER, "Analysis of Numerical Methods," pp. 109–119. Wiley, New York.
15. A. P. MAIKA, "Minimax Piecewise Linear Approximation of Functions," *MSE Thesis*, Department of Electrical Engineering, Princeton University, August 1971.
16. P. RABINOWITZ, Applications of linear programming to numerical analysis, *SIAM Review* **10** (1968), 121–159.