1975

On the Computational Complexity of Approximation Operators II

John R. Rice
Purdue University, jrr@cs.purdue.edu

Report Number:
75-136

ON THE COMPUTATIONAL COMPLEXITY OF
APPROXIMATION OPERATORS II

John R. Rice
Purdue University

CSD-TR 136

March 14, 1975

ON THE COMPUTATIONAL COMPLEXITY OF APPROXIMATION OPERATORS II

John R. Rice*
Purdue University

INTRODUCTION. Computational complexity is a measure of the number of operations that some abstract machine requires to carry out a task. The task considered here is to compute an approximation to a real function $f(x)$ and the only operations that we count are evaluations of $f(x)$. Thus, we consider all other arithmetic performed to be negligible. We have already considered this topic in a previous paper [7], but we recast the terminology and notation to be more natural. We also sharpen many of the results of [7], and establish some new results.

We consider approximation by polynomials and piecewise polynomials in some norm (primarily L_2 and L_∞) . For a given number N of parameters (coefficients or knots) let $P_N^*(x)$ denote the best approximation and let $\varepsilon(N)$ denote its error $||f-P_N^*||$. Throughout we assume the approximation is on a standard interval. Note that P_N^* and $\varepsilon(N)$ depend on the norm, but the norm used is always clear from the context. It is generally impossible to compute $P_N^*(x)$ exactly, so we must consider estimates of $P_N^*(x)$. These estimates are produced by various computational algorithms and we have

Definition 1. An algorithm A which produces an estimate $P_L(x)$ of $P_N^*(x)$ so that, as N and $L(N) \to \infty$,

$$||f-P_L|| = \mathcal{O}(\varepsilon(N))$$

is called an optimal order L-parameter algorithm. The letter $M = M(A,N)$ denotes throughout the number of $f(x)$ evaluations required by A to compute $P_L(x)$. If $L = N$ and $M(A,N) = \mathcal{O}(N)$ then A is simply called an optimal algorithm.

The complexity of the algorithm is measured by M.

*This work was partially supported by NSF grant GP32940X

We denote the best approximation operator by T_N: $f(x) \to P_N^*(x)$ and we measure the complexity of T_N for a class C of functions by

$$M^*(N,C) = \inf_{A} \sup_{f\in C} M(A,N)$$

It is easy to believe (but not proved here) that M^* cannot be less than $\mathcal{O}(N)$ for any interesting class of functions.

Our ideal objective is to show that $M^*=N$ for various norms (e.g., L_1, L_2 and L_∞), approximation forms (e.g., polynomials, splines) and classes of functions (e.g., $C^p[-1,1]$, analytic in $|z| < 2$). Of course, we also wish to identify a corresponding optimal algorithm. We are able to do this in some cases and to come close in others. A significant conclusion derived from the results here is that asymptotically it is as easy to compute L_∞ approximations as L_2 approximations for most functions. A second significant conclusion is that, for a wide class of functions, piecewise polynomial approximations are no more complex to compute (even perhaps less complex) than ordinary polynomial approximations of comparable accuracy. We note that piecewise polynomials are much less complex to use than ordinary polynomials.

2. DISCRETIZATION.

The first algorithm we consider is

Algorithm 1 (Discretization) Set $X = \{ih \mid i=0, 1, 2,\ldots,1/h\}$, evaluate $f(x)$ on X and then compute $P_L(x)$ as the best approximation to $f(x)$ on X.

This algorithm is directly applicable to L_1, L_2 and L_∞ approximations. It was pointed out in [7] that a minor variant algorithm is not very useful for smooth functions and that one obtains $M^* = N^p$ for the class $C^p[0,1]$. Since then, Dunham [3] has shown that if the end points 0 and 1 are included in X (as they are) then a better results holds. We have

Theorem 1. Consider the class $C^p[0,1]$, $p \geq 2$, and polynomial approximation in the L_1, L_2 or L_∞ norms. Then discretization (Algorithm 1) is an optimal order N-parameter algorithm with

$$M = N^{p/2}$$

Proof. Let $P_N(x)$ be approximation produced by Algorithm 1. Dunham has shown [3] that $||P_N - P_N^*|| = \mathcal{O}(h^2)$. We have then that $\varepsilon(N) = N^{-p}$ and $M = 1/h$ and we may eliminate h from the relation $h^2 = N^{-p}$ to obtain the conclusions stated.

Corollary. *Algorithm 1 is an optimal algorithm for L_1, L_2 or L_∞ approximation by polynomials for the class $C^2[0,1]$.*

3. LEAST SQUARES APPROXIMATION BY POLYNOMIALS.

There appear to be two main algorithms for estimating least squares approximations by polynomials. For convenience we do least squares approximation with respect to the weight function $(1-x^2)^{-1/2}$ on $[-1,1]$. They are

Algorithm 2 (Gauss Quadrature for Fourier Coefficients). Estimate the coefficients

$$a_k^* = \int_{-1}^{1} \frac{f(x)\ T_k(x)\ dx}{\sqrt{1-x^2}}$$

by the Gauss quadrature formula:

$$a_k = \frac{2}{m} \sum_{i=1}^{m} f(\xi_i^{(m)})\ T_k(\xi_i^{(m)})$$

Thus a_k^* is the coefficient of the k-th Tchebycheff polynomial $T_k(x)$ in the Tchebycheff expansion of $f(x)$. The points $\xi_i^{(m)}$ are the m-point Gauss quadrature abscissa. The use of this algorithm and closely related ones is discussed in some detail by Rivlin [11, Section 3.5].

Algorithm 3 (Interpolation at the Tchebycheff points). Determine the polynomial $P_L(x)$ so that

$$P_L(\xi_i^{(L+1)}) = f(\xi_i^{(L+1)}) \qquad i = 1, 2, \ldots, L+1$$

It is well known that for P_L determined by Algorithm 3 we have

$$||f - P_L||_\infty \le ||f - P_L^*||_\infty\ (2/\pi \log L+1)$$

We note that if $m = L$, then the polynomials obtained by Algorithms 2 and 3 are the same [11].

Our first result on least squares is

<u>Theorem 2 Consider the class $C^p[-1, 1]$ $p \geq 3$ and least squares approximation by polynomials. Then Algorithm 2 is an optimal order N-parameter algorithm with</u>

$$M = N^{\frac{p}{p-1}}$$

<u>Proof</u>: We restrict our attention to $m \geq N$ and we have from [11, Theorem 3.12] that

$$||P_N^* - P_N|| < \sum_{j=1}^{\infty} \sum_{i=2jm-N}^{2jm+N} |a_i|$$

Now if $f(x) \in C^p[-1,1]$ we have that $|a_i| = \mathcal{O}(i^{-p})$ and we may estimate the inner sum, for some constant c, by

$$\sum_{i=2jm-N}^{2jm+N} |a_i| \leq \sum_{i=2jm-N}^{2jm+N} \frac{c}{i^p} \leq \sum_{i=(2j-1)m}^{(2j-1)m+2N} \frac{c}{i^p}$$

$$\leq \frac{c}{(p-1)[(2j-1)m]^{p-1}}$$

For $p \geq 3$ we then have that, for some constant c',

$$||P_N^* - P_N|| \leq \sum_{j=1}^{\infty} \frac{c}{(p-1)[(2j-1)m]^{p-1}} \leq \frac{c'}{m^{p-1}}$$

We now choose $m = M = N^{\frac{p}{p-1}}$ to obtain the correct order in the error $||f-P_N||$ and this concludes the proof.

We note that the previous result in [7] corresponds to obtaining $M = N^{\frac{p}{p-1.5}}$ and thus this sharpens that result. It seems likely that slightly more care in the proof would allow one to include the case $p = 2$, but then Algorithm 1 is already known to optimal for the class $C^2[-1,1]$.

Theorem 3. Consider the class $C^p[-1,1]$ and least squares approximation by polynomials. Then Algorithm 3 is an optimal order L-parameter algorithm with

$$L = M = N \sqrt[p]{\log N}$$

Proof. We have already noted that $||f-P_L||_\infty = \mathcal{O}(\varepsilon(L) \log L)$ and we also have that $||f-P_L||_2 \leq ||f-P_L||_\infty$. We have that $\varepsilon(L) = L^{-p}$ and $\varepsilon(N) = N^{-p}$. We claim that if $L = N \sqrt[p]{\log N}$ then $\varepsilon(L) \log L$ is $\mathcal{O}(N^{-p})$ because

$$(N \sqrt[p]{\log N})^{-p} \log (N \sqrt[p]{\log N}) = N^{-p}(\log N)^{-1} [\log N + 1/p \log N]$$

$$= N^{-p}(1+1/p)$$

This concludes the proof.

We see that Algorithm 3 uses fewer $f(x)$ evaluations than Algorithm 2, but it does not result in an Nth degree polynomial.

The non-optimality in Theorems 2 and 3 arises from functions in $C^p[-1,1]$ where the Tchebycheff expansion coefficients a_k^* are the order of k^{-p}. These functions are rather special since we must also have $\Sigma_{j=k}^{\infty} |a_k^*|$ the order of k^{-p}. Thus these functions have a very few large coefficients and the rest are comparitively negligible. The bulk of the funtions in $C^p[-1,1]$ would seem to be covered by the next Theorem.

Theorem 4. Consider the subclass of $C^p[-1,1]$ which has $|a_k| = \mathcal{O}(k^{-p-1})$ and least squares approximation by polynomials. Then Algorithms 2 and 3 produce $P_N(x)$ with $||f-P_N|| = \mathcal{O}(N^{-p})$ and $M = N$.

Proof. We must, of course, take $m = N$ in Algorithm 2 and $L = N$ in Algorithm 3. We have already noted that the two algorithms produce the same polynomial in this case, so we may restrict our attention to Algorithm 2. If we repeat the proof of Theorem 2 with $|a_i| = \mathcal{O}(i^{-p-1})$ instead of $\mathcal{O}(i^{-p})$ we see that the final estimate turns out to be

$$||P_N^* - P_N|| \le \frac{c'}{m^p}$$

and the choice $m = M = N$ produces the specified order in the error which concludes the proof.

Note that Theorem 4 does not state that Algorithms 2 and 3 are optimal for the subclass considered. They are not optimal because one sees that

$$||f-P_N^*||^2 = \sum_{j=N+1}^{\infty} a_i^2 \le c \sum_{j=N+1}^{\infty} \frac{1}{i^{2(p+1)}} \sim \frac{c}{N^{p+1/2}}$$

and hence $\varepsilon(N)$ is not N^{-p}.

There are classes of smoother functions where Algorithm 2 and 3 are optimal. We have

<u>Theorem 5</u>. <u>Consider $f(x)$ analytic in a region containing $[-1,1]$ and least squares polynomial approximation</u>. <u>Then Algorithms 2 and 3 are optimal</u>.

<u>Proof</u>. It is known that the hypothesis on $f(x)$ implies that $|a_i| < c\rho^i$ for some constants $\rho < 1$ and c. Of course, we have again taken $m = L = N$ in these algorithms. We use the same estimate as in the proof of Theorem 2 to obtain that

$$\sum_{i=(2j-1)N}^{(2j+1)N} |a_i| \le \frac{c}{1-\rho} \rho^{(2j-1)N}$$

and hence we have

$$||P_N^* - P_N|| \le \frac{c\rho}{(1-\rho)(1-\rho^2)} \rho^N$$

It is well known that $\varepsilon(N) = \mathcal{O}(\rho^N)$ and hence we have established that $||f - P_N|| = \mathcal{O}(\varepsilon(N))$ with $M = N$. This concludes the proof.

<u>4. TCHEBYCHEFF APPROXIMATION BY POLYNOMIALS</u>. We first note that Algorithm 3 (Interpolation at the Tchebycheff points) is equally applicable to Tchebycheff approximation and, in fact, we have

Theorem 6. Consider the class $C^p[-1,1]$ and Tchebycheff approximation by polynomials. Then Algorithm 3 is an optimal order L-parameter algorithm with

$$L = M = N \sqrt[p]{\log N}$$

The proof is essentially identical with that of Theorem 3.

It is well known [11] that the best least squares approximation is asymptotically as good as the best Tchebycheff approximation for analytic functions. Thus we immediately obtain from Theorem 5

Theorem 7. Consider f(x) analytic in a region containing [-1,1] and Tchebycheff approximation by polynomials. Then Algorithm 3 is optimal.

We now turn to the most common algorithm for computing Tchebycheff approximations:

Algorithm 4. (Remes Algorithm). Take a large number (say 2N) of points in [-1,1] and apply Algorithm 1 to obtain the best Tchebycheff approximation on this discrete set. Then apply the Remes algorithm [5], with this as initial guess and use the Murnaghan and Wrench procedure [4] to locate local maxima. Once convergence is achieved within the specified tolerance, check the error curve for extraneous maxima that invalidate the approximation obtained. The check is performed by sampling the error curve at a number of points proportional to N.

This statement of the Remes algorithm is one useful in practice. It is known [6], [12] that the Remes algorithm is Newton's method for a particular set of equations. As such it has two weaknesses: it might converge to a local solution that is not a global one and we do not know the number of iterations required before quadratic convergence sets in. In fact, the latter number is unbounded on the set $f(x) \in C^p[-1,1]$. Its strength is that it is quadratically convergent. In [7] we introduced some rather abstruse function classes in order to identify those f(x) where the Remes algorithm (Newton's method) behaves well. The fact of the matter is that one cannot identify such classes of functions with natural mathematical terms. The following definition allows us to make a more direct and intuitive presentation of the result.

Definition 3. Consider $f(x) \in C^3[-1,1]$ with $||f||_\infty \le 1/2$. Let $P_N^{(i)}(x)$ be the approximation obtained by the algorithm at the ith iteration and set $\delta_i = ||P_N^*(x) - P_N^{(i)}(x)||_\infty$. We say that the Remes algorithm converges normally with constant α for $f(x)$ if

(i) $\delta_i \le \alpha(\delta_0)^{2^i}$ (quadratic convergence)

(ii) the a posteriori check validates the approximation obtained (convergence to the global solution)

With this we may reformulate Theorem 4 of [7] as follows:

Theorem 8. Consider the class of functions in $C^p[-1,1]$ $p \ge 3$ where Algorithm 4 converges normally with constant $\alpha \le \alpha_0$. Then, for Tchebycheff approximation by polynomials, Algorithm 4 is an optimal order N-parameter algorithm with

$$M = \mathcal{O}(N \log\log N)$$

This more direct reformulation allows us to give a simpler proof than the one previously outlined.

Proof. The initial calculation of $P_N^{(0)}(x)$ requires $\mathcal{O}(N)$ evaluations of $f(x)$. Each iteration of the standard Remes algorithm requires 4N evaluations (3N are for the Murnaghan-Wrench estimation of local maxima of the error curve). The number of iterations required is determined by the condition that $\delta_i \le N^{-p}$. Since $\delta_0 \le 1/2$ we find that $i = \log\log N + c$ is a sufficient number where c is constant depending on α_0 and p. The validation check requies a further $\mathcal{O}(N)$ evaluation of $f(x)$ and the total number required is $\mathcal{O}(N \log\log N)$ as claimed.

Note that while Theorem 8 asymptotically specifies fewer evaluations than Theorem 6 (or Theorems 2 and 3 for least squares), this relation does not hold for problems likely to occur in practice. With the optimistic assumptions that the initialization and checking only require N evalutions each and that 4 iterations are required (independent of NI) we find that the Remes algorithm leads to 18N $f(x)$ evaluations. The values of N where Theorems 2, 3 and 6 start to require more evaluations are, for $p = 4$, $N = 324$ (Theorem 2) and $N = 10^{45590}$ (Theorems 4 and 6).

In a similar manner we may establish

Theorem 9. Consider the class of functions analytic in a region containing $[-1,1]$ where Algorithm 4 converges normally with constant $\alpha \leq \alpha_0$. Then, for Tchebycheff approximation by polynomials, Algorithm 4 is an optimal order N-parameter algorithm with

$$M = \mathcal{O}(N \log N)$$

Proof. The proof is the same as Theorem 8 except for bounding the number of iterations in the Remes algorithm. The requirement that $\delta_i \leq \rho^N$ (where $\rho < 1$ is associated with the size of the region of analyticity of $f(x)$) leads to $i = \log N + c$ (c = constant depending on ρ and α_0) as a sufficient number of iterations. The theorem now follows immediately.

5. PIECEWISE POLYNOMIAL APPROXIMATION. In our previous paper we proved that the spline projection operator of deBoor [1] is an optimal algorithm for $C^p[-1,1]$, for L_∞ approximation by piecewise pth degree polynomials with N knots. A recent adaptive approximation algorithm of Rice [8], [9] allows us to substantially enlarge the domain of functions where an optimal algorithm is known. We do not describe the algorithm here, but we do define a class of functions for which this algorithm is applicable.

Definition 4. The class $S_q^p[-1,1]$ of functions has the following properities:

a) Each $f(x)$ is bounded in the L_q norm on $[-1,1]$
b) Each $f(x)$ has a finite number of singularities

$$s_i,\ i = 1, 2, \ldots, R$$

We set $w(x) = \prod_{i=1}^{R} (x-s_i)$

c) $f^{(p)}(x)$ is continuous between the signularities
d) There are constants K and $\alpha > -1/q$ so that $|f^{(p)}(x)| \leq K|w(x)|^{\alpha-p}$ if $x \neq s_i$.

e) <u>For any interval $[x,x+\rho]$ we let $F_p(x,\rho)$ denote the L_q norm of $f^{(p)}(x)$ on this interval.</u> <u>Let $E(x,\rho)$ denote the error in the quadrature formula used by the adaptive approximation algorithm. This is typically a Gauss formula of precision p.</u> <u>There is a number $\lambda = \lambda(f)$ called the characteristic length so that if $\rho \leq \lambda$ we have</u>

(i) $E(x,\rho) \leq K\, F_p(x,\rho)\rho^{p+1}$ <u>if</u> $F_p(x,\rho) < \infty$

(ii) <u>otherwise $E(x,\rho) \leq K\rho^{1+\delta}$ for some $\delta > 0$.</u>

There are three pertinent remarks to be made about this definition. The first is that S_q^p contains essentially <u>all</u> functions of practical interest in approximation. The second is that S_q^p is a subset of the functions involved in the work of Burchard [2]. Finally, the somewhat lengthly part (e) of the definition is included to ensure that the algorithm is computationally effective. We note that <u>none</u> of the previous algorithms have this feature and computationally effective versions of them must have <u>at least</u> one additional fact about f(x), a fact analoguous to the characteristic length. The typical example of such a fact is the actually numerical value of the norm of $f^{(p)}(x)$. These facts provide a priori bounds on the oscillations of f(x) and its derivatives.

The work of Rice [10] and Burchard [2] shows that $\varepsilon(N) = N^{-p}$ for the class S_q^p and it has been shown by Rice [8], [9] that his adaptive algorithm achieves this degree of convergence. A simple inspection of that algorithm shows that the number of function evaluations is proportional to the number N of knots. The factor of proportionality is typically 8 or 10 although this would grow with larger p. These result imply

<u>Theorem 10</u>. <u>Consider the class $S_q^p[-1,1]$ and L_q approximation</u>, $1 \leq q \leq \infty$, <u>by piecewise polynomials of degree p.</u> <u>Then the adaptive approximation algorithm is optimal.</u>

REFERENCES

1. deBoor, Carl, On uniform approximation by splines, J. Approx. Thy., 1(1968) pp. 219-235.

2. Burchard, H. G., On the degree of convergence of piecewise polynomial approximations on optimal meshes II, to appear.

3. Dunham, C. B., Efficiency of Chebyshev approximation on finite subsets, J. Assoc. Comp. Mach., 21(1974) 311-313.

4. Murnaghan, F. D. and Wrench, J. W., The determination of the Chebyshev approximating polynomial of a differentiable function. Math. Comp., 13(1959) pp. 185-193.

5. Rice, J. R., The Approximation of Functions, Vol. I, Addison Wesley, Reading, Mass. (1964) Chapter 6.

6. ______________, The Approximation of Functions, Vol. II, Addison Wesley, Reading, Mass. (1969) Chapter 9.

7. ______________, On the computational complexity of approximation operators, in Approximation Theory (G. G. Lorentz, ed.) Academic Press, New York, (1973) pp. 449-456.

8. ______________, Adaptive approximation, J. Approx. Thy., to appear.

9. ______________, An algorithm for adaptive piecewise polynomial approximation, to appear.

10. ______________, On the degree of convergence for nonlinear spline approximations, in Approximations with Special Emphasis on Spline Functions (I.J. Schoenberg, ed.), Academic Press, New York, (1969) pp. 349-365.

11. Rivlin, T. J., The Chebyshev Polynomials, John Wiley, New York (1974).

12. Werner, H., Die contrucktive Ermittlung der Tschebyscheff-Approximierenden im Bereich der rationalen Funktionen, Arch. Rat. Mech. Anal., 11(1962) pp. 368-384.