Order Constrained Chebyshev Rational Approximation

by

J. Douglas Lawson
Department of Computer Science
University of Waterloo
Waterloo, Ontario, N2L 3G1

Terence C. Lau
Geac Canada Ltd.
350 Steelcase West
Unionville, Ontario

Research Report CS-80-10 (December, 1980)

Order Constrained Chebyshev Rational Approximation

J. Douglas Lawson

Department of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1

Terence C. Lau

Geac Canada Ltd
350 Steelcase West
Unionville, Ontario

ABSTRACT

In this paper, we study Chebyshev rational approximations which have specified derivatives at one end point of the interval of approximation. They then share the properties of best uniform and Padé approximations.

We prove a generalized de la Vallée-Poussin theorem, discuss the existence of such approximations, and provide a characterization of best approximations of this kind. This characterization uses the classical alternating set, but removes one point of alternation for each constraint.

Finally, we recommend an algorithm for computation of such approximants and illustrate this by an example.

December 8, 1980

Order Constrained Chebyshev Rational Approximation

J. Douglas Lawson

Department of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1

Terence C. Lau

Geac Canada Ltd
350 Steelcase West
Unionville, Ontario

1. Introduction

In this paper, we study rational approximations which share the properties of Padé and best uniform approximants. We require that the rational approximations be best in the Chebyshev sense, but we perform the optimization over subsets of the rational functions which have specified derivatives at one end point of the interval of approximation.

Let π_m denote the collection of all real polynomials of degree at most m and let $\pi_{m,n}$ denote the collection of all real rational functions $r_{m,n}(x)$ of the form:

$$r_{m,n}(x) = q_n^{-1}(x)p_m(x), \tag{1.1}$$

where $p_m \in \pi_m$ and $q_n \in \pi_n$. We normalize by prescribing $q(0) = 1$ and we assume that $q(x)$ does not vanish on the interval of approximation.

We assume that members of $\pi_{m,n}$ will be used to approximate a given function $f(x)$ on an interval $[0,b]$, that $f(x) \in C^k$ at $x = 0$, and that $f(x) \in C$ for $x \in [0,b]$, $0 < b < \infty$. Let

$$\frac{d^i}{dx^i} f(x)\Big|_{x=0} = i!c_i, \; i = 0, 1, \ldots, k. \tag{1.2}$$

Further, let $\pi_{k,m,n}$ be a subset of $\pi_{m,n}$ such that for $0 \leqslant k \leqslant m + n$,

$$r_{k,m,n}(x) \in \pi_{k,m,n} \iff \frac{d^i}{dx^i} r_{k,m,n}(x)\Big|_{x=0} = i!c_i, \; i = 0, 1, \ldots, k. \tag{1.3}$$

Now consider the error $\lambda_{k,m,n}$ associated with the best Chebyshev rational approximation of $f(x)$ by members of $\pi_{k,m,n}$ on $[0,b]$,

$$\lambda_{k,m,n} \equiv \inf_{r_{k,m,n} \in \pi_{k,m,n}} \left\{ \max_{0 \leqslant x \leqslant b} |r_{k,m,n}(x) - f(x)| \right\}. \tag{1.4}$$

In §2 we show that a generalized de la Vallée-Poussin theorem holds for such approximations. We prove in §3 the existence of an optimal approximation under certain conditions; and in §4 its characterization by an alternating set. An algorithm for the construction of such approximations is suggested and an example is presented in §5.

2. Generalized de la Vallée-Poussin Theorem

Theorem 1

Let $\mu \leqslant m$, $\nu \leqslant n$, $r(x) \in \pi_{k,m-\mu,n-\nu}$ *and* $q(x) \in \pi_{k,m,n}$ *and let* $H_q \equiv \max_{0 \leqslant x \leqslant b} |q(x) - f(x)|$. *Suppose that* $r(x) - f(x)$ *takes the values* $\lambda_1, -\lambda_2, \lambda_3, \ldots, (-1)^{N-1}\lambda_N$ *at the points* $0 < x_1 < x_2 < \cdots < x_N \leqslant b$, *with* $\lambda_i > 0$ *and* $N = m + n + 1 - k - min(\mu, \nu)$. *Then,*

$H_q \geqslant min(\lambda_1, \lambda_2, \ldots, \lambda_N)$ for all $q(x) \in \pi_{k,m,n}$.

Proof:

Let there exist $q^*(x) \in \pi_{k,m,n}$ such that $H_{q^*} < min(\lambda_1, \lambda_2, \ldots, \lambda_N)$. Form $\Delta(x) = r(x) - q^*(x) = [r(x) - f(x)] - [q^*(x) - f(x)]$. Clearly $\Delta(x_i) \neq 0$ and $\Delta(x)$ has alternating signs on $(x_1, x_2, \ldots, x_N)$. $\Delta(x)$ is continuous and has, therefore, at least $N-1$ zeros in (x_1, x_N). Further, $\Delta(x)$ has a zero of multiplicity $k + 1$ at $x = 0$ and thus possesses at least $m + n + 1 - min(\mu, \nu)$ zeros on $[0, b]$. However, $\Delta(x)$ is a rational function of which the numerator has degree at most $m + n - min(\mu, \nu)$, contradicting the existence of such a $q^*(x)$. □

3. Existence

Theorem 2

If $\pi_{k,m,n}$ *is non-empty, then there exists* $\bar{r}(x) \in \pi_{k,m,n}$ *for which* $\max_{0 \leqslant x \leqslant b} |\bar{r}(x) - f(x)| = \lambda_{k,m,n}$.

Proof:

Consider the set $\left\{ \max_{0 \leqslant x \leqslant b} |r(x) - f(x)|, \; r(x) \in \pi_{k,m,n} \right\}$.

Since $\pi_{k,m,n}$ is non-empty, the set is defined and bounded. Let ρ be its greatest lower bound. If we re-normalize the members of $\pi_{k,m,n}$ in a way similar to that in the proof for the classical unconstrained case [see 1, §33] and employ the same technique, it can be shown that there exists a convergent sequence $\bar{v}_i = (a_{i0}, \ldots, a_{im}, b_{i0}, \ldots, b_{in})$, $i = 1, 2, \ldots$, such that the rational functions $r_i(x) = \sum_{j=0}^{m} a_{ij} x^j / \sum_{j=0}^{n} b_{ij} x^j$, $i = 1, 2, \ldots$, are in $\pi_{k,m,n}$, $\lim_{i \to \infty} \bar{v}_i = (a'_0, a'_1, \ldots, a'_m, b'_0, \ldots, b'_n)$ and the function $\bar{r}(x) = \sum_{j=0}^{m} a'_j / \sum_{j=0}^{n} b'_j x^j$ is *bounded* in $[0,b]$. Hence, after being reduced to its lowest terms, $\bar{r}(x)$ will assume a form $\bar{p}(x)/\bar{q}(x)$ where $\bar{p}(x) = \sum_{j=0}^{m-\mu} a_j x^j$, $\bar{q}(x) = \sum_{j=0}^{n-\nu} b_j x^j$, $a_{m-\mu} \neq 0$, $b_0 \neq 0$ and $\bar{q}(x)$ does not vanish in $[0,b]$. This in turn implies $\frac{d^i}{dx^i} (\bar{r}(x))\big|_{x=0}$ exists for $i = 0, 1, \ldots, k$. Furthermore, at this particular point $x = 0$, these derivatives will be continuous functions of the coefficients of $\bar{r}(x)$. Therefore, $r_i(x)$, $i = 1, 2, \ldots$ converges uniformly to $\bar{r}(x)$, and further, $\frac{d^j}{dx^j} (r_i(x))\big|_{x=0} = \frac{d^j}{dx^j} (\bar{r}(x))\big|_{x=0} = j! c_j$, $j = 0, 1, \ldots, k$ and $\bar{r}(x)$ is in $\pi_{k,m,n}$.

In fact, as in [1, §33], $\bar{r}(x)$ could be shown to attain the greatest lower bound ρ, and is hence the best approximation. □

We shall call $\bar{r}(x)$ the best approximation to $f(x)$ in $\pi_{k,m,n}$. It may easily be shown that in certain situations, $\pi_{k,m,n}$ may be empty. For instance, let $f(x)$ be the exponential function $exp(x)$, $m = 0$, $n = 1$ and $k = 1$. Obviously, $\pi_{k,m,n}$ has only one member, namely, the $(0,1)$-Padé approximation $1/(1-x)$, over the interval $[0,b]$, $b < 1$. For $b \geqslant 1$, $\pi_{k,m,n}$ will be empty. In fact, for any $r(x) = \sum_{i=0}^{m} a_i x^i / \sum_{i=0}^{j} b_i x^i$, $b_0 = 1$, $r(x)$ will be in $\pi_{k,m,n}$ provided the following two conditions are satisfied:

Condition (1): $a_j - \sum_{i=0}^{j} b_i c_{j-i} = 0$, $j = 0, 1, \ldots, k$,

$a_j = 0$, $j > m$,

$b_j = 0$, $j > n$, and

Condition (2): $\sum_{i=0}^{n} b_i x^i$ does not vanish in $[0,b]$.

The [illegible] is a subset of the usual order-constraints used in Padé approximations.

Theorem 3

If $k \leqslant \max\{m, n-1\}$, *then* $\pi_{k,m,n}$ *is non-empty.*

Proof:

If $k \leqslant m$, it is possible first to choose b_i, $i = 1, \ldots, n$ satisfying condition (2). These, together with a_i, $i = 1, \ldots, k$ obtained next from condition (1) and a_i, $i = k+1, \ldots, m$ chosen arbitrarily, will yield an $r(x)$ in $\pi_{k,m,n}$.

If $m < k \leqslant n-1$, we can first solve the linear system in condition (1) for a_i, $i = 1, \ldots, m$ and b_i, $i = 1, \ldots, k$. Next, by choosing appropriate values for b_i, $i = k+1, \ldots, n$, say sufficiently large value with a correct sign for b_n, we shall get an $r(x)$ whose poles are all away from the non-negative real axis. □

Theorem 4

For $k > \max\{m, n-1\}$, *if the linear system* (*in* b_i) $\sum_{i=1}^{n} b_i c_{j-i} = -c_j$, $j = m+1, m+2, \ldots, m+n$ *is non-singular and yields* b_i, $i = 1, \ldots, n$ *satisfying condition* (2), *then* $\pi_{k,m,n}$ *is non-empty.*

Proof:

Obvious. □

Hence, for any reasonably smooth $f(x)$, low order-constrained best approximation can always be found. The existence of high order-constrained approximations, on the other hand, will depend on the function to be approximated and its higher derivatives at the origin.

The case when $f(x)$ is the exponential function $exp(-x)$ over the non-negative axis is of particular interest recently because of the usefulness of rational exponential approximation in the numerical solutions of systems of differential equations, especially to heat-conduction type problems or to problems which are classified "stiff". (See for example [3,4,5,6,7,9]). In [7], Lawson points out some applications of order-constrained Chebyshev rational approximation to $exp(-x)$. Ehle [4] establishes that each Padé approximant entry $R_{m,n}(z)$ for $exp(-z)$ on the first two subdiagonals of the Padé table has all its poles in the open left half-plane. Saff, et.al. [8] extends this result to the first four subdiagonals, and to entries sufficiently far out on *any* subdiagonals. Hence, we have

Remark 1

There exists an optimal approximation in $\pi_{k,m,n}$ *to* $exp(-z)$ *for* $m \leqslant n-4$, *and* $k \leqslant m+n$, $n = 0, 1, 2, \ldots$ *over the non-negative real axis.*

Remark 2

For any integer τ, *there exists an integer* n *such that, in each of* $\{\pi_{k,m,m+\tau},\ k \leqslant 2m+\tau\}_{m=n}^{\infty}$, *there exists an optimal approximation to* $exp(-z)$ *over the non-negative axis.*

4. Characterization

Theorem 5

The rational function $\overline{r}_{k,m-\mu,n-\nu}(x)$ is optimal in $\pi_{k,m,n}$ in the Chebyshev sense if and only if there exists a set of points $0 < x_1 < x_2 < \cdots < x_N \leqslant b$, $N = m + n + 1 - k - \min(\mu, \nu)$ and a constant λ for which

$$\overline{r}_{k,m-\mu,n-\nu}(x_i) - f(x_i) = (-1)^i \lambda, \quad i = 1, 2, \ldots, N.$$

Proof:

That the existence of an alternating set of N or more points is a sufficient condition for $\overline{r}(x)$ to be optimal is an immediate consequence of *Theorem 1*.

To establish the necessity, we assume that $r_{k,m-\mu,n-\nu}(x)$ is optimal in $\pi_{k,m,n}$ and is in reduced form, but that it possesses an alternating set of $N' < N$ points. That is,

$$\pm[r_{k,m-\mu,n-\nu}(x_i) - f(x_i)] = (-1)^i \lambda_r, \quad i = 1, 2, \ldots, N', \lambda_r > 0.$$

We divide $[0,b]$ into N' partial intervals $[0, \xi_1], [\xi_1, \xi_2], \ldots, [\xi_{N'-1}, b]$, such that in each sub-interval the following inequalities hold alternately:

$$-\lambda_r \leqslant r(x) - f(x) < \lambda_r - \alpha;$$

$$-\lambda_r + \alpha < r(x) - f(x) \leqslant \lambda_r.$$

Define $\Phi(x) = \prod_{i=1}^{N'-1}(x - \xi_i)$. Since the numerator $p(x)$ and denominator $q(x)$ of $r(x)$ have no common factors, we may find $\phi(x)$ and $\psi(x)$ of degrees at most n and m respectively such that:

$$x^{k+1}\Phi(x) = [q(x)\psi(x) - p(x)\phi(x)].$$

Now, let

$$r^*(x) = \frac{p^*}{q^*} = \frac{p(x) + \omega\psi(x)}{q(x) + \omega\phi(x)}.$$

Further, (omitting argument x),

$$r^* - f = r - f + \left(\frac{p^*}{q^*} - \frac{p}{q}\right).$$

$$\frac{p^*}{q^*} - \frac{p}{q} = \frac{p^*q - pq^*}{qq^*}$$

$$= \frac{q(p + \omega\psi) - p(q + \omega\phi)}{qq^*}$$

$$= \frac{\omega(q\psi - p\phi)}{qq^*}$$

$$= \frac{\omega\,\Phi\,x^{k+1}}{qq^*}.$$

Choose ω small enough that $qq^* \geqslant \beta > 0$, and $|\ \omega\,\Phi\,x^{k+1}\ | \leqslant \frac{\alpha\beta}{2}$ for $x \in [0,b]$. Since Φ alternates on the subintervals, an appropriate choice of sign for ω ensures that r^* is a better approximation than r.

From the construction it is clear that $r^* - r = O(x^{k+1})$. Thus, $r \in \pi_{k,m,n}$ implies that $r^* \in \pi_{k,m,n}$, and the hypothesis that r is optimal in $\pi_{k,m,n}$ is contradicted.

5. An algorithm and examples

In the construction of an algorithm for generating constrained Chebyshev approximations, we are guided by two considerations:

i) The approximations are characterized by an alternating set of $m + n + 1 - k$ points, assuming no degeneracy;

ii) The constraints may be represented by a set of $k + 1$ *linear* equations linking the coefficients of the numerator and denominator of the approximating rational function.

One of the best current algorithms for the construction of Chebyshev rational approximations is given as an ALGOL procedure by Cody, Fraser and Hart [2]. This algorithm proceeds from a given approximation to the alternating set of critical points using the equations:

$$p_m(x_i) - q_n(x_i)f(x_i) - (-1)^i \lambda q_n(x_i) = 0,\ i = 1, 2, \ldots, m + n + 2. \tag{5.1}$$

These equations are linear in the $m + n + 1$ coefficients $(a_0, a_1, \ldots, a_m)$ and $(b_1, b_2, \ldots, b_n)$ where $p_m(x) = \sum_{i=0}^{m} a_i x^i$ and $q_n(x) = \sum_{i=1}^{n} b_i x^i + 1$. The algorithm is:

i) Solve (5.1) for $\{a_i\}$, $\{b_i\}$ and λ, using an interation on λ;

ii) Search for the critical points of the rational approximation thus produced;

iii) Repeat with these new critical points until convergence or choose a new starting approximation if the process diverges.

It is a relatively straightforward procedure to replace equations (5.1) in the algorithm by:

$$a_j - \sum_{i=0}^{j} b_i c_{j-i} = 0,\ j = 0, 1, \ldots, k;$$

$$p_m(x_i) - q_n(x_i)f(x_i) - (-1)^i \lambda q_n(x_i) = 0,\ i = 1, 2, \ldots, m + n + 1 - k. \tag{5.2}$$

The resulting scheme has proven to be effective on a number of test problems. In some cases, however, there was either no convergence or cycling in the iterations even for very good initial approximations to the critical points.

As an example of the use of this scheme, we consider the approximation of $f(x) = exp[-x/(1-x)]$, $x \in [0, 1)$, $f(1) = 0$. This problem arose from an attempt to generalize the uniform approximations of e^{-x} on $[0, \infty)$, considered in [3], to approximations having specified order at $x = 0$.

A single-precision ALGOL code was prepared by imbedding equations (5.2) in the procedure given in [2]. For $k = m = n$, $k = 2, 3, 4, 5$, approximations were computed, with convergence for all cases from the initial approximations to the critical points given by:

$$x_1 = 0.7,\ x_{i+1} = (1+x_i)/2,\ i = 1, 2, \ldots, n.$$

The resulting approximation for $k = 2$, for example, was

$$r(x) = \frac{1.0 - 2.2284800x + 1.2342096x^2}{1.0 - 1.2284800x + 0.5057296x^2}$$

The critical points were 0.6612573, 0.9039874 and 0.9985108 with associated errors 0.01965, -0.01967 and 0.01936.

Note that the first set of equations of (5.2) prescribing the order are exactly satisfied, so that nothing would be gained by double precision computations; a more precise levelling of the error curve would be of no practical importance.

In table I, we show the values of the error norm λ computed for these approximations. For comparison we include the values of the error norm λ^* associated with the unconstrained approximations of the same degree, given in [3].

n	λ (constrained)	λ (unconstrained)
2	1.96 (-02)	7.36 (-03)
3	2.86 (-03)	7.99 (-04)
4	4.11 (-04)	8.65 (-05)
5	5.9 (-05)	9.5 (-06)

6. Acknowledgements

One of the authors (JDL) gratefully acknowledges the hospitality of the University of Dundee in the preparation of part of this work and the assistance of Mr. K. Brodlie in providing a program for the algorithm of [2]. We also thank Professor Charles Dunham for his helpful comments on an earlier version of the paper and Chuan K. Chee for his assistance with typesetting of the paper. A portion of the work was supported by NRC Grant A1244.

7. References

[1] N.I. Achieser, Theory of Approximation, trans. by Charles J. Hyman, Frederick Ungar, New York, 1956.

[2] W.J. Cody, W. Fraser and J.F. Hart, *Rational Chebyshev approximation using linear equations*, Numer. Math., 12 (1968), pp. 242-251.

[3] W.J. Cody, G. Meinardus and R.S. Varga, *Chebyshev rational approximation to e^{-x} in $[0, \infty)$ and application to heat conduction problems*, J. Approx. Theor., 2 (1969), pp. 50-65.

[4] Ehle, B.L., *A-stable methods and Padé approximation to the exponential.* SIAM J. Math. Anal., 4 (1973), pp. 671-680.

[5] Ehle, B.L., Lawson, J.D., *Generalized Runge-Kutta processes for stiff initial-value problems*, J. Inst. Maths. Applics., 16 (1975), pp. 11-21.

[6] Lau, T.C.Y., *A class of approximations to the exponential function for the numerical solution of stiff differential equations.* Research Report #CS-74-13, University of Waterloo, Canada, August, 1974.

[7] Lawson, J.D., *Some numerical methods for stiff ordinary and partial differential equations*, Proc. Second Manitoba Conference on Numerical Math., 1972, pp. 27-34.

[8] Saff, E.B., Varga, R.S., *On the zeros and poles of Padé approximants to $\exp(x)$*, Numer. Math., 25 (1975), pp. 1-14.

[9] Saff, E.B., Schönhage, A., Varga, R.S., *Geometric convergence to $\exp(-x)$ by rational* [illegible] 7-32.