LINEAR PROGRAMMING METHOD FOR RATIONAL APPROXIMATION

BY

JONATHAN HAR-EL AND SHMUEL KANIEL

ABSTRACT

Development of a new algorithm, based on linear programming, for the computation of the best rational approximation of a continuous function.

1. Introduction

Let $f(x)$ be a continuous function defined on some finite interval $[a, b]$. Let $R_{l,m}$ denote the class of rational functions of the form

$$R(x) = \frac{P(x)}{Q(x)} = \left(\sum_{j=0}^{l} p_j x^j\right) \Big/ \left(\sum_{j=0}^{m} q_j x^j\right) \tag{1}$$

where $Q(x) > 0$ in $[a, b]$.

It is well known that there exists a unique function, $R^* \in R_{l,m}$ of best approximation in the maximum norm [1].

Various algorithms for computing R^* are suggested. The Remes algorithms (see [4], [6], [9]) are fast, but their convergence is assured only if a sufficiently good initial approximation is used. The differential correction algorithm (given in [2], [3]) always converges, but the volume of computation is prohibitive. Other algorithms (given in [2], [5], [8]) are neither convergent nor fast. Here we develop a new algorithm, based on linear programming. It is relatively fast and convergence is assured in all cases.

2. The linear programming algorithm

Given

$$R \in R_{l,m},\ R(x) = \frac{P(x)}{Q(x)} = \left(\sum_{j=0}^{l} p_j x^j\right) \Big/ \left(\sum_{j=0}^{m} q_j x^j\right),\ Q(x) > 0 \text{ in } [a, b],$$

denote $p = (p_0, \cdots, p_l)$, $q = (q_0, \cdots, q_m)$, and denote $S(x, p, q) = R(x) - f(x)$. Denote $S^*(x) = R^*(x) - f(x)$ where R^* is the best approximation to f in $R_{l,m}$. Let $R^{(1)} \in R_{l,m}$ be an initial approximation to f.

Let $x_i(R)$, $i = 1, \cdots, N(R)$ be the positive local maxima and the negative local minima of $S(x) = R(x) - f(x)$. Denote:

$$\text{(2)} \qquad \alpha_{ij}(R) = \frac{\partial S}{\partial p_j}(x_i(R)) = \frac{\partial R}{\partial p_j}(x_i) = \frac{x_i^j}{Q(x_i)} \text{ and}$$

$$\text{(3)} \qquad \beta_{ij}(R) = \frac{\partial S}{\partial q_j}(x_i(R)) = \frac{\partial R}{\partial q_j}(x_k) = \frac{-x_i^j \cdot P(x_i)}{[Q(x_i)]^2}.$$

Denote by $R^{(K)}$ the Kth approximation to f. Denote $S^{(K)}(K) = R^{(K)}(x) - f(x)$. Denote $\alpha_{ij}^{(K)}$ and $\beta_{ij}^{(K)}$ the quantities defined in (2) and (3) where $R^{(K)}$ is substituted instead of R.

For the variables ξ_j, $j = 0, \cdots, l$, η_j, $j = 0, \cdots, m$ and ε solve the following linear programming problem.

LINEAR PROGRAMMING PROBLEM (LP1). Maximize ε under the following constraints:

$$\text{(4)} \qquad 0 \leqq \varepsilon \leqq \| S^{(K)} \|,$$

$$\text{(5)} \qquad -1 \leqq \xi_j \leqq 1 \text{ for } j = 0, \cdots, l,$$

$$\text{(6)} \qquad -1 \leqq \eta_j \leqq 1 \text{ for } j = 0, \cdots m,$$

$$\text{(7)} \qquad \operatorname{sign} S^{(K)}(x_i^{(K)}) [S^{(K)}(x_i^{(K)}) + \sum_{j=0}^{l} \alpha_{ij}^{(K)} \xi_j + \sum_{j=0}^{m} \beta_{ij}^{(K)} \eta_j] + \varepsilon \leqq \| S^{(K)} \| \text{ for } i = 1, \cdots, N(K).$$

Then set

$$\text{(8)} \qquad P^{(K+1)}(x) = P^{(K)}(x) + \lambda_K \sum_{j=0}^{l} \xi_j^{(K)} x^j \text{ and}$$

$$\text{(9)} \qquad Q^{(K+1)}(x) = Q^{(K)}(x) + \lambda_K \sum_{j=0}^{m} \eta_j^{(K)} x^j$$

where $\xi_j^{(K)}$, $\eta_j^{(K)}$ is a solution (not necessarily unique) of LP1, and λ_K minimizes

$$\text{(10)} \qquad d(\lambda) = \| S(\cdot, p^{(K)} + \lambda \xi^{(K)}, q^{(K)} + \lambda \eta^{(K)}) \|.$$

Clearly $d(\lambda_K) \leqq d(0) = \| S^{(K)} \|$. It will be shown that $d(\lambda_K)$ tends to $d^* = \| R^* - f \|$.

3. Convergence

We prove convergence with the following mild assumption. For any $R \in R_{l,m}$ the number of positive maxima as well as negative minima of $S = R - f$ in $[a, b]$ is finite. In this case the linear programming problem LP1 is well defined. Denote $d_K = \| S^{(K)} \|$. Then by (10) it follows that $d_{K+1} \leqq d_K$. We will prove that d_K tends to $d^* = \| S^* \|$.

PROPOSITION 1. *There exists a sequence K_n for which $R^{(K_n)}$ is convergent.*

PROOF.

$$\| R^{(K)} \| = \| R^{(K)} - f + f \| \leqq \| R^{(K)} - f \| + \| f \| = d_K + \| f \|.$$

Since $d_K \leqq d_1$ it follows that $R^{(K)}$ is bounded. Hence by a standard argument, it is possible to choose a convergent subsequence. Denote the limit of $R^{(K_n)}$ by $\bar{R}$. Assume that $\bar{R} \neq R^*$. Any standard proof of the characterization of R^* (see [1]) is based on the following two propositions.

PROPOSITION 2. *Denote by M the set of points x_i in $[a, b]$ for which $|S(x_i, p, q)| = \| S(\cdot, p, q) \|$. Let $A(x) = \sum_{j=0}^{l} a_j x^j$ and $B(x) = \sum_{j=0}^{m} b_j x^j$ be such that*

$$\operatorname{sign}[Q(x_i)A(x_i) - P(x_i)B(x_i)] \underset{x_i \in M}{\neq} \operatorname{sign} S(x_i, p, q). \tag{11}$$

Then for all λ sufficiently small

$$\| S(\cdot, p + \lambda a, q + \lambda b) \| < \| S(\cdot, p, q) \|. \tag{12}$$

PROPOSITION 3. *If $R \neq R^*$ then there exist $A(x)$, $B(x)$ satisfying* (11) *and afortiori* (12).

The choice of the linear programming problem LP1 was motivated by the following reasoning. We can look at the notations (2) and (3), the linear programming problem LP1, and the problem (8)–(9) for a general function $R = P/Q \in R_{l,m}$ and deviation $S(x) = R(x) - f(x)$. Now we establish the following proposition.

PROPOSITION 4. *Let $A(x) = \sum_{j=0}^{l} a_j x^j$, $B(x) = \sum_{j=0}^{m} b_j x^j$. Then inequalities* (4)–(7) *are satisfied for some $\varepsilon > 0$ (not necessarily the best) and $\xi = \omega a$, $\eta = \omega b$ where $\omega > 0$ if and only if* (11) *is satisfied for $A(x)$ and $B(x)$.*

PROOF. Suppose that there exist $A(x)$ and $B(x)$ satisfying (11).

$$Q(x_i)A(x_i) - P(x_i)B(x_i)$$

$$= Q^2(x_i)\left[\frac{A(x_i)}{Q(x_i)} - \frac{P(x_i)}{Q^2(x_i)}B(x_i)\right] = Q^2(x_i)\left[\sum_{j=0}^{l} a_j \frac{x_i^j}{Q(x_i)} - \sum_{j=0}^{m} b_j \frac{x_i^j P(x_i)}{Q^2(x_i)}\right].$$

Now recall that

$$\frac{x_i^j}{Q(x_i)} = \alpha_{ij}, \quad \frac{-x_i^j P(x_i)}{Q^2(x_i)} = \beta_{ij}.$$

Hence

$$Q(x_i)A(x_i) - P(x_i)B(x_i) = Q^2(x_i)\left[\sum_{j=0}^{l} \alpha_{ij}a_j + \sum_{j=0}^{m} \beta_{ij}b_j\right]. \tag{13}$$

Define now $\xi = \omega a$, $\eta = \omega b$ where $\omega > 0$. Since $Q^2(x_i) > 0$ it follows from (11) and (13) that

$$\operatorname{sign}(x_i) \neq \operatorname{sign}\left[\sum_{j=0}^{l} \alpha_{ij}\xi_j + \sum_{j=0}^{m} \beta_{ij}\eta_j\right], \qquad x_i \in M. \tag{14}$$

Let ω be chosen to be so small that (5) and (6) are satisfied. Restrict it, furthermore, so that

$$0 < \left|\sum_{j=0}^{l} \alpha_{ij}\xi_j + \sum_{j=0}^{m} \beta_{ij}\eta_j\right| \leqq |S(x_i)| = \|S\|, \qquad x_i \in M. \tag{15}$$

It follows that (14) and (15) imply the existence of $\varepsilon_1 > 0$ for which (7) is true for all $x_i \in M$. As for $x_i \notin M$ $|S(x_i)| < \|S\|$. Denote now $\delta = \min_{x_i \notin M} \|S\| - |S(x_i)|$. Then restrict ω again, if necessary, so that

$$\left|\sum_{j=0}^{l} \alpha_{ij}\xi_j + \sum_{j=0}^{m} \beta_{ij}\eta_j\right| \leqq \tfrac{1}{2}\delta, \qquad x_i \notin M. \tag{16}$$

Thus (4)–(7) will be satisfied for $\varepsilon = \min\{\varepsilon_1, \frac{1}{2}\delta\}$. Conversely, if (4)–(7) are satisfied for some $\varepsilon > 0$ then

$$\operatorname{sign}\left[\sum_{j=0}^{l} \alpha_{ij}\xi_j + \sum_{j=0}^{m} \beta_{ij}\eta_j\right] \neq \operatorname{sign} S(x_i, p, q), \qquad x_i \in M. \tag{17}$$

Thus if $a_j = \xi_j$, $b_j = \eta_j$ then for the corresponding polynomials $A(x)$, $B(x)$ relation (11) holds via (13).

Proposition 5. *Suppose that $\bar{R} \neq R^*$; then there exist $\bar{\delta} > 0$, $\bar{\varepsilon} > 0$ and $\bar{\mu} > 0$ so that for all $R \in R_{l,m}$ such that $\|R - \bar{R}\| < \bar{\delta}$ there exists a solution $(\varepsilon, \xi, \eta) = (\varepsilon, \xi, \eta)(R)$ of (4)–(7) with $\varepsilon \geqq \frac{1}{2}\bar{\varepsilon}$ so that*

$$\min_{\lambda} \|S(\cdot, p + \lambda\xi, q + \lambda\xi)\| \leqq \|S(\cdot, p\ q)\| - \bar{\mu}. \tag{18}$$

Proof. We use a standard compactness argument. If $\bar{R} \neq R^*$, then by Proposition 5 it is possible to solve the inequalities (4)–(7) for $\bar{R}$, $\bar{\varepsilon} > 0$, $\bar{\xi}, \bar{\eta}$. Since the extreme points of $S(x, p, q)$ are continuous functions of p and q it follows, for δ sufficiently small that

$$\operatorname{sign} S(y_i)\left[S(y_i) + \sum_{j=0}^{l} \alpha_{ij}\bar{\xi}_j + \sum_{j=0}^{m} \beta_{ij}\bar{\eta}_j\right] + \tfrac{1}{2}\bar{\varepsilon} \leqq \| S \| \tag{19}$$

where $\| R - \bar{R} \| < \delta$ and y_i are the extreme points of $S(x)$. If we maximize as we do in the algorithm, we certainly can derive $\varepsilon(R) \geqq \frac{1}{2}\bar{\varepsilon}$.

Suppose now that (18) is false. This means that there exist $R^{(K)}$, $\xi^{(K)}$, $\eta^{(K)}$ (not coinciding with those generated in the algorithm) such that $\| R^{(K)} - \bar{R} \| < \delta$ and for which (4)–(7) hold with $\varepsilon_K \geqq \frac{1}{2}\bar{\varepsilon}$ while

$$\min_{\lambda} \| S(\cdot, p^{(K)} + \lambda\xi^{(K)}, q^{(K)} + \lambda\eta^{(K)}) \| \geqq \| S(\cdot, p^{(K)}, q^{(K)} \| - \mu_K \tag{20}$$

where $\mu_K \to 0$ as $K \to \infty$.

Extract now a subsequence K_n for which $p^{(K_n)}$, $\xi^{(K_n)}$, $q^{(K_n)}$ and $\eta^{(K_n)}$ converge to p^0, ξ^0, q^0 and η^0 respectively. For R^0 we can solve the inequalities (4)–(7) with $\varepsilon^0 \geqq \frac{1}{2}\bar{\varepsilon}$. Therefore, by Propositions 2 and 3, for the polynomials $A^0(x) = \sum_{j=0}^{l} \xi_j^0 x^j$ and $B^0(x) = \sum_{j=0}^{m} \eta_j^0 x^j$ there exist $\lambda^0 > 0$ and $\mu^0 > 0$ so that

$$\| S(\cdot, p^0 + \lambda^0\xi^0, q^0 + \lambda^0\eta^0) \| \leqq \| S(\cdot, p^0, q^0) \| - \mu^0. \tag{21}$$

By continuity for $K = K_n$ sufficiently large we obtain

$$\| S(\cdot\ p^{(K)} + \lambda^0\xi^{(K)}, q^{(K)} + \lambda^0\eta^{(K)}) \| \leqq \| S(\cdot, p^{(K)}, q^{(K)}) \| - \mu^0 \tag{22}$$

contrary to (20).

Now the proof of convergence is clear. If the sequence $R^{(K_n)}$ of Proposition 1 converges to $\bar{R} \neq R^*$ then for $K = K_n$ sufficiently large

$$\min \| S(\cdot, p^{(K)} + \lambda\xi^{(K)}, q^{(K)} + \lambda\eta^{(K)}) \| \leqq \| S(\cdot, p^{(K)}, q^{(K)}) \| - \bar{\mu}. \tag{23}$$

Hence for $K = K_n$, by (8) and (9) and Proposition 5,

$$\| S(\cdot, p^{(K+1)}, q^{(K+1)}) \| \leqq \| S(\cdot, p^{(K)}, q^{(K)}) \| - \bar{\mu}. \tag{24}$$

We assumed the sequence $S^{(K_n)}$ to be convergent, thus (24) can hold for only a finite number of indices, a contradiction.

4. Application

Theoretically, it is sufficient to consider in (7) only the global extrema, that is, the points $x_i^{(K)}$ for which $| S^{(K)}(x_i^{(K)}) | = \| S^{(K)} \|$. The inclusion of all positive local maxima and negative local minima has been done just to achieve better practical efficiency.

Enlarging the number of points $x_i^{(K)}$ in (7) can only increase the improvement, $\| S^{(K)} \| - \| S^{(K+1)} \|$, and thereby accelerate convergence. However, adding too many points leads to a big system of constraints in the linear programming problem.

Another possible strategy is to consider all local extrema if their number does not exceed $l + m + 2$, and if there are more than $l + m + 2$ local extrema, to take only those local extrema x for which $|S^{(K)}(x)| > c\| S^{(K)} \|$ where $0 < c < 1$ is some constant. A fair choice could be $c = \frac{1}{4}$.

EXAMPLE. The following is an example of rational approximations computed by the proposed algorithm. Chose $f(x) = \sin x$, $a = -1$, $b = 1$, $l = 4$, $m = 4$.

(i) For $R^{(1)}(x) = \dfrac{.99749x - .15652x^3}{1}$,

the local extrema of $S^{(1)}(x)$ are as follows.

Point	Value
−1.0000E + 00	4.9953E − 04
−8.0767E − 01	−4.9953E − 04
−3.0768E − 01	4.9953E − 04
3.0768E − 01	−4.9953E − 04
8.0767E − 01	4.9953E − 04
1.0000E + 00	−4.9953E − 04

(ii) For $R^{(3)}(x) = \dfrac{x - .10713x^3}{1 + .05956x^2 + .0015244x^4}$,

the local extrema of $S^{(3)}(x)$ are as follows.

Point	Value
−1.0000E + 00	1.5990E − 06
−8.7078E − 01	−1.5255E − 06
−5.1551E − 01	8.6111E − 07
−9.7562E − 02	−4.2207E − 08
9.7564E − 02	4.2267E − 08
5.1551E − 01	−8.6105E − 07
8.7079E − 01	1.5256E − 06
1.0000E + 00	−1.5989E − 06

(iii) For $R^{(8)}(x) = \dfrac{x - .10476x^3}{1 + .061905x^2 + .0019952x^4}$, the local extrema of $S^{(8)}(x)$ are as follows.

Point	Value
$-1.0000E+00$	$2.4363E-08$
$-9.3847E-01$	$-2.4363E-08$
$-7.6255E-01$	$2.4363E-08$
$-4.9590E-01$	$-2.4363E-08$
$-1.7182E-01$	$2.4363E-08$
$1.7180E-01$	$-2.4363E-08$
$4.9588E-01$	$2.4363E-08$
$7.6254E-01$	$-2.4363E-08$
$9.3847E-01$	$2.4363E-08$
$1.0000E+00$	$-2.4363E-08$

References

1. N. I. Achieser, *Theory of Approximation*, Frederick Ungar Publishing Company, New York, 1956.
2. E. W. Cheney, *Introduction to Approximation Theory*, McGraw-Hill, New York, 1966.
3. E. W. Cheney, and H. L. Loeb, *On rational Chebyshev approximation*, Numer. Math. **4** (1962), 124–127.
4. W. J. Cody, W. Fraser, and J. F. Hart, *Rational Chebyshev approximation using linear equations*, Numer. Math. **12** (1968), 242–251.
5. H. J. Maehly, *Methods for fitting rational approximations II, III*, J. Assoc. Comp. Mach. **10** (1963), 257–277.
6. A. Ralston, *Rational Chebyshev approximation by Remes' algorithms*, Numer. Math. **7** (1965), 322–330.
7. A. Ralston, and H. S. Wilf, eds.: *Mathematical Methods for Digital Computers*, Vol. II, John Wiley & Sons, Inc.
8. J. R. Rice, *The Approximation of Functions*, Vol. II, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1964.
9. H. Werner, J. Stoer, and W. Bommas, *Rational Chebyshev approximation*, Numer. Math. **10** (1967), 289–306.

THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM, ISRAEL