REVUE D'ANALYSE NUMÉRIQUE ET DE THÉORIE DE L'APPROXIMATION
Tome 23, N° 1, 1994, pp. 55–62

A MODELLING BY RATIONAL APPROMAXIONS

PAVOL CHOCHOLATÝ
(Bratislava)

1. INTRODUCTION

A scientist who has compiled tables of data wishes to reduce them to a more convenient and comprehensible form. He accomplishes this by representing the data in functional form. He selects a class of functions and chooses from this class the one that best fits his data.

In a computational structure sense, polynomials are the simplest "finitely representable" functions that can be used in approximating continuous functions. It seems reasonable to consider the next simplest class to be the rational functions, and they are usually a somewhat more efficient form of approximation. A rational function $R(t) = P(t)/Q(t)$ is one which can be evaluated as the quotient of two polynomials. If $Q(t) \neq 0$, then a single division yields the final result.

Actually, a rather different algorithm may be less time consuming depending on the time required for the multiplication and division operations in particular computer being used. The alternative algorithm is derived by transforming $R(t)$ into a continued fraction [4], [5]. These last years some theoretical aspects of continued fractions have been discussed in many papers, f.e. [2], [3]. If the division and multiplication operations take about the same time in the computer, then there is a decided advantage in using a continued-fraction expansion.

The purpose of this paper is to demonstrate a method for evaluting the coefficients of functions $R(t)$:

$$R(t) \equiv \sum_{i_1=1}^{k_1} \frac{A_{i_1}}{t + m_{i_1}} \tag{1.1}$$

and

$$R(t) = R(t^{q-1}, K_1, K_2, \ldots, K_{q-2}) \equiv \cfrac{A}{t + \cfrac{K_1}{t + \ldots \cfrac{}{t + \cfrac{K_{q-2}}{t + m_{k_{q-1}}}}}} \tag{1.2}$$

where $(q-1)$ denotes the stage of the polynomial $Q(t)$ and K_j, $j = 1, 2, \ldots, q-2$, are prescribed values or functions $K_j = K_j(m_{k_{q-1}})$,

and

$$R(t) = R(t, L_{p_1}, L_{p_2}, \ldots, L_{p_{q-2}}) \equiv$$

$$(1.3)\qquad \sum_{i_1=1}^{k_1} \cfrac{A_{1,i_1}}{t + \sum_{i_2=1}^{k_2} \cfrac{A_{2,i_2}}{L_{i_1}^{(1)} + \ldots \cfrac{}{L_{i_{q-3}}^{(q-3)} + \sum_{i_{q-1}=1}^{k_{q-1}} \cfrac{A_{q-1,i_{q-1}}}{L_{i_{q-2}}^{(q-2)} + m_{i_{q-1}}}}}}$$

where $L_{p_j} = (L_1^{(j)}, L_2^{(j)}, \ldots, L_{k_j}^{(j)})$, $j = 1, 2, \ldots, q-2$, are prescribed values, $L_{r+1}^{(j)} = L_r^{(j)} + h^{(j)}$, $r = 1, 2, \ldots, k_j - 1$, $h^{(j)} > 0$.

We now formulate the core of our method. Suppose that k_0 values z_{1,i_0} as a monotonous sequence of real numbers at some k_0 real equidistant points t_{1,i_0} are given, where $t_{1,a+1} = t_{1,a} + h^{(0)}$, $a = 1, 2, \ldots, k_0 - 1$, $h^{(0)} > 0$.

We propose to develop a method for determining $R(t)$ as an approximating function in terms of the values z_{1,i_0}.

Define

$$(1.4)\qquad z_{s,i_{s-1}} = T_{s,i_{s-1}}(t_{s,i_{s-1}}, x_s, T_{s+1})$$

for $s = 1, 2, \ldots, q-1$ and unknowns $x_s = (x_{s,i_1}, \ldots, x_{s,i_{k_s}})$, where $T_{q,i_{q-1}}$ are elements of the vector $T_q = (m_1, \ldots, m_{k_{q-1}})$.

Now, suppose that

$$(1.5)\qquad T_{s,i_{s-1}} = \sum_{i_s=1}^{k_s} \frac{x_{s,i_s}}{t_{s,i_{s-1}} + T_{s+1,i_s}}$$

is justified for $s = 1, 2, \ldots, q-1$, where $2k_s < k_{s-1} - 1$, $x_{s,i_s} = A_{s,i_s}$ and for $b = 2, 3, \ldots, q-1$ we denote

$$t_{b,i_b} = L_{i_{b-1}}^{(b-1)}.$$

Under the assumption that $k_1 = 1$ holds $k_j = 1$ too, for $j = 2, \ldots, q-1$, $x_{j,1} = K_{j-1}$ and $t_{r,i_{r-1}}$, $r = 2, 3, \ldots, q-1$, take the values of t_{1,i_0}, $i_0 = 1, 2, \ldots, k_0$, $x_{1,1} = A$.

Our method determines values of vectors x_s and T_q by solving of a linear overdetermined system and a polynomial equation of the k_{q-1} -th order, which may be in some cases realized parallel. Details of computational implementation and results are provided.

2. THE MAIN NUMERICAL CONSTRUCTION

Thus the suitable approximation $R(t)$ has some form of (1.1), (1.2), (1.3), we get a system of k_0 equations of the form (1.4), (1.5). For one, we determine the vector T_q under the assumption that $q = 2$ for simplicity.

The first differences from values z_{1,i_0} we can express in the form

$$(2.1)\qquad z_{1,r+1} - z_{1,r} = -h^{(0)} \sum_{i_1=1}^{k_1} \frac{x_{1,i_1}}{(t_{1,r} + T_{2,i_1})(t_{1,r+1} + T_{2,i_1})},$$

$$r = 1, 2, \ldots, k_0 - 1,$$

or, in short,

$$(2.2)\qquad z_{1,r+1} - z_{1,r} = \sum_{i_1=1}^{k_1} d_{r,i_1}$$

where the relation

$$(2.3)\qquad d_{r+1,i_1} = \frac{t_{1,r} + T_{2,i_1}}{t_{1,r+2} + T_{2,i_1}} \cdot d_{r,i_1} \equiv u_{r,i_1} \cdot d_{r,i_1}$$

holds for $r = 1, 2, \ldots, k_0 - 2$, and $i_1 = 1, 2, \ldots, k_1$.

Evidently for fixed $r (r = 1, 2, \ldots, k_0 - k_1 - 1)$, (2.1) to (2.3) give

$$(2.4)\qquad z_{1,r+p+1} - z_{1,r+p} = \sum_{i_1=1}^{k_1} \left(\prod_{e=0}^{p-1} u_{r+e,i_1} \right) d_{r,i_1}, \quad p = 1, 2, \ldots, k_1,$$

(if $p = 0$, then (2.4) becomes (2.2)).

Note that (2.3) implies the following result:

$$(2.5)\qquad u_{r+e,i_1} = C_e(u_{r,i_1}), \quad e = 0, 1, \ldots, k_1 - 1,$$

where

$$(2.6)\qquad C_e(y) = \frac{e(1-y) + 2y}{(e+2)(1-y) + 2y}.$$

Hence, instead of (2.4) we have

$$(2.7)\qquad z_{1,r+p+1} - z_{1,r+p} = \sum_{i_1=1}^{k_1} \left(\prod_{e=0}^{p-1} C_e(u_{r,i_1}) \right) d_{r,i_1}, \quad p = 1, 2, \ldots, k_1.$$

Let $n_{k_1-p,r}$, $p = 0, 1, \ldots, k_1$ and $r = 1, 2, \ldots, k_0 - k_1 - 1$ be parameters where $n_{0,r} = 1$. By adding a multiple of the r-th equation of (2.2) to multiple of the first equation of (2.4) we can produce a new equation containing $z_{1,r+1} - z_{1,r}$, $z_{1,r+2} - z_{1,r+1}$ and by adding this to a multiple of the second equation of (2.4) we produce a new equation containing $z_{1,r+1} - z_{1,r}$, $z_{1,r+2} - z_{1,r+1}$, $z_{1,r+3} - z_{1,r+2}$. The obvious continuation of this process utimately results in a new set of equations not containing d_{r,i_1}:

$$(2.8)\qquad \sum_{p=0}^{k_1} n_{k_1-p,r}(z_{1,r+p+1} - z_{1,r+p}) = 0, \quad r = 1, 2, \ldots, k_0 - k_1 - 1,$$

where the parameters $n_{v,r}$ are the mentioned 'multipliers'.

Using relations (2.2) and (2.7) in (2.8) for each i_1, $i_1 = 1, 2, \ldots, k_1$, we obtain immediately the equations

$$n_{k_1,r} + \sum_{p=1}^{k_1} n_{k_1-p,r}\left(\prod_{c=0}^{p-1} C_c(u_{r,\cdot})\right) = 0. \tag{2.9}$$

$$r = 1, 2, \ldots, k_0 - k_1 - 1.$$

Now, using (2.6), equations (2.9) become

$$\sum_{p=0}^{k_1} J_p\, u_{r,\cdot}^{k_1-p} = 0, \tag{2.10}$$

for $r = 1, 2, \ldots, k_0 - k_1 - 1$, where J_p depends linearly on $n_{0,r}, n_{1,r}, \ldots$ $\ldots, n_{k_1 r}$, and $u_{r,i_1} \in (0, 1)$, $i_1 = 1, 2, \ldots, k_1$. We know from algebra that the coefficients J_p of each polynomial equation in (2.10) can be expressed in terms of the roots.

In practice, it is possible to find expressions in the form

$$n_{b,r} = Q_b(u_{r,1}, u_{r,2}, \ldots, u_{r,k_1}), \tag{2.11}$$

$r = 1, 2, \ldots, k_0 - k_1 - 1$; $b = 1, 2, \ldots, k_1$. Using equality (2.5) for u_{r,i_1}, $i_1 = 1, 2, \ldots, k_1$, and applying the property of roots of the polynomials (2.10), (2.11) becomes

$$n_{b,r} = X_b(u_{1,1}, u_{1,2}, \ldots, u_{1,k_1}), \tag{2.12}$$

$r = 1, 2, \ldots, k_0 - k_1 - 1$; $b = 1, 2, \ldots, k_1$, where u_{1,i_1} are roots of the first polynomial equation in (2.10). The roots of this polynomial equation can be expressed in terms of the coefficients J_p, i.e. $n_{0,1}, n_{1,1}, \ldots, n_{k_1,1}$. Turning to our equations (2.12) we have

$$n_{b,r} = Y_b(n_{0,1}, n_{1,1}, \ldots, n_{k_1,1}), \tag{2.13}$$

$$r = 1, 2, \ldots, k_0 - k_1 - 1;\quad b = 1, 2, \ldots, k_1.$$

Substituting (2.13) in (2.8) we obtain a linear overdetermined system

$$\sum_{p=0}^{k_1} n_{k_1-p,1}\, Z_{r,p}(z_{1,r}, z_{1,r+1}, \ldots, z_{1,r+k_1+1}) = 0, \tag{2.14}$$

$r = 1, 2, \ldots, k_0 - k_1 - 1$, containing k_1 unknown parameters $n_{1,1}$, $n_{2,1}, \ldots, n_{k_1,1}$. Such overdetermined system can be solved in the L_1 sense. Background material on L_1 minimization can be found in Barrodale and Roberts [1]. We propose to solve system (2.14) in the least squares sense

$$E = \sum_{r=1}^{k_0-k_1-1}\left(\sum_{p=0}^{k_1} n_{k_1-p,1}\, Z_{r,p}(z_{1,r}, z_{1,r+1}, \ldots, z_{1,r+k_1+0})\right)^2. \tag{2.15}$$

We try to determine parameters $n_{b,1}$, $b = 1, 2, \ldots, k_1$, from the system

$$\frac{\partial E}{\partial n_{b,1}} = 0. \tag{2.16}$$

It is now clear that the accuracy under which we can compute roots $u_{1,1}, u_{1,2}, \ldots, u_{1,k_1}$ of the first polynomial equation in (2.10) depends essentially on the accuracy under which we can evalute the coefficients J_p of the polynomial.

We see T_{2,i_1}, $i_1 = 1, 2, \ldots, k_1$ as a result of relation (2.3) (for $r = 1$) is well defined. In our special case was $q = 2$ that means m_{i_1}, $i_1 = 1, 2, \ldots, k_1$ for (1.1) are calculated.

Returning now to relations (1.2) and (2.3), i.e. if we take $q = 3$, it follows that the coefficient m_{k_2} can be computed by solving the nonlinear equation

$$u_{1,1} = \frac{(K_1 - (t_{1,1} + h^{(0)} + m_{k_2})(t_{1,1} + 2h^{(0)} + m_{k_2}))}{(K_1 - (t_{1,1} + m_{k_2})(t_{1,1} + h^{(0)} + m_{k_2}))} \cdot \tag{2.17}$$

$$\cdot \frac{(t_{1,1}(t_{1,1} + m_{k_2}) + K_1)}{((t_{1,1} + 2h^{(0)})(t_{1,1} + 2h^{(0)} + m_{k_2}) + K_1)}.$$

In order to calculate the components x_{1,i_1}, $i_1 = 1, 2, \ldots, k_1$, of vector x_1, we use the values of vector T_2. Then equations (1.4), (1.5) form the linear overdetermined system of k_0 equations with k_1 unknowns x_{1,i_1}, $i_1 = 1, 2, \ldots, k_1$. This overdetermined system is solved in the least squares sense under the assumption that k_1 is the rank of its matrix. Using our preceding results we obtain the approximating function $R(t)$ written in the form of (1.1) and (1.2).

Since we shall confine our attention to the approximating function $R(t)$ written in the form of (1.3). Therefore we must consider anew the same approach to evaluate T_2 as before. This is, at the same time as (1.4), possible to write in the form

$$z_{2,i_1} = T_{2,i_1}(t_{2,i_1}, x_2, T_3),\quad i_1 = 1, 2, \ldots, k_1.$$

It is quite easy to see that the computation of T_3 can be realized parallel to the computation of x_1. It seems plausible, therefore, to repeat the procedure (1.4) with T_4, x_2 replacing T_3, x_1 and continue to repeat the process until the desired values T_q, x_{q-1} are achieved. This completes the evaluation of coefficients for the function $R(t)$ given in the form of (1.3).

3. NUMERICAL EXPERIENCE

One may obtain the data by observing situations occurring in the nature, or one may set up experiments in which conditions are controlled as to favour the process of observation. Suppose we have a monotonous sequence of k_0 measurements z_{1,i_0}, $i_0 = 1, 2, \ldots, k_0$, recorded when the clock indicated times t_{1,i_0}, and we are interested on finding the coefficients of the function $R(t)$ given by relation (1.1).

Using the results of (2.16) for $k_1 = 2,3$ we can obtain from (2.10) the next formulas, i.e. u_{r,i_1}, $i_1 = 1, 2, \ldots, k_1$ are roots of the system of

$(k_0 - k_1 - 1)$ equations :

a) $k_1 = 2$

$$(1 - n_{1,r})\,u_{r,.}^2 + (1 + 3n_{1,r} - n_{2,r})\,u_{r,.} + 3n_{2,r} = 0 \tag{3.1}$$

b) $k_1 = 3$

$$(n_{2,r} - n_{1,r})u_{r,.}^3 + (1 + n_{3,r} - 5n_{2,r} + n_{1,r})\,u_{r,.}^2 + \\ + (1 - 5n_{3,r} + 6n_{2,r} + 2n_{1,r})u_{r,.} + 6n_{3,r} = 0 \tag{3.2}$$

where $r = 1, 2, \ldots, k_0 - k_1 - 1$.

Owing to the reason of the simplicity, our technique will be presented for $k_0 = 15$ and $k_1 = 2$. Applying the property of roots of the polynomials in (3.1), (2.13) becomes

$$n_{1,r+1} = \frac{2n_{1,r} + n_{2,r} - 3}{2n_{1,r} + n_{2,r} + 6}, \qquad n_{2,r+1} = \frac{n_{2,r} - n_{1,r}}{2n_{1,r} + n_{2,r} + 6}, \tag{3.3}$$

where $r = 1, 2, \ldots, 11$, or exactly

$$n_{1,r} = \frac{Bn_{1,1} + Cn_{2,1} + D}{En_{1,1} + Fn_{2,1} + G}, \qquad n_{2,r} = \frac{Hn_{1,1} + In_{2,1} + J}{En_{1,1} + Fn_{2,1} + G}, \tag{3.4}$$

where the coefficients $B, \ldots, J$ are given in Table 1.

Table 1

r	B	C	D	E	F	G	H	I	J
1	1	0	0	0	0	1	0	1	0
2	2	1	−3	2	1	6	−1	1	0
3	−1	0	−8	5	3	10	−1	0	1
4	−2	−1	−5	3	2	5	0	0	1
5	−13	−8	−24	14	10	21	2	1	6
6	−22	−15	−35	20	15	28	5	3	10
7	−11	−8	−16	9	7	12	3	2	5
8	−46	−31	−63	35	28	45	14	10	21
9	−61	−48	−80	44	36	55	20	15	28
10	−26	−21	−33	18	15	22	9	7	12
11	−97	−80	−120	65	55	78	35	28	45
12	−118	−99	−143	77	66	91	44	36	55

Now, we complete our expression by working forwards in (2.8) (substitution of (3.4)) and obtain a linear overdetermined system

$$n_{2,1}\,[-Iz_{1,r} + (I - C)z_{1,r+1} + (C - F)z_{1,r+2} + Fz_{1,r+3}] + \\ + n_{1,1}[-Hz_{1,r} + (H - B)z_{1,r+1} + (B - E)z_{1,r+2} + Ez_{1,r+3}] + \\ + [-Jz_{1,r} + (J - D)z_{1,r+1} + (D - G)z_{1,r+2} + Gz_{1,r+3}] = 0, \tag{3.5}$$

$r = 1, 2, \ldots, 12.$

The unknown parameters $n_{1,1}$, $n_{2,1}$ are calculated by minimizing (3.5) in the least-squares sense. Hence we can compute roots $u_{1,1}$, $u_{1,2}$ of the first ($r = 1$) polynomial equation in (3.1) by the subroutine POLRT from SSP (Scientific Subroutine Package).

We complete the calculation of $R(t)$ by solving the linear overdetermined system (2.1) for unknown parameters $x_{1,1}$, $x_{1,2}$.

We give two illustrative problems to which our techniques may be applied. The first applies our method to the case with positive m_{i_1}, $i_1 = 1, 2, \ldots, k_1$. The second one handles negative m_{i_1}.

Example 1. Suppose that 15 values z_{1,i_0} as the values of the function $1/(t_{1,i_0} + 1) + 1/(t_{1,i_0} + 5)$ at 15 equidistant points t_{1,i_0} are given. As the first case we calculate the approximation $R_2(t)$ for $k_1 = 2$, $t_{1,1} = 1$, $h^{(0)} = 5$. For the second case, if we take $k_1 = 2$, $t_{1,1} = 0.5$, $h^{(0)} = 1.4$ a convenient approximation $\bar{R}_2(t)$ is obtained. Some results are presented in Table 2.

Table 2

t	$R_2(t)$, $\bar{R}_2(t)$	$\lvert\text{exact} - R_2(t)\rvert$	$\lvert\text{exact} - \bar{R}_2(t)\rvert$
1	0.666667	0.160475−13	0.455191−14
2	0.476191	0.269507−13	0.832667−15
3	0.375000	0.229261−13	0.165840−14
10	0.157576	0.104361−15	0.102660−16
20	0.087619	0.101491−17	0.103356−18
30	0.060829	0.706125−19	0.725122−20

Example 2. Here we consider the values of the function $0.5/(t_{1,i_0} - 1) + 0.1/(t_{1,i_0} - 2)$, at the 15 equidistant points t_{1,i_0}, for $k_1 = 2$, $t_{1,1} = 5$, $h^{(0)} = 0.3$. Some results for $R_2(t)$ are presented in Table 3.

Table 3

t	$R_2(t)$	exact $- R_2(t)$
3	0.350000	0.111369−14
4	0.216667	0.156480−14
5	0.158333	0.595010−15
10	0.068056	0.246114−16
20	0.031871	0.132696−17
30	0.020813	0.203859−19

These results are typical of our wider experience with this method applied to a variety of test problems. The quality of the approximation $R(t)$ depends on the determination of values k_j, $j = 1, \ldots, q - 1$, and t_{1,i_0}, $i_0 = 1, 2, \ldots, k_0$. Prior to the question of which of forms (1.1), (1.2), (1.3) to use in computing a best approximation is the question of whether to prefer rational to polynomial approximation, or possibly other forms

such as piecewise-polynomial or piecewise-rational functions or spline functions. The answer depends to some extend on the intended use of the approximation $R(t)$ and in our case on prescribed coefficients $K_1, K_2, \ldots$ $\ldots, K_{q-2}$ or $L_1^{(j)}, L_2^{(j)}, \ldots, L_{k_j}^{(j)}$ too.

REFERENCES

1. Barrodale, I., Roberts, F.D.K., *An improved algorithm for discrete L_1 linear approximation*, SIAM J. Numer. Anal., **10** (1973), 839—848.
2. Hovstad, R. M., *Continued fraction tails and irrationality*, The Rocky Mountain J. Math. **19**, *4* (1989), 1035—1041.
3. Jacobsen, L., Waadeland, H., *An asymptotic property for tails of limit periodic continued fractions*, The Rocky Mountain J. Math. **20**, *1* (1990), 151—163.
4. Jones, W. B., Thorn, W. J., *Continued fractions, analytic theory and applications*, Encyclopedia of Mathematics and its Application. (Addison-Wesley, Reading, Massachusetts, 1980).
5. Waadeland, H., *Local properties of continued fractions*, Lecture Notes in Math., 1237 (1987) 239—250.

Received 1 VII 1992

Department of Numerical Analysis
and Optimization
Comenius University Bratislava
Mlynská dolina
CS—842 15 *Bratislava*
Slovakia