SIAM J. NUMER. ANAL.
Vol. 18, No. 5, October 1981

0036-1429/81/1805-0009 $01.00/0

ON THE EVALUATION OF RATIONAL FUNCTIONS IN INTERVAL ARITHMETIC*

G. ALEFELD† AND J. G. ROKNE‡

Abstract. We discuss some new expressions for rational functions for which the Hausdorff distance between the range over a given interval and the interval arithmetic evaluation has the order of the square of the width of the given interval.

1. Preliminaries and introduction. Let there be given a real-valued rational function of the real variable x,

$$f(x)=\frac{g(x)}{h(x)}=\frac{\sum_{\gamma=0}^{r} a_\gamma x^\gamma}{\sum_{\gamma=0}^{s} b_\gamma x^\gamma}, \tag{1}$$

and a compact interval $X=[x_1, x_2]$. If $g(X)$ and $h(X)$ denote the interval arithmetic evaluation of one of the infinitely many equivalent representations of g and h (for example, by using Horner's method) and if $0 \notin h(X)$, then

$$f(x) \in f(X) := \frac{g(X)}{h(X)}, \qquad x \in X. \tag{2}$$

Let

$$q(A, B)=\max\{|a_1-b_1|, |a_2-b_2|\}$$

denote the Hausdorff metric for real intervals $A=[a_1, a_2]$, $B=[b_1, b_2]$ and let $d(X)=x_2-x_1$ denote the width of the interval $X=[x_1, x_2]$. Then if

$$W(f, X) := [f(u), f(v)]$$

is the range of f over X it holds that

$$q(W(f, X), f(X)) \leqq \alpha d(X), \tag{3}$$

provided we have

$$d(g(X)) \leqq c_1 d(X), \qquad d(h(X)) \leqq c_2 d(X).$$

These last two relations are valid if, for example, we use Horner's scheme in evaluating $g(X)$ and $h(X)$. In the sense of (3), $W(f, X)$ is therefore linearly approximated by $f(X)$.

In his book [5, § 6.2], R. E. Moore has demonstrated by some examples that using the so-called centered form of f allows one to replace $d(X)$ by the square of $d(X)$ on the right-hand side of (3). Furthermore, he made the conjecture that this is true in general. The centered form is described by Moore in the following manner:

Let there be given a rational function $f(x)$ and a real number $c=m(X)$ (=the midpoint of the interval X) at which f is defined. Then we have

$$f(x)=f(c)+t(x-c), \tag{4}$$

where the real function t is defined by

$$t(y)=f(y+c)-f(c). \tag{5}$$

* Received by the editors July 14, 1980.

† Fachbereich 3 - Mathematik, Technische Universität Berlin, 1 Berlin 12, Germany.
‡ Department of Computer Science, University of Calgary, Calgary, Alberta, Canada T2N 1N4.

The representation (4) of f is called the centered form. For t the only assumption is that the number of occurrences of the variable y in the expression chosen for $t(y)$ cannot be further reduced by cancellations. For example,

$$t(y) = y^2 - y^2 + 2y + 1$$

is not allowed. In this case we have to use

$$t(y) = 2y + 1.$$

Moore's conjecture now reads (in a different but because of [1, eq. (21), p. 24] equivalent form):

If

$$f(X) := f(c) + t(X - c),$$

then

$$q(W(f, X), f(X)) \leqq \alpha\, d(X)^2 \tag{6}$$

holds.

Although it is perhaps generally known, we first show by a simple example that without additional assumptions on the expression used for $t(y)$, (6) does not hold.

Let

$$f(x) = \frac{1+x}{2+x},$$

$c = m(X) = 0$, $X = [-r, r]$, $r < 2$, and hence $d(X) = 2r$. We have $f(c) = f(o) = \frac{1}{2}$, and therefore (4) reads

$$f(x) = \tfrac{1}{2} + t(x - c), \tag{7}$$

where, from (5),

$$t(y) = \frac{1+y}{2+y} - \frac{1}{2}.$$

Using this expression for $t(y)$ we get from (7)

$$f(X) = \frac{1}{2} + \frac{1+[-r, r]}{2+[-r, r]} - \frac{1}{2} = \begin{cases} \left[\dfrac{1-r}{2+r}, \dfrac{1+r}{2-r}\right], & r < 1, \\ \left[\dfrac{1-r}{2-r}, \dfrac{1+r}{2-r}\right], & 1 \leqq r < 2. \end{cases}$$

Since in this case

$$W(f, X) = \left[\frac{1-r}{2-r}, \frac{1+r}{2+r}\right],$$

a simple computation shows that we only have

$$q(W(f, X), f(X)) = \frac{2r}{4-r^2}(1+r) = O(d(X)).$$

If, however, we use the expression

$$t(y) = \frac{y}{2(2+y)} = y \cdot w(y), \qquad w(y) = \frac{1}{2(2+y)}$$

for t and evaluate (7) by first evaluating $w(X-c)$ then multiplying this interval by $X-c$ and adding $\frac{1}{2}$, then we actually have the estimation (6).

In general, the following is true. If

$$f(x)=f(c)+(x-c)\cdot w(x-c), \tag{8}$$

then for

$$f(X)=f(c)+(X-c)\cdot w(X-c)$$

Moore's conjecture is true. This has been proven first by Hansen [3, p. 102 ff.]; see also [1, p. 37 ff.]. Hansen's proof shows that the conjecture is also true if the factor $x-c$ is "multiplied into the dividend" of $w(x-c)$. If, for example,

$$f(x)=f(c)+(x-c)\left\{\frac{1+(x-c)^2+(x-c)^3}{1+(x-c)^2}\right\},$$

then the relation (6) holds for

$$f(X):=f(c)+(X-c)\left\{\frac{1+(X-c)^2+(X-c)^3}{1+(X-c)^2}\right\}$$

as well as for

$$f(X):=f(c)+\frac{(X-c)+(X-c)^3+(X-c)^4}{1+(X-c)^2},$$

no matter how the polynomials involved in these expressions are evaluated.

For polynomials

$$f(x):=\sum_{\nu=0}^{r} a_\nu x^\nu$$

the centered form (8) can be computed by representing f as a Taylor's series at $x=c$ and writing the nonconstant terms as $(x-c)\cdot w(x-c)$. For rational functions (1) one can—after Ratschek [8]—proceed in the following manner.

Let

$$g(x)=\sum_{\nu=0}^{r} a'_\nu(x-c)^\nu$$

and

$$h(x)=\sum_{\nu=0}^{s} b'_\nu(x-c)^\nu$$

be the Taylor polynomials of g and h at $x=c$.

Then we have

$$f(x)=f(c)+(x-c)\cdot w(x-c), \tag{9}$$

where

$$w(y)=\frac{\sum_{\nu=1}^{\max(r,s)}(a'_\nu-f(c)b'_\nu)y^{\nu-1}}{\sum_{\nu=0}^{s} b'_\nu y^\nu}. \tag{10}$$

Despite the good approximation of $W(f,X)$ by (9) and (10), its use has one great disadvantage: if $r \ll s$ then the evaluation of (9) needs many more operations than (1), for example.

Therefore, the question arises: Do there exist "simpler" expressions for f which nevertheless possess the property (6)?

We first note that we can write (10) in the form

$$w(y)=\frac{\sum_{\nu=1}^{r} a'_\nu y^{\nu-1}-f(c)\sum_{\nu=1}^{s} b'_\nu y^{\nu-1}}{\sum_{\nu=0}^{s} b'_\nu y^\nu}. \tag{10'}$$

Exactly in the same way as it was done for (9), (10) in [1, p. 44 ff.], for example, one can show that (6) holds for (9), (10'). We omit the details.

Since $\sum_{\nu=0}^{s} b'_\nu (X-c)^\nu$ can be computed from $\sum_{\nu=1}^{s} b'_\nu (X-c)^{\nu-1}$ by multiplying with $X-c$ and then adding b'_0, the disadvantage of (9), (10) for $r \ll s$ does not exist for (9), (10').

On the other hand it follows by the property of subdistributivity (see [6, p. 13]) that the evaluation of (9), (10) is always contained in the evaluation of (9), (10').

In [1, p. 41 ff.] it was proven that the so-called mean-value form,

$$f(X) := f(c)+(X-c)f'(X), \tag{11}$$

which was introduced by Moore [5, § 6.3], has the property (6) provided $d(f'(X)) \leqq \beta d(X)$. If, for example, f is given by (1), then one can use

$$f'(X) := \frac{h(X)g'(X)-g(X)h'(X)}{h(X)^2}$$

or

$$f'(X) := \frac{g'(X)}{h(X)}-\frac{g(X)h'(X)}{h(X)^2}.$$

In both cases one has to evaluate not only the derivatives of g and h, but also g and h themselves at the interval X.

2. Some new quadratic convergent cases. In the sequel we will show that under certain conditions for g and h, there exist expressions that are simpler than the centered form (9) or the mean-value form (11) for which (6) holds.

In order to simplify the notation, we assume without loss of generality that $h(c)=1$: $h(c)\neq 0$ is necessary in order that f is defined for all $x \in X$. Then we can write f, given by (1) as

$$f(x)=\frac{1}{b'_0}\cdot\frac{\sum_{\nu=0}^{r} a'_\nu (x-c)^\nu}{1+\sum_{\nu=1}^{s}(b'_\nu/b'_0)(x-c)^\nu}.$$

The range of f can be obtained by division of the range of $b'_0 f$ by b'_0. Correspondingly, one obtains an interval-arithmetic evaluation of f by dividing the evaluation of $b'_0 f$ by b'_0. Therefore (6) is proven for f if we have proven this relation for $b'_0 f$. We can therefore assume that f given by (1) has also the following representation:

$$f(x)=\frac{\sum_{\nu=0}^{r} a'_\nu (x-c)^\nu}{1+\sum_{\nu=1}^{s} b'_\nu (x-c)^\nu}. \tag{12}$$

We are now ready to prove the following:

THEOREM 1. (a) *Let $c=m(X)$ ($=$ midpoint of the interval X) and let*

$$0 \notin 1+\sum_{\nu=1}^{s} b'_\nu (X-c)^\nu =: h(X).$$

If

$$f(X) := \frac{g(X)}{h(X)},$$

where

$$g(X) = \sum_{\gamma=0}^{r} a'_\gamma (X-c)^\gamma,$$

then (6) *holds provided we have*

$$\text{sign}\,(a'_1) \cdot \text{sign}\,(b'_1 a'_0) \leqq 0. \tag{13}$$

(Here $g(X)$ and $h(X)$ are supposed to be computed by first computing the powers of $X-c$, then multiplying by the coefficients a'_γ and b'_γ and finally adding these terms. The statement is also true if we use Horner's scheme for the evaluation of the dividend and divisor of (12). More generally, the statement holds if we have $d(g(X)) \leqq c_1 d(X)$ and $d(h(X)) \leqq c_2 d(X)$, where g and h have the representation used in (12), respectively.)

(b) *Let* $c = m(X)$ (= *midpoint of the interval* X) *and let*

$$0 \notin 1 + (X-c)h'(X).$$

Then under the condition (13) *the relation* (6) *holds for*

$$f(X) := \frac{a'_0 + (X-c)g'(X)}{1 + (X-c)h'(X)}. \tag{14}$$

($g'(X)$ is any evaluation of the derivative of g for which $d(g'(X)) \leqq \beta d(X)$ holds. For example, one can use the given representation of g in (1), form the derivative and then use Horner's scheme. The same is true for $h'(X)$.)

Proof. We prove (b). (The proof of (a) can be performed in an analogous way and is in some parts even easier.) First we note that for a real interval

$$A = 1 + [-r, r], \qquad r < 1,$$

the following holds:

$$\frac{1}{A} = \left[\frac{1}{1+r}, \frac{1}{1-r}\right] = A + r^2 \left[\frac{1}{1+r}, \frac{1}{1-r}\right]. \tag{15}$$

Because of $c = m(X)$ the divisor of (14) can be written as

$$1 + (X-c)h'(X) = 1 + [-r, r],$$

where

$$r = \tfrac{1}{2}\, d(X)|h'(X)|.$$

The absolute value $|h'(X)|$ is defined as $|h'(X)| := q(h'(X), 0)$. Using (15) and the subdistributive law of interval arithmetic (see, e.g., [6, p. 13]), we have for (14)

$$f(X) = (a'_0 + (X-c)g'(X))\left\{1 + (X-c)h'(X) + r^2\left[\frac{1}{1+r}, \frac{1}{1-r}\right]\right\}$$

$$\subseteq a'_0 + a'_0(X-c)h'(X) + (X-c)g'(X) + a'_0 r^2 \left[\frac{1}{1+r}, \frac{1}{1-r}\right]$$

$$+ (X-c)^2 g'(X)h'(X) + (X-c)g'(X)r^2\left[\frac{1}{1+r}, \frac{1}{1-r}\right].$$

From this it follows immediately that the width of $f(X)$ satisfies the relation

$$d(f(X)) \leqq |a_0' h'(X)|\, d(X) + |g'(X)|\, d(X) + O(d(X)^2). \tag{16}$$

Correspondingly, we have for the range of f over the interval X the representation

$$W(f, X) = [f(u), f(v)], \tag{17}$$

where $u, v \in X$. Using (12) we have

$$\begin{aligned} f(x) &= \frac{a_0' + a_1'(x-c) + O((x-c)^2)}{1 + b_1'(x-c) + O((x-c)^2)} \\ &= a_0' - a_0' b_1'(x-c) + a_1'(x-c) + O((x-c)^2), \end{aligned}$$

and therefore, by (17),

$$d(W(f, X)) = -a_0' b_1'(v-u) + a_1'(v-u) + O((u-c)^2) + O((v-c)^2).$$

If we have

$$-a_0' b_1' \geqq 0, \qquad a_1' \geqq 0,$$

then, replacing v by x_2 and u by x_1 in the last equation, we have

$$d(W(f, X)) \geqq |a_0' b_1'|\, d(X) + |a_1'|\, d(X) + O(d(X)^2). \tag{18}$$

If, however,

$$-a_0' b_1' \leqq 0, \qquad a_1' \leqq 0,$$

we again get (18) by replacing v by x_1 and u by x_2.

In order to complete the proof, we use the following facts:

1. If $A \subseteq B$, then $q(A, B) \leqq d(B) - d(A)$ (see [1, p. 24]).

2. If $|A| := q(A, 0)$ then $|A| - |B| \leqq q(A, B)$. This follows from the triangle inequality: $q(A, 0) \leqq q(A, B) + q(B, 0)$.

3. Finally, we note that

$$a_0' b_1' \in a_0' h'(X), \qquad a_1' \in g'(X).$$

Using (16) and (18) we have

$$\begin{aligned} q(W(f, X), f(X)) &\leqq d(f(X)) - d(W(f, X)) \\ &\leqq (|a_0' h'(X)| - |a_0' b_1'|)\, d(X) + (|g'(X)| - |a_1'|)\, d(X) + O(d(X)^2) \\ &\leqq \{q(a_0' h'(X), a_0' b_1') + q(g'(X), a_1')\}\, d(X) + O(d(X)^2) \\ &\leqq \{d(a_0' h'(X)) + d(g'(X))\}\, d(X) + O(d(X)^2) \\ &\leqq \alpha\, d(X)^2 \end{aligned}$$

□

As a special case of the preceding theorem, we have the following:

COROLLARY. (1) *If* $h(x) \equiv 1$ *then all the statements of Theorem* 1 *hold.*

(2) *If* $g(x) \equiv 1$ *then all the statements of Theorem* 1 *hold.*

Proof. The proof follows immediately, since in both cases condition (13) holds.

While statement (1) of this corollary is equivalent either to the statements about the centered form of polynomials (if g is evaluated as in part (a) of Theorem 1) or to the statements about the mean-value form (if g is evaluated as in part (b) of the theorem) the second part of the corollary seems to be a new result.

We now illustrate our results by some simple examples.

Example 1.

$$f(x)=\frac{1-x^2}{1+x}, \quad c=m(X)=0, \quad X=[-r,r], \quad r<1.$$

We have $W(f,X)=[1-r,1+r]$ and (13) holds.

The evaluation of f as described in part (a) of Theorem 1 yields

$$f(X)=\left[1-r,\frac{1+r^2}{1-r}\right],$$

and therefore

$$q(W(f,X),f(X))=\frac{2r^2}{1-r}=O(d(X)^2),$$

as predicted by Theorem 1.

If we use the expression defined in part (b) of Theorem 1, namely

$$f(X):=\frac{1-2X\cdot X}{1+X}=\left[\frac{1-2r^2}{1+r},\frac{1+2r^2}{1-r}\right],$$

we also get

$$q(W(f,X),f(X))=\max\left(\frac{r^2}{1+r},\frac{3r^2}{1-r}\right)=O(d(X)^2),$$

as predicted by theory.

Example 2. Although condition (13) seems at first sight to be rather artificial and only necessary for the given proof of Theorem 1, the following example shows that without (13) the statements of Theorem 1 are not true in general. Let

$$f(x)=\frac{1+x}{2+x}, \quad c=m(X)=0, \quad X=[-r,r], \quad r<2.$$

The relation (13) does not hold in this example. We have

$$W(f,X)=\left[\frac{1-r}{2-r},\frac{1+r}{2+r}\right],$$

and both (a) and (b) of Theorem 1 give

$$f(X)=\frac{1+X}{2+X}=\left[\frac{1-r}{2+r},\frac{1+r}{2-r}\right].$$

From this it follows that in this case we only have

$$q(W(f,X),f(X))=\frac{2r(1+r)}{4-r^2}=O(d(X)).$$

Example 3. For the centered form (8) it is well known that (6) is true not only for $c=m(X)$ but for an arbitrary point $c\in X$. The same is true for the mean-value form (11). For these general cases the proofs have been given in [1].

A slight modification of the proof of Theorem 1 shows that the statements of this theorem are also true if $c\neq m(X)$, $c\in X$. We illustrate this as follows. Consider again

the first example,

$$f(x)=\frac{1-x^2}{1+x},$$

and $X=[-r_1, r_2]$. $1>r_1>r_2$. The point $c=0$ is in X but $m(X)\neq 0$ if $r_1\neq r_2$. The evaluation of the expression in part (a) of Theorem 1 gives

$$f(X)=\left[\frac{1-r_1^2}{1+r_2}, \frac{1+r_1r_2}{1-r_1}\right],$$

and since in this case

$$W(f, X)=[1-r_2, 1+r_1]$$

it follows that

$$q(W(f, X), f(X))=\max\left(\frac{r_1^2-r_2^2}{1+r_2}, \frac{r_1(r_1+r_2)}{1-r_1}\right)=O(d(X)^2).$$

3. The *n*-dimensional case. In the remainder of this paper, we discuss the generalization of Theorem 1 to the multidimensional case. Let $x=(x_i)$ be a real n-vector, and let

$$f(x)=\frac{g(x)}{h(x)}$$

be a rational function of the n variables $x_1, x_2, \cdots, x_n$. Then, in the same way as for $n=1$, f can be written as

$$f(x)=\frac{\sum_{\gamma=0}^{r} a_\gamma x^\gamma}{\sum_{\gamma=0}^{s} b_\gamma x^\gamma}. \tag{19}$$

Here, again, a_0 and b_0 represent real numbers but a_γ, $1\leqq\gamma\leqq r$, and b_γ, $1\leqq\gamma\leqq s$, are γ-linear operators from $\mathbb{R}^n$ into $\mathbb{R}$ (see, for example, [7, Def. 17.2]). Furthermore, let there be given n compact real intervals X_i, $1\leqq i\leqq n$. If we denote by $X=(X_i)$ a vector which has the intervals X_i as components, then analogous to (3) we have

$$q(W(f, X), f(X))\leqq\alpha\|(d(X_i))\| \tag{20}$$

for an arbitrary vector norm. Here $f(X)$ is obtained by evaluating (19) for the interval vector $X=(X_i)$.

If one uses the centered form defined analogously to (8), then again it holds that

$$q(W(f, X), f(X))\leqq\alpha\|(d(X_i))\|^2. \tag{21}$$

In [8] Ratschek has given an explicit formula for the centered form in the multidimensional case, which is similar to (10).

Let c be the vector $c=(m(X_i))$, and suppose that $f(x)$ can be represented as

$$f(x)=\frac{\sum_{\gamma=0}^{r} a'_\gamma(x-c)^\gamma}{1+\sum_{\gamma=1}^{s} b'_\gamma(x-c)^\gamma},$$

which—as in the case $n=1$—we can assume without loss of generality.

In order to formulate the next theorem, recall that the linear operators a_1' and b_1' are n-dimensional row vectors:

$$a_1' = (a_{11}', a_{12}', \cdots, a_{1n}'),$$

$$b_1' = (b_{11}', b_{12}', \cdots, b_{1n}').$$

THEOREM 2. (a) *Let* $c = (m(X_i))$ *and*

$$0 \notin 1 + \sum_{\nu=1}^{s} b_\nu'(X-c)^\nu =: h(X).$$

Furthermore, let

$$g(X) := \sum_{\nu=0}^{r} a_\nu'(X-c)^\nu.$$

If

$$\operatorname{sign}(a_{1i}') \cdot \operatorname{sign}(b_{1i}' a_0') \leqq 0, \qquad 1 \leqq i \leqq n, \tag{22}$$

then (21) *holds for*

$$f(X) := \frac{g(X)}{h(X)}.$$

(b) *Let* $c = (m(X_i))$ *and*

$$0 \notin 1 + h'(X)(X-c).$$

Then (21) *holds for*

$$f(X) := \frac{a_0' + g'(X)(X-c)}{1 + h'(X)(X-c)}$$

if (22) *is true.*

(For the evaluation of $g(X)$ and $h(X)$ in part (a) and $h'(X)$ and $g'(X)$ in part (b), the same remarks as in Theorem 1 hold.)

The proof of this theorem is analogous to that of Theorem 1. We omit the details.

The corollary can also be formulated in the multidimensional case in a completely analogous manner.

REFERENCES

[1] G. ALEFELD AND J. HERZBERGER, *Einführung in die Intervallrechnung*, Bibliographisches Institut, Mannheim, 1974.

[2] W. CHUBA AND W. MILLER, *Quadratic convergence in interval arithmetic. I.*, BIT, 12 (1972), pp. 284–290.

[3] E. HANSEN, *The centered form*, in Topics in Interval Analysis, E. Hansen, ed., University Press, Oxford, 1963, pp. 102–106.

[4] W. MILLER, *Quadratic convergence in interval arithmetic. II*, BIT, 12 (1972), pp. 291–298.

[5] R. E. MOORE, *Interval Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1966.

[6] ———, *Methods and Applications of Interval Analysis*, SIAM Studies in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, 1979.

[7] L. B. RALL, *Computational Solution of Nonlinear Operator Equations*, John Wiley, New York, 1969.

[8] H. RATSCHEK, *Centered forms*, this Journal, 17 (1980), pp. 656–662.