Application of Continued Fractions for Fast Evaluation of Certain Functions on a Digital Computer

AMNON BRACHA-BARAK

Abstract—The purpose of this paper is to develop a method for evaluation of certain elementary functions on a digital computer by the use of continued fractions. The time required for this evaluation is drastically reduced by using "short" operations like shift and add, instead of multiplications. Functional consistency is the most important factor that allows the expansion of a function into a continued fraction. Several cases are discussed; in particular the solution of the quadratic equation is discussed in more detail to demonstrate the convergence of the method.

Index Terms—Bilinear transformation, binary arithmetic, continued fractions, quadratic equation, Riccati equation, selection rules.

I. Introduction

The idea of using continued fraction representations for generating a solution to a limited class of quadratics was first introduced by Robertson [3].

Consider the finite continued fraction with k partial numerators p_i and k partial denominators $q_i = 1, 2, \cdots, k$, whose value is A_k/B_k, i.e.,

$$\frac{A_k}{B_k} = \cfrac{p_1}{q_1 + \cfrac{p_2}{q_2 + \cfrac{p_3}{q_3 + \cdots + \cfrac{p_k}{q_k}}}}. \tag{1.1}$$

A standard way of writing (1.1) is

$$\frac{A_k}{B_k} = \frac{p_1}{q_1 +} \frac{p_2}{q_2 +} \frac{p_3}{q_3 +} \cdots + \frac{p_k}{q_k}.$$

A_k and B_k are determined from the recursions:

$$A_i = q_i A_{i-1} + p_i A_{i-2}$$

$$B_i = q_i B_{i-1} + p_i B_{i-2} \qquad i = 2, 3, \cdots, k, \tag{1.2}$$

with initial values:

$$\begin{aligned} A_0 &= 0 \qquad A_1 = p_1 \\ B_0 &= 1 \qquad B_1 = q_1. \end{aligned} \tag{1.3}$$

It is clear that A_k and B_k can be separately and simultaneously determined in two binary arithmetic units in $k - 1$ addition times if the p_i and q_i are chosen to be simple in the binary sense.

The digit set for p_i and q_i for the purposes of this paper is $\{\frac{1}{2}, 1\}$. Since an arithmetic unit of a digital computer is used, all the approximations in this paper are up to a finite number of binary digits. We will therefore define binary values as all the values that a given binary arithmetic unit can assume. It will be proved in Section II that the continued fraction A_k/B_k assumes in the limit all the binary values over the interval $[(\sqrt{2} - 1)/2, \sqrt{2}]$. This range includes $[\frac{1}{2}, 1]$, the range of normalized floating point binary fractions. This property indicates that a suitable continued fraction representation exists, such that conversion to conventional binary can be achieved by repetitive use of two binary adders in parallel, followed by a division to determine the quotient A_k/B_k.

The main reason for selecting $p_i, q_i \in \{\frac{1}{2}, 1\}, i = 1, 2, \cdots$, is that the four multiplicative operations required for each iteration in (1.2) are reduced to "shift" and "add" operations. These operations will be called "short" operations throughout this paper, mainly because the time required to perform these operations is shorter than the time required to perform "long" operations, e.g., multiplication, division.

The purpose of this paper is to develop algorithms for fast evaluation of certain elementary functions by using "short" operations in several registers simultaneously. In order to be able to do so we make use of functional consistency which will be defined in Section II.

Determination of selection rules for p and q in each iteration is an important step for the development of the algorithm. Selection rules were extensively studied by Trivedi [5], where a complete set of such rules were developed for the quadratic equation. The set of selection rules that is used in this paper is described in Section III.

In Section IV, we generalize our results to a higher degree

Manuscript received May 15, 1972; revised July 9, 1973. This work was supported in part by the National Science Foundation under Grant US NSF GJ813.

The author was with the Department of Computer Science, University of Illinois, Urbana-Champaign, Ill. He is now with the Hebrew University, Jerusalem, Israel.

polynomial, and in Section V, we show two more cases where our analysis is applicable.

II. Preliminaries

In this section we develop some concepts of continued fractions that will be used throughout this paper.

The general continued fraction will be regarded as a sequence of bilinear transformations of the form:

$$f_k = \frac{p_k}{q_k + f_{k+1}}, \quad k = 1, 2, \cdots, \tag{2.1}$$

where $f_k(x)$ is a function of x.

$$f_1 = \frac{p_1}{q_1 +}\frac{p_2}{q_2 +}\cdots + \frac{p_n}{q_n + f_{n+1}} \tag{2.2}$$

$$= \frac{A_n + f_{n+1}A_{n-1}}{B_n + f_{n+1}B_{n-1}}, \quad n = 1, 2, \cdots, \tag{2.3}$$

where the functions A_n and B_n satisfy the recursion (1.2) with the initial values given by (1.3).

Let $f_1(x) = F(x, 1)$ be defined over the interval $[m, M]$. We will expand $f_1(x)$ into a continued fraction such as (2.1) and require that the choice of p_1 and q_1 be made such that $f_2(x) = F(x, 2)$ is also defined over $[m, M]$. Since $f_2(x)$ is a continued fraction we can use the same rules of selection for p_2 and q_2.

We now define the term functional consistency.

Definition: For a substitution of the form

$$f_k(x) = \frac{p_k}{q_k + f_{k+1}(x)}, \quad k = 1, 2, \cdots$$

where $f_k(x) = F(k, x)$ is a function of k and x, p_k and q_k are constants; if $f_k(x)$ and $f_{k+1}(x)$ are defined over the same interval, then we have functional consistency.

We now give a general proof of convergence for continued fractions with positive elements. Functional consistency will be required in order to assure that only one set of selection rules for p and q are used.

Theorem 1: Let

$$f_1(x) = \frac{A}{B} = \frac{p_1}{q_1 +}\frac{p_2}{q_2 +}\cdots + \frac{p_n}{q_n + f_{n+1}(x)}, \quad n = 1, 2, \cdots$$

be the continued fraction expansion of $f_1(x)$ with positive elements, p and q. Assume that functional consistency exists for all $f_i(x)$, $i = 1, 2, \cdots$. Let A_n/B_n be the nth approximation to $f_1(x)$.

Then for every $\epsilon > 0$, there exists N, such that for all $n > N$,

$$\delta_n = \left| \frac{A}{B} - \frac{A_n}{B_n} \right| < \epsilon.$$

Proof: We will study relations between δ_{2n}, δ_{2n-2} and δ_{2n+1}, δ_{2n-1}, because of the well-known property that in continued fractions with positive elements, the convergents of even order generate a monotonically increasing sequence, which has a limit, and the odd convergents generate a monotonically decreasing sequence with a limit. The value of the continued fraction is greater than that of any of its even convergents and less than that of any of its odd convergents [1], [2].

Define

$$T_{2n} = \frac{\delta_{2n}}{\delta_{2n-2}}, \quad T_{2n+1} = \frac{\delta_{2n+1}}{\delta_{2n-1}}, \quad n = 1, 2, \cdots,$$

then if we prove that for all n, $T_n < 1$, it will follow that the limit of convergents of even order and the limit of convergents of odd order are equal to the value of the continued fraction. We note that the values in the numerator and denumerator of T have equal signs; therefore the absolute value sign can be omitted.

We have [2, ch. 1]:

$$T_n = \frac{A/B - A_n/B_n}{A/B - A_{n-2}/B_{n-2}}$$

$$= \frac{(-1)^{n+2}p_1p_2 \cdots p_n p_{n+1}/BB_n}{(-1)^n p_1 p_2 \cdots p_{n-1}(q_{n+1}q_n + p_{n+1})/BB_{n-2}}$$

$$= \frac{p_n \cdot p_{n+1}}{q_{n+1} \cdot q_n + p_{n+1}} \cdot \frac{B_{n-2}}{B_n}.$$

Since $p_{n+1} = f_{n+1} = p_n/f_n - q_n$, $q_{n+1} = 1$ and $B_n = q_nB_{n-1} + p_nB_{n-2}$, we conclude that:

$$T_n = \frac{p_n - q_n f_n}{p_n + q_n \dfrac{B_{n-1}}{B_{n-2}}} = \frac{1 - \dfrac{q_n}{p_n} f_n}{1 + \dfrac{q_n}{p_n}\dfrac{B_{n-1}}{B_{n-2}}}.$$

All the quantities and T_n are positive, $T_n < 1$, and the result follows. Q.E.D.

The assumption in Theorem 1 is that an algorithm for finding p and q in each step exists. In the remaining part of this section we will show one such algorithm.

First, we find extreme values, m and M, for a continued fraction of the form (1.1).

Theorem 2: Let the k approximant to $f_1(x)$ be

$$\frac{A_k}{B_k} = \frac{p_1}{q_1} + \frac{p_2}{q_2} + \cdots + \frac{p_k}{q_k};$$

and let $p_i, q_i \in \left\{\frac{1}{2}, 1\right\}, i = 1, 2, \cdots$.

Then we have in the limit:

$$M = \max f_1(x) = \max \lim_{k \to \infty} A_k/B_k = \sqrt{2}$$

$$m = \min f_1(x) = \min \lim_{k\to\infty} A_k/B_k = (\sqrt{2}-1)/2.$$

Proof: For the maximum value M we have,

$$M = \frac{\max p_1}{\min q_1} + \frac{\min p_2}{\max q_2 + M}.$$

We substitute now the maximum and minimum values for p and q and solve for M. The result is $M^2 = 2$, and since the first and second convergents [1, theorem 154] are positive, it follows that $M = +\sqrt{2}$.

Similarly we have for m:

$$m = \frac{\min p_1}{\max q_1} + \frac{\max p_2}{\min q_2 + m}.$$

m is positive and can be found by solving the quadratic equation

$$4m^2 + 4m - 1 = 0.$$

The result is $m = (\sqrt{2}-1)/2$. Note that in general, if all p and q are positive, so are m and M. Q.E.D.

We will use now the analysis of Theorem 1 to study the rate of convergence, by finding an upper bound to T_n:

$$\max T_n = \max \frac{1 - \frac{q_n}{p_n} f_n}{1 + \frac{q_n}{p_n}\frac{B_{n-1}}{B_{n-2}}}$$

$$\leqslant \frac{1 - \min\left(\frac{q_n}{p_n} f_n\right)}{1 + \min\left(\frac{q_n}{p_n}\frac{B_{n-1}}{B_{n-2}}\right)},$$

$$\min \frac{q_n}{p_n} f_n = \tfrac{1}{2}\min f_n = \frac{\sqrt{2}-1}{4} \cong 0.1035;$$

$$\min\left(\frac{q_n}{p_n}\cdot\frac{B_{n-1}}{B_{n-2}}\right) \leqslant \tfrac{1}{2} + \min\frac{B_{n-1}}{B_{n-2}},$$

but B_{n-1}/B_{n-2} is a continued fraction whose value is

$$\frac{B_{n-1}}{B_{n-2}} = q_{n-1} + \frac{p_{n-1}}{q_{n-2}} + \frac{p_{n-2}}{q_{n-3}} + \cdots + \frac{p_2}{q_1}.$$

Therefore,

$$\min\frac{B_{n-1}}{B_{n-2}} \leqslant \tfrac{1}{2} + \frac{\sqrt{2}-1}{2} = \frac{\sqrt{2}}{2}.$$

Finally we have,

$$\max T_n \leqslant \frac{1 - \frac{\sqrt{2}-1}{4}}{1 + \frac{\sqrt{2}}{4}} = \frac{3-\sqrt{2}}{4+\sqrt{2}} \cong 0.2929.$$

Therefore, for n sufficiently large, the absolute error δ_n is reduced by a factor which is less than or equal to 0.2929 for each pair of additional iterations.

We will prove now that all values, $y, y \in [m, M]$, can be approximated by a continued fraction of the form (1.1). In particular we show a method, also primitive, for selecting p and q in each step.

Algorithm 1: Let y be a given real constant, $y \in [m, M]$. Then there exists a continued fraction x, of the form (1.1) with $p, q \in \{\tfrac{1}{2}, 1\}$, such that in the limit $x = y$.

We first note that in practice we approximate y with only a finite number of steps and by Theorem 1, if such x exists, than with a sufficient number of steps and given ϵ, $|y - x| < \epsilon$. Also note that y is a real constant that can be represented by a binary computer.

In order to find x we define continued fractions x_i, of length i, $i = 1, 2, \cdots$.

$$x_i = \frac{p_1}{q_1} + \frac{p_2}{q_2} + \cdots + \frac{p_i}{q_i + \theta},$$

where $\theta \in \{m, M\}$.

We start with $i = 1$, and a given y in the interval $[m, M]$. Define $x_1 = p_1/(q_1 + \theta)$, and substitute all possible values for (p, q, θ) in the following order: $(\tfrac{1}{2}, 1, m)$; $(\tfrac{1}{2}, 1, M)$; $(\tfrac{1}{2}, \tfrac{1}{2}, m)$; $(\tfrac{1}{2}, \tfrac{1}{2}, M)$; $(1, 1, m)$; $(1, 1, M)$; $(1, \tfrac{1}{2}, m)$ and $(1, \tfrac{1}{2}, M)$. These values when substituted in x_1 define four subintervals over $[m, M]$. Simple analysis shows that the subintervals cover the entire interval $[m, M]$, with some regions of overlap between each consecutive subintervals. The result is that y is included in at least one such subinterval, and therefore we select the corresponding p and q as p_1 and q_1. We increase i by one and study $x_2 = p_1/q_1 + p_2/q_2 + \theta$, where now p_1 and q_1 are fixed and we substitute all values for p_2, q_2, and θ. Again there will be at least one subinterval which will contain y and we can select p_2 and q_2. Our process can be carried now for increasing values of i until a given precision is reached.

Algorithm 1 is inefficient because there are several multiplication and division operations in each step. In the next section we develop a more efficient set of selection rules for p and q.

III. Solution of $ax^2 + bx - c = 0$

We now show how to solve a quadratic equation with two distinct roots of opposite signs and in particular, a square root problem.

Let

$$a_1 x_1^2 + b_1 x_1 - c_1 = 0 \qquad (3.1)$$

be a given quadratic equation, a_1 and c_1 are positive and b_1 is a nonnegative constant.

The substitution we use is of the form

$$x_i = \frac{p_i}{q_i + x_{i+1}} \tag{3.2}$$

where

$$p_i, q_i \in \left\{\tfrac{1}{2}, 1\right\} \qquad i = 1, 2, \cdots.$$

For the kth step we have

$$a_k \frac{p_k^{\,2}}{(q_k + x_{k+1})^2} + b_k \frac{p_k}{q_k + x_{k+1}} - c_k = 0 \qquad k = 1, 2, \cdots$$

or

$$c_k x_{k+1}^{\,2} + (2c_k q_k - b_k p_k) x_{k+1} + c_k q_k^{\,2} - a_k p_k^{\,2} - b_k p_k q_k = 0.$$

The recursion that follows is

$$\begin{aligned} \lambda a_{k+1} &= c_k \\ \lambda b_{k+1} &= 2c_k q_k - b_k p_k \\ \lambda c_{k+1} &= -c_k q_k^{\,2} + (a_k p_k + b_k q_k) p_k \qquad k = 1, 2, \cdots \end{aligned} \tag{3.3}$$

where λ is a nonzero constant that can be used for normalization.

The resulting quadratic equation is

$$a_{k+1} x_{k+1}^{\,2} + b_{k+1} x_{k+1} - c_{k+1} = 0. \tag{3.4}$$

This method of approximating the solution of (3.1) can be used if we develop a technique for selecting p_k and q_k, $k = 1, 2, \cdots$, from the coefficients of the kth quadratic equation, i.e., a_k, b_k, and c_k.

Using the results of Section II it can be seen that functional consistency of the procedure can be achieved in each step if

$$m \leqslant x_k \leqslant M \qquad k = 1, 2, \cdots. \tag{3.5}$$

By imposing condition (3.5) we need only one set of selection rules for p_k and q_k, $k = 1, 2, \cdots$ for the range $[m, M]$.

We develop now a set of selection rules for p_k and q_k, $k = 1, 2, \cdots$, for the quadratic equation.

We write below a version of (3.1).

Let

$$x_1 = \frac{c_1}{b_1 + a_1 x_1} \tag{3.6}$$

where it is assumed that $c_1 > 0$, $b_1 \geqslant 0$, $a_1 > 0$, and $m \leqslant x_1 \leqslant M$.

We will find p_1 and q_1 such that

$$x_1 = \frac{p_1}{q_1 + x_2} \tag{3.7}$$

where $m \leqslant x_2 \leqslant M$ and $p_1, q_1 \in \left\{\tfrac{1}{2}, 1\right\}$.

Clearly, we have four possibilities, and for each pair of p_1 and q_1 we get a different x_2.

We take now the inverse approach. We assume that condition (3.5) exists for x_2, and find the range of x_1 for each pair of p_1 and q_1. We start with a pair $p_1 = 1$ and $q_1 = \tfrac{1}{2}$. From (3.7) we have for $x_2 = \sqrt{2}$ (lower bound for x_1)

$$x_1 \geqslant \frac{1}{\tfrac{1}{2} + \sqrt{2}} = \frac{2(2\sqrt{2} - 1)}{7} \cong 0.522, \tag{3.8}$$

and for $x_2 = (\sqrt{2} - 1)/2$ (upper bound for x_1)

$$x_1 \leqslant \frac{1}{\tfrac{1}{2} + \dfrac{\sqrt{2} - 1}{2}} = \sqrt{2} \cong 1.414. \tag{3.9}$$

The result is that for x_1 in the range defined by (3.8)-(3.9), we can choose $p_1 = 1$ and $q_1 = \tfrac{1}{2}$. Since x_1 is the unknown we use (3.6) in order to find the allowable range for $p_1 = 1$ and $q_1 = \tfrac{1}{2}$.

We have

$$1.414 \geqslant \frac{c_1}{b_1 + a_1 x_1} \geqslant 0.522. \tag{3.10}$$

Since (3.10) is possible for any x_1 in the range (3.8)-(3.9), we conclude that for the range

$$0.522 b_1 + 0.522^2 a_1 \leqslant c_1 \leqslant \sqrt{2}\, b_1 + 2a_1 \tag{3.11}$$

we choose $p_1 = 1$ and $q_1 = \tfrac{1}{2}$.

Similarly we write below the ranges for each of the remaining possibilities.

For

$$(\sqrt{2} - 1) b_1 + (\sqrt{2} - 1)^2 a_1 \leqslant c_1 \leqslant 2(\sqrt{2} - 1) b_1 + 4(\sqrt{2} - 1)^2 a_1, \quad \text{choose } p_1 = 1, q_1 = 1; \tag{3.12}$$

$$\frac{2\sqrt{2} - 1}{7} b_1 + \left(\frac{2\sqrt{2} - 1}{7}\right)^2 a_1 \leqslant c_1 \leqslant \frac{\sqrt{2}}{2} b_1 + \tfrac{1}{2} a_1, \quad \text{choose } p_1 = \tfrac{1}{2}, q_1 = \tfrac{1}{2}; \tag{3.13}$$

$$\frac{\sqrt{2} - 1}{2} b_1 + \left(\frac{\sqrt{2} - 1}{2}\right)^2 a_1 \leqslant c_1 \leqslant (\sqrt{2} - 1) b_1 + (\sqrt{2} - 1)^2 a_1, \quad \text{choose } p_1 = \tfrac{1}{2}, q_1 = 1. \tag{3.14}$$

The result is that the entire range $[m, M]$ is divided into four sections, (3.11)-(3.14), and for each section we can choose a pair of p_1 and q_1 such that condition (3.5) for x_2 be satisfied.

Clearly, if we have to do two multiplications for each selection range in order to find p_1 and q_1, our procedure is inefficient. In the analysis that follows we make use of an important feature of the ranges defined in (3.11)-(3.14); this is the existence of overlapping between any two consecutive ranges.

This means that in the overlap regions we have a freedom of selecting the pairs p_1 and q_1 between the two sets of such constants. We will use this freedom in order to simplify our selection algorithm by defining a line of selection inside the overlap region such that the coefficients of b_1 and a_1 will be simple in the binary sense.

Before we define the selection lines, we note that rate of convergence of the method was found to be strongly dependent on these lines. The first set of selection lines can be the upper lines in each range, for

$$\begin{aligned} &c_1 - 0.414b_1 \leqslant 0.1713a_1 && \text{then } p_1 = \tfrac{1}{2},\ q_1 = 1;\\ &c_1 - 0.707b_1 \leqslant 0.5a_1 && \text{then } p_1 = \tfrac{1}{2},\ q_1 = \tfrac{1}{2};\\ &c_1 - 0.828b_1 \leqslant 0.686a_1 && \text{then } p_1 = 1,\ q_1 = 1;\\ & && \text{otherwise } p_1 = 1,\ q_1 = \tfrac{1}{2}. \end{aligned} \quad (3.15)$$

Experimentally, this set of rules gave the best rate of convergence. Although written for (3.1), we see that it is valid for (3.4) for the general subscript k.

To simplify the constants that appear in (3.15) we use simple binary constants with at most two nonzero binary digits.

We have, for (3.4) with $k = 1, 2, \cdots$,

$$\begin{aligned} &c_k - 0.375b_k \leqslant 0.15625a_k && \text{then } p_k = \tfrac{1}{2},\ q_k = 1;\\ &c_k - 0.625b_k \leqslant 0.5a_k && \text{then } p_k = \tfrac{1}{2},\ q_k = \tfrac{1}{2};\\ &c_k - 0.75b_k \leqslant 0.625a_k && \text{then } p_k = 1,\ q_k = 1;\\ & && \text{otherwise } p_k = 1,\ q_k = \tfrac{1}{2}. \end{aligned} \quad (3.16)$$

These selection rules involve only short operations.

The algorithm described in this section involves the following steps, for the k iterations, starting with $k = 1$.

Step 1: Equation (3.4) is given, then use selection rules (3.16) to find p_k and q_k.

Step 2: Use the results of Step 1 and iterate on (1.2).

Step 3: Use the recursion (3.3) and find (3.4) for $k + 1$.

Step 4: Check if A_k/B_k reached the required precision. This check can be done only once if the number of iterations required to achieve certain precision is known. The analysis above for the rate of convergence gives the necessary information to find such numbers. If the required precision is not reached proceed to Step 1 for $k + 1$.

Theorem 1 of the last section assures convergence to the solution. Table I is a numerical example.

IV. Solution of a Higher Degree Polynomial

We now show how one solution of the cubic equation can be found by the method of Section III.

Let

$$a_1x_1^3 + b_1x_1^2 + c_1x_1 - d_1 = 0 \quad (4.1)$$

be a given cubic equation. We use the substitution (3.2) and we get for the kth step:

$$a_k \frac{p_k^3}{(q_k + x_{k+1})^3} + b_k \frac{p_k^2}{(q_k + x_{k+1})^2} + c_k \frac{p_k}{q_k + x_{k+1}} - d_k = 0$$

or

$$d_kx_{k+1}^3 + (3d_kq_k - c_kp_k)x_{k+1}^2 + (3d_kq_k^2 - 2c_kp_kq_k - b_kp_k^2)x_{k+1} - (a_kp_k^3 + b_kp_k^2q_k + c_kp_kq_k^2 - d_kq_k^3) = 0.$$

The recursion relations between the coefficients of the kth cubic equation and the $(k + 1)$st cubic equation are, therefore,

$$\begin{aligned} a_{k+1} &= d_k\\ b_{k+1} &= 3d_kq_k - c_kp_k\\ c_{k+1} &= 3d_kq_k^2 - 2c_kp_kq_k - b_kp_k^2\\ d_{k+1} &= a_kp_k^3 + b_kp_k^2q_k + c_kp_kq_k^2 - d_kq_k^3 \qquad k = 1, 2, \cdots. \end{aligned} \quad (4.2)$$

The resulting cubic equation is

$$a_{k+1}x_{k+1}^3 + b_{k+1}x_{k+1}^2 + c_{k+1}x_{k+1} - d_{k+1} = 0.$$

For the selection rules we use an analysis similar to that of Section III. First observe that the bounds given in (3.8)-(3.9) for the case $p = 1$ and $q = \frac{1}{2}$ are valid. Therefore, we can write an expression, similar to (3.10) for the cubic equation:

$$\sqrt{2} \geqslant \frac{d_k}{a_1x_1^2 + b_1x_1 + c_1} \geqslant 0.522. \quad (4.3)$$

The result is that for

$$0.522c_1 + 0.522^2b_1 + 0.522^3a_1 \leqslant d_1 \leqslant \sqrt{2}c_1 + 2b_1 + 2\sqrt{2}a_1,$$

we choose $p_1 = 1$ and $q_1 = \frac{1}{2}$.

For the remaining cases we have

$$\begin{aligned} &(\sqrt{2} - 1)c_1 + (\sqrt{2} - 1)^2b_1 + (\sqrt{2} - 1)^3a_1 \leqslant d_1\\ &\qquad \leqslant 2(\sqrt{2} - 1)c_1 + 4(\sqrt{2} - 1)^2b_1 + 8(\sqrt{2} - 1)^3a_1\\ &\qquad\qquad \text{choose } p_1 = 1, q_1 = 1;\\ &\frac{2\sqrt{2} - 1}{7}c_1 + \left(\frac{2\sqrt{2} - 1}{7}\right)^2b_1 + \left(\frac{2\sqrt{2} - 1}{7}\right)^3a_1 \leqslant d_1\\ &\qquad \leqslant \frac{\sqrt{2}}{2}c_1 + \tfrac{1}{2}b_1 + \frac{\sqrt{2}}{4}a_1\\ &\qquad\qquad \text{choose } p_1 = \tfrac{1}{2}, q_1 = \tfrac{1}{2}; \end{aligned}$$

TABLE I

Solution of (x-0.4)(x+0.5) = 0

k	a_k	b_k	c_k	p_k	q_k	A_k	B_k	A_k/B_k	Error
1	0.100000D 01	0.100000D 00	0.200000D 00	0.5	0.5	0.500000D 00	0.500000D 00	0.1000000000000000D 01	-0.600000D 00
2	0.200000D 00	0.150000D 00	0.225000D 00	1.0	1.0	0.500000D 00	0.150000D 01	0.3333333333333333D 00	0.666667D-01
3	0.225000D 00	0.300000D 00	0.125000D 00	0.5	1.0	0.750000D 00	0.175000D 01	0.4285714285714286D 00	-0.285714D-01
4	0.125000D 00	0.100000D 00	0.812500D-01	0.5	0.5	0.625000D 00	0.162500D 01	0.3846153846153846D 00	0.153846D-01
5	0.812500D-01	0.312500D-01	0.359375D-01	0.5	0.5	0.687500D 00	0.168750D 01	0.4074074074074074D 00	-0.740741D-02
6	0.359375D-01	0.203125D-01	0.191406D-01	0.5	0.5	0.656250D 00	0.165625D 01	0.3962264150943396D 00	0.377359D-02
7	0.191406D-01	0.898437D-02	0.927734D-02	0.5	0.5	0.671875D 00	0.167188D 01	0.4018691588785046D 00	-0.186916D-02
8	0.927734D-02	0.478516D-02	0.471191D-02	0.5	0.5	0.664063D 00	0.166406D 01	0.3990610328638498D 00	0.938967D-03
9	0.471191D-02	0.231934D-02	0.233765D-02	0.5	0.5	0.667969D 00	0.166797D 01	0.4004683840749414D 00	-0.468384D-03
10	0.233765D-02	0.117798D-02	0.117340D-02	0.5	0.5	0.666016D 00	0.166602D 01	0.3997655334114888D 00	0.234467D-03
11	0.117340D-02	0.584412D-03	0.585556D-03	0.5	0.5	0.666992D 00	0.166699D 01	0.4001171646162859D 00	-0.117165D-03
12	0.585556D-03	0.293350D-03	0.293064D-03	0.5	0.5	0.666504D 00	0.166650D 01	0.3999414005273952D 00	0.585995D-04
13	0.293064D-03	0.146389D-03	0.146461D-03	0.5	0.5	0.666748D 00	0.166675D 01	0.4000292954445584D 00	-0.292954D-04
14	0.146461D-03	0.732550D-04	0.732481D-04	0.5	0.5	0.666626D 00	0.166663D 01	0.3999853512048634D 00	0.146488D-04
15	0.732481D-04	0.366151D-04	0.366196D-04	0.5	0.5	0.666687D 00	0.166669D 01	0.4000073241293441D 00	-0.732413D-05
16	0.366196D-04	0.183120D-04	0.183109D-04	0.5	0.5	0.666656D 00	0.166666D 01	0.3999963378682731D 00	0.366213D-05
17	0.183109D-04	0.915490D-05	0.915518D-05	0.5	0.5	0.666672D 00	0.166667D 01	0.4000018310490996D 00	-0.183105D-05
18	0.915518D-05	0.457773D-05	0.457766D-05	0.5	0.5	0.666664D 00	0.166666D 01	0.3999990844712593D 00	0.915529D-06
19	0.457766D-05	0.228880D-05	0.228881D-05	0.5	0.5	0.666668D 00	0.166667D 01	0.4000004577533226D 00	-0.457763D-06
20	0.228881D-05	0.114442D-05	0.114441D-05	0.5	0.5	0.666666D 00	0.166667D 01	0.3999997711180767D 00	0.228882D-06
21	0.114441D-05	0.572203D-06	0.572204D-06	0.5	0.5	0.666667D 00	0.166667D 01	0.4000001144408961D 00	-0.114441D-06
22	0.572204D-06	0.286103D-06	0.286102D-06	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999427795355D 00	0.572205D-07
23	0.286102D-06	0.143051D-06	0.143051D-06	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000286102281D 00	-0.286102D-07
24	0.143051D-06	0.715256D-07	0.715256D-07	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999856948849D 00	0.143051D-07
25	0.715256D-07	0.357628D-07	0.357628D-07	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000071525573D 00	-0.715256D-08
26	0.357628D-07	0.178814D-07	0.178814D-07	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999964237213D 00	0.357628D-08
27	0.178814D-07	0.894070D-08	0.894070D-08	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000017881393D 00	-0.178814D-08
28	0.894070D-08	0.447035D-08	0.447035D-08	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999991059303D 00	0.894070D-09
29	0.447035D-08	0.223517D-08	0.223517D-08	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000004470348D 00	-0.447035D-09
30	0.223517D-08	0.111759D-08	0.111759D-08	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999997764826D 00	0.223517D-09
31	0.111759D-08	0.558794D-09	0.558794D-09	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000001117587D 00	-0.111759D-09
32	0.558794D-09	0.279397D-09	0.279397D-09	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999999441206D 00	0.558794D-10
33	0.279397D-09	0.139698D-09	0.139698D-09	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000000279397D 00	-0.279397D-10
34	0.139698D-09	0.698492D-10	0.698492D-10	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999999860302D 00	0.139598D-10
35	0.698492D-10	0.349246D-10	0.349246D-10	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000000069849D 00	-0.698491D-11
36	0.349246D-10	0.174623D-10	0.174623D-10	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999999965075D 00	0.349246D-11
37	0.174623D-10	0.873114D-11	0.873116D-11	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000000017462D 00	-0.174623D-11
38	0.873116D-11	0.436559D-11	0.436557D-11	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999999991269D 00	0.873121D-12
39	0.436557D-11	0.218277D-11	0.218279D-11	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000000004355D 00	-0.436554D-12
40	0.218279D-11	0.109141D-11	0.109139D-11	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999999997817D 00	0.218284D-12
41	0.109139D-11	0.545683D-12	0.545704D-12	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000000001091D 00	-0.109135D-12
42	0.545704D-12	0.272862D-12	0.272842D-12	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999999999454D 00	0.545675D-13
43	0.272842D-12	0.136411D-12	0.136431D-12	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000000000273D 00	-0.272837D-13
44	0.136431D-12	0.682257D-13	0.682053D-13	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999999999854D 00	0.136419D-13
45	0.682053D-13	0.340924D-13	0.341129D-13	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000000000068D 00	-0.681399D-14
46	0.341129D-13	0.170667D-13	0.170462D-13	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999999999956D 00	0.341394D-14
47	0.170462D-13	0.851287D-14	0.853333D-14	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000000000017D 00	-0.170697D-14
48	0.853333D-14	0.427590D-14	0.425644D-14	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999999999991D 00	0.860423D-15
49	0.425644D-14	0.211799D-14	0.213845D-14	0.5	0.5	0.666667D 00	0.166667D 01	0.4000000000000004D 00	-0.430211D-15
50	0.213845D-14	0.107946D-14	0.105899D-14	0.5	0.5	0.666667D 00	0.166667D 01	0.3999999999999998D 00	0.208167D-15

$$\frac{\sqrt{2}-1}{2}c_1 + \left(\frac{\sqrt{2}-1}{2}\right)^2 b_1 + \left(\frac{\sqrt{2}-1}{2}\right)^3 a_1 \leqslant d_1$$

$$\leqslant (\sqrt{2}-1)c_1 + (\sqrt{2}-1)^2 b_1 + (\sqrt{2}-1)^3 a_1,$$

$$\text{choose } p_1 = \tfrac{1}{2},\ q_1 = 1.$$

We are ready now to write a set of selection rules similar to (3.15), i.e., the upper line in each range will be our selection line:

$$\begin{aligned}
&d_1 - 0.414c_1 - 0.1713b_1 \leqslant 0.071a_1 && \text{then } p_1 = \tfrac{1}{2}, q_1 = 1;\\
&d_1 - 0.707c_1 - 0.5b_1 \leqslant 0.3535a_1 && \text{then } p_1 = \tfrac{1}{2}, q_1 = \tfrac{1}{2};\\
&d_1 - 0.828c_1 - 0.686b_1 \leqslant 0.5683a_1 && \text{then } p_1 = 1, q_1 = 1;\\
&\text{otherwise } p_1 = 1, q_1 = \tfrac{1}{2}. && \qquad (4.4)
\end{aligned}$$

We note that as in (3.16) the constants which appear in (4.4) can be simplified.

For the proof of convergence and the rate of convergence we can use the analysis of Section II, and therefore we developed a method to approximate one positive solution of a cubic equation.

The procedure can be generalized now to higher degree polynomials. The necessary steps are as follows.

Step 1: To write the recursion for the coefficients of the polynomial.

Step 2: To develop the selection rules by using an argument similar to (3.10) and (4.3).

Step 3: To simplify the coefficients in the selection rules.

The result is an algorithm which always converges to one positive solution. Table II is a numerical example.

V. Riccati Equation

The main purpose of this section is to find the family of functions for which bilinear transformations of the form (2.1) can be used with functional consistency.

Consider the Riccati equation

$$f_1{}' + a_1 f_1{}^2 + b_1 f_1 + c_1 = 0$$

where $f_1(x)$ is a function of the variable x, and a_1, b_1, and c_1 are functions of x or constants. The property of this equation as noted by Wynn [7] is that if the dependent variable f_1 is

TABLE II

Solution of (x-0.6)(x+0.5)(x+0.9) = 0

k	a_k	b_k	c_k	d_k	p_k	q_k	A_k	B_k	A_k/B_k	Error
1	0.10000D 01	0.80000D 00	-0.39000D 00	0.27000D 00	0.5	0.5	0.50000D 00	0.50000D 00	0.1000000000000000D 01	-0.40000D 00
2	0.27000D 00	0.60000D 00	0.19750D 00	0.14250D 00	0.5	1.0	0.50000D 00	0.10000D 01	0.5000000000000000D 00	0.10000D 00
3	0.14250D 00	0.32880D 00	0.80000D-01	0.14000D 00	0.5	0.5	0.50000D 00	0.75000D 00	0.6666666666666666D 00	-0.66670D-01
4	0.14000D 00	0.17000D 00	-0.17190D-01	0.51410D-01	0.5	0.5	0.50000D 00	0.87500D 00	0.5714285714285714D 00	0.28570D-01
5	0.51410D-01	0.85700D-01	0.46480D-02	0.30180D-01	0.5	0.5	0.50000D 00	0.81250D 00	0.6153846153846154D 00	-0.15380D-01
6	0.30180D-01	0.42940D-01	-0.11180D-02	0.13950D-01	0.5	0.5	0.50000D 00	0.84380D 00	0.5925925925925926D 00	0.74070D-02
7	0.13950D-01	0.21480D-01	0.28500D-03	0.72560D-02	0.5	0.5	0.50000D 00	0.82810D 00	0.6037735849056604D 00	-0.37740D-02
8	0.72560D-02	0.10740D-01	-0.70570D-04	0.35570D-02	0.5	0.5	0.50000D 00	0.83590D 00	0.5981308411214953D 00	0.18690D-02
9	0.35570D-02	0.53710D-02	0.17730D-04	0.17950D-02	0.5	0.5	0.50000D 00	0.83200D 00	0.6009389671361502D 00	-0.93900D-03
10	0.17950D-02	0.26860D-02	-0.44210D-05	0.89370D-03	0.5	0.5	0.50000D 00	0.83400D 00	0.5995316159250585D 00	0.46840D-03
11	0.89370D-03	0.13430D-02	0.11070D-05	0.44800D-03	0.5	0.5	0.50000D 00	0.83300D 00	0.6002344665885111D 00	-0.23450D-03
12	0.44800D-03	0.67140D-03	-0.27650D-06	0.22370D-03	0.5	0.5	0.50000D 00	0.83350D 00	0.5998828353837141D 00	0.11720D-03
13	0.22370D-03	0.33570D-03	0.69150D-07	0.11190D-03	0.5	0.5	0.50000D 00	0.83330D 00	0.6000585994725047D 00	-0.58600D-04
14	0.11190D-03	0.16780D-03	-0.17280D-07	0.55940D-04	0.5	0.5	0.50000D 00	0.83340D 00	0.5999707045554415D 00	0.29300D-04
15	0.55940D-04	0.83920D-04	0.43210D-08	0.27980D-04	0.5	0.5	0.50000D 00	0.83330D 00	0.6000146487951365D 00	-0.14650D-04
16	0.27980D-04	0.41960D-04	-0.10800D-08	0.13990D-04	0.5	0.5	0.50000D 00	0.83330D 00	0.5999926758706559D 00	0.73240D-05
17	0.13990D-04	0.20980D-04	0.27010D-09	0.69940D-05	0.5	0.5	0.50000D 00	0.83330D 00	0.6000036621317269D 00	-0.36620D-05
18	0.69940D-05	0.10490D-04	-0.67520D-10	0.34970D-05	0.5	0.5	0.50000D 00	0.83330D 00	0.5999981689509004D 00	0.18310D-05
19	0.34970D-05	0.52450D-05	0.16880D-10	0.17480D-05	0.5	0.5	0.50000D 00	0.83330D 00	0.6000009155287407D 00	-0.91550D-06
20	0.17480D-05	0.26230D-05	-0.42200D-11	0.87420D-06	0.5	0.5	0.50000D 00	0.83330D 00	0.5999995422365774D 00	0.45780D-06
21	0.87420D-06	0.13110D-05	0.10550D-11	0.43710D-06	0.5	0.5	0.50000D 00	0.83330D 00	0.6000002288819232D 00	-0.22890D-06
22	0.43710D-06	0.65570D-06	-0.26370D-12	0.21860D-06	0.5	0.5	0.50000D 00	0.83330D 00	0.5999998855591038D 00	0.11440D-06
23	0.21860D-06	0.32780D-06	0.65920D-13	0.10930D-06	0.5	0.5	0.50000D 00	0.83330D 00	0.6000000572204544D 00	-0.57220D-07
24	0.10930D-06	0.16390D-06	-0.16470D-13	0.54640D-07	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999713897719D 00	0.28610D-07
25	0.54640D-07	0.81960D-07	0.41040D-14	0.27320D-07	0.5	0.5	0.50000D 00	0.83330D 00	0.6000000143051151D 00	-0.14310D-07
26	0.27320D-07	0.40980D-07	-0.10130D-14	0.13660D-07	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999928474427D 00	0.71530D-08
27	0.13660D-07	0.20490D-07	0.24020D-15	0.68300D-08	0.5	0.5	0.50000D 00	0.83330D 00	0.5000000035762787D 00	-0.35760D-08
28	0.68300D-08	0.10240D-07	-0.46980D-16	0.34150D-08	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999982118607D 00	0.17880D-08
29	0.34150D-08	0.51220D-08	-0.13140D-17	0.17070D-08	0.5	0.5	0.50000D 00	0.83330D 00	0.5000000008940697D 00	-0.89410D-09
30	0.17070D-08	0.25610D-08	0.13390D-16	0.85370D-09	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999995529651D 00	0.44700D-09
31	0.85370D-09	0.12810D-08	-0.16410D-16	0.42690D-09	0.5	0.5	0.50000D 00	0.83330D 00	0.6000000002235174D 00	-0.22350D-09
32	0.42690D-09	0.64030D-09	0.17160D-16	0.21340D-09	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999998882413D 00	0.11180D-09
33	0.21340D-09	0.32010D-09	-0.17350D-16	0.10670D-09	0.5	0.5	0.50000D 00	0.83330D 00	0.6000000000558794D 00	-0.55880D-10
34	0.10670D-09	0.16010D-09	0.17400D-16	0.53360D-10	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999999720603D 00	0.27940D-10
35	0.53360D-10	0.80040D-10	-0.17410D-16	0.26680D-10	0.5	0.5	0.50000D 00	0.83330D 00	0.6000000000139698D 00	-0.13970D-10
36	0.26680D-10	0.40020D-10	0.17410D-16	0.13340D-10	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999999930151D 00	0.69850D-11
37	0.13340D-10	0.20010D-10	-0.17410D-16	0.66700D-11	0.5	0.5	0.50000D 00	0.83330D 00	0.6000000000034924D 00	-0.34920D-11
38	0.66700D-11	0.10000D-10	0.17410D-16	0.33350D-11	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999999982538D 00	0.17460D-11
39	0.33350D-11	0.50020D-11	-0.17410D-16	0.16670D-11	0.5	0.5	0.50000D 00	0.83330D 00	0.6000000000008731D 00	-0.87310D-12
40	0.16670D-11	0.25010D-11	0.17410D-16	0.83370D-12	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999999995634D 00	0.43660D-12
41	0.83370D-12	0.12510D-11	-0.17410D-16	0.41690D-12	0.5	0.5	0.50000D 00	0.83330D 00	0.6000000000002183D 00	-0.21830D-12
42	0.41690D-12	0.62530D-12	0.17410D-16	0.20840D-12	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999999998908D 00	0.10910D-12
43	0.20840D-12	0.31260D-12	-0.17410D-16	0.10420D-12	0.5	0.5	0.50000D 00	0.83330D 00	0.5000000000000545D 00	-0.54570D-13
44	0.10420D-12	0.15630D-12	0.17410D-16	0.52100D-13	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999999999727D 00	0.27280D-13
45	0.52100D-13	0.78140D-13	-0.17410D-16	0.26060D-13	0.5	0.5	0.50000D 00	0.83330D 00	0.6000000000000136D 00	-0.13640D-13
46	0.26060D-13	0.39100D-13	0.17410D-16	0.13020D-13	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999999999932D 00	0.68140D-14
47	0.13020D-13	0.19520D-13	-0.17410D-16	0.65190D-14	0.5	0.5	0.50000D 00	0.83330D 00	0.5000000000000034D 00	-0.34140D-14
48	0.65190D-14	0.97870D-14	0.17410D-16	0.32510D-14	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999999999983D 00	0.17070D-14
49	0.32510D-14	0.48680D-14	-0.17410D-16	0.16340D-14	0.5	0.5	0.50000D 00	0.83330D 00	0.5000000000000008D 00	-0.86040D-15
50	0.16340D-14	0.24600D-14	0.17410D-16	0.80840D-15	0.5	0.5	0.50000D 00	0.83330D 00	0.5999999999999996D 00	0.43020D-15

replaced by the bilinear transformation (2.1), then the functions f_k, $k = 1, 2, \cdots$ also satisfy the Riccati equation

$$f_k' + a_k f_k^2 + b_k f_k + c_k = 0. \tag{5.1}$$

We develop below the recursion for the coefficients of the $(k+1)$st equation by means of the coefficients of the kth equation.

Let (5.1) be designated as the kth Riccati equation, and assume that this is satisfied by $y_k = f_k$.

From (2.1) we have

$$y_k = \frac{p_k}{q_k + y_{k+1}}, \quad p_k, q_k \in \{\tfrac{1}{2}, 1\}$$

then since

$$y_k' = \frac{-p_k y_{k+1}'}{(q_k + y_{k+1})^2}$$

we have

$$\frac{-p_k y_{k+1}'}{(q_k + y_{k+1})^2} + a_k \frac{p_k^2}{(q_k + y_{k+1})^2} + b_k \frac{p_k}{q_k + y_{k+1}} + c_k = 0.$$

If we multiply by $-(q_k + y_{k+1})^2/p_k$ to normalize the coefficient of y_{k+1}', we get

$$y_{k+1}' - \frac{c_k}{p_k} y_{k+1}^2 - \left(b_k + \frac{2c_k q_k}{p_k}\right) y_{k+1} - \left(a_k p_k + b_k q_k + \frac{c_k q_k^2}{p_k}\right) = 0$$

and y_{k+1} will satisfy the $(k+1)$th Riccati equation if

$$\begin{aligned} a_{k+1} &= -\frac{c_k}{p_k} \\ b_{k+1} &= -b_k - \frac{2c_k q_k}{p_k} \\ c_{k+1} &= -a_k p_k - b_k q_k - \frac{c_k q_k^2}{p_k} \qquad k = 1, 2, \cdots. \end{aligned} \tag{5.2}$$

We note that all the operations involved in (5.2) require only "short" operations, since both p_k and q_k are simple binary constants.

Theorem 3: If $\{y_k\}$ satisfies the Riccati equation, then $\Delta = b_k^2 - 4a_kc_k$ is independent of k [4].

Proof: We use the recursion (5.2) and get

$$b_{k+1}^2 - 4a_{k+1}c_{k+1} = b_k^2 + \frac{4b_kc_kq_k}{p_k} + \frac{4c_k^2q_k^2}{p_k^2} - 4a_kc_k$$

$$- \frac{4b_kc_kq_k}{p_k} - \frac{4c_k^2q_k^2}{p_k^2} = b_k^2 - 4a_kc_k.$$

Q.E.D.

In the analysis that follows we find continued fraction expansions of certain functions by the use of the general solution of the Riccati equation. Let

$$y' + ay^2 + by + c = 0 \tag{5.3}$$

be a Riccati equation, with a, b, and c constants.

In order to find the solution of (5.3) we integrate by parts:

$$\frac{dy}{ay^2 + by + c} = -dx \tag{5.4}$$

and the solutions are

$$\int \frac{dy}{ay^2 + by + c} = \frac{1}{\sqrt{(b^2 - 4ac)}} \ln \frac{2ay + b - \sqrt{(b^2 - 4ac)}}{2ay + b + \sqrt{(b^2 - 4ac)}};$$

$$= \frac{1}{\sqrt{(b^2 - 4ac)}} \operatorname{arctanh} \frac{b + 2ay}{\sqrt{(b^2 - 4ac)}},$$

$$\text{when } b^2 - 4ac > 0;$$

$$= \frac{-2}{b + 2ay}, \quad \text{when } b^2 - 4ac = 0;$$

$$= \frac{2}{\sqrt{(4ac - b^2)}} \arctan \frac{2ay + b}{\sqrt{(4ac - b^2)}},$$

$$\text{when } b^2 - 4ac < 0. \tag{5.5}$$

The preceding solutions can be used now for the continued fraction expansion of the inverse functions which appear explicitly in the solution.

We start with the case

$$y = \tan x. \tag{5.6}$$

The Riccati equation for (5.6) is

$$y' = y^2 + 1, \quad y(0) = 0.$$

We note that $-\Delta = \sqrt{4ac - b^2} = 2$.

Now we use a bilinear transformation of the form (2.1). The result is a differential equation of the type (5.1) with the recursion (5.2). By Theorem 3 it follows that the solution for each equation k is of the type (5.5) with $-\Delta > 0$, and therefore we get for the kth step:

$$-x_k - d_k = \arctan \frac{2a_ky_k + b_k}{2}$$

where d_k is a constant of integration.

The solution is therefore

$$y_k = -\frac{b_k}{2a_k} - \frac{1}{a_k} \tan (x_k + d_k) \qquad k = 1, 2, \cdots.$$

Except for the first part of the solution which is a linear transformation, we see the consistency of the method, because if a set of selection rules are developed for $\tan x$ it can be used for each step and therefore evaluation of this function will be possible.

Another important function which can be included is e^x. We have

$$y' = y, \quad \Delta = 1.$$

The kth step solution is

$$-x_k - d_k = \ln \frac{2a_ky_k + b_k - 1}{2a_ky_k + b_k + 1},$$

or

$$y_k = -\frac{b_k + 1}{2a_k} - \frac{1}{a_k(e^{-x_k} - 1)} \qquad k = 1, 2, \cdots.$$

Again we note that if a set of selection rules can be developed for e^x then it is possible to carry the process for each step and therefore to find the continued fraction expansion for the exponential function.

For the case where $\Delta = 0$ we have several possibilities.

1) $b = a = 0, y' + c = 0$ with the solution $y = -cx + d$.
2) $b = c = 0, y' + ay^2 =$ with the solution

$$y_k = \frac{1}{ax + b}.$$

3) $a \neq 0, b \neq 0, c \neq 0$, and $b^2 - 4ac = 0$.

Case 3) is particularly interesting because there exists two constants S and T such that

$$ay^2 + by + c = (Sy + T)^2$$

and by Theorem 3, this relation is true in each step of the iteration. The resulting Riccati equation has the form:

$$y_k' + (-1)^{k+1} V(x)(S_ky_k + T_k)^2 = 0 \quad k = 1, 2, \cdots$$

where

$$S_k^2 = a_k;$$

$$2S_kT_k = b_k;$$

$$T_k^2 = c_k;$$

and $V(x)$ is a function of x or a constant.

The recursion relations for S_k and T_k that follow from (5.2) are

$$S_{k+1} = T_k/\sqrt{p_k}$$

and

$$T_{k+1} = (S_k p_k + T_k q_k)/\sqrt{p_k}.$$

Since $\sqrt{p_k}$ is not a desired feature, $p_k = 1$ can be assumed in each step.

It is anticipated that the solution of many functions can be expanded as a continued fraction, provided that an adequate set of selection rules for p_k and q_k can be found.

Acknowledgment

The author wishes to thank Prof. J.E. Robertson of the University of Illinois, for many helpful discussions.

References

[1] G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers*. New York: Oxford, 1954.

[2] A.N. Khovanskii, *The Application of Continued Fractions and Their Generalizations to Problems in Approximation Theory*. Groningen, The Netherlands: Noordhoff, 1963.

[3] J.E. Robertson, Quarterly Progress Rep., Dept. Computer Science, Univ. of Illinois, Urbana, pp. 121-126, July 1970.

[4] —, private communication, Jan. 1972.

[5] K.S. Trivedi, "An algorithm for the solution of a quadratic equation using continued fractions," M.S. thesis, Dept. Comput. Sci., Univ. of Illinois, Urbana, 1972.

[6] H.S. Wall, *Analytic Theory of Continued Fractions*. New York: Van Nostrand , 1948.

[7] P. Wynn, "Five lectures on the numerical application of continued fractions," Mathematics Res. Center, Univ. of Wisconsin, Madison,

[8] —, "On some recent development in the theory and application of continued fractions," *SIAM J. Numerical Analysis*, vol. 1, pp. 177-197, 1964.

Amnon Bracha-Barak was born in Tel-Aviv, Israel, on August 19, 1942. He received the B.Sc. degree in mathematics from the Technion, Israel Institute of Technology, Haifa, Israel, in 1967 and the M.S. and Ph.D. degrees in computer science from the University of Illinois, Urbana-Champaign, in 1970 and 1971, respectively.

From 1963 to 1968 he was a scientific programmer in the Department of Applied Mathematics, the Weizmann Institute of Science, Israel. From 1968 to 1972 he was a Research Assistant and Postdoctorate in the Department of Computer Science, University of Illinois, and he is presently an Assistant Professor of Computer Science, Hebrew University, Jerusalem, Israel. His research interests are computer arithmetic, numerical analysis and pattern recognition.

Dr. Bracha-Barak is a member of Sigma Xi and the Association for Computing Machinery.

Data Manipulating Functions in Parallel Processors and Their Implementations

TSE-YUN FENG

Abstract—**This paper shows that there exists a class of functions called data manipulating functions (DMF's), in sequential as well as parallel processors. The circuits used to achieve these functions can be considered to form an independent functional block, called a data manipulator. A basic organization applicable to both sequential and parallel processors is then suggested. The main deviation of a parallel processor organization from the conventional Von Neumann organization is seen to be in the bit-slice (bis) manipulating functions. A comprehensive set of bis manipulating functions from the categories of permuting, replicating, spacing and masking is given. Implementation of the last category, the masking functions, is usually through a mask register by defining its content (mask pattern). It is found that for many operations the required mask patterns are periodic and/or monotonic. The upper bounds of generating these patterns are found. The techniques and designs of two data manipulators for the first three categories of DMF's (permuting, replicating, spacing) are given. Periodic and monotonic mask patterns are also used to help in implementing some of these functions. In addition, it is shown that the data manipulator designs presented in this paper are extremely flexible to suit the requirements of various parallel processors.**

Index Terms—**Cell communications, data manipulating functions, data manipulator, logic design, parallel processing, parallel processor organization, processing characteristics.**

Manuscript received August 20, 1972; revised May 24, 1973. This work was supported by the Rome Air Development Center under Contract F30602-72-C-0281.

The author is with the Department of Electrical and Computer Engineering, Syracuse University, Syracuse, N.Y.

Introduction

It is well known that as the switching speeds of computer devices approach a limit, any further improvement in computer throughput has to be in increasing the number of bits which can be processed simultaneously. Thus, given the same cycle time the slowest method of processing is by *bit-serial* (one bit at a time). The processing speed is increased by an order of magnitude or more when a number of bits, called a word, can be processed simultaneously. This