Description of Generalized Continued Fractions by Finite Automata

Jeffrey Shallit

In memory of Alf van der Poorten, the sorcerer of continued fractions

Abstract A generalized continued fraction algorithm associates every real number x with a sequence of integers; x is rational iff the sequence is finite. For a fixed algorithm A, call a sequence of integers *valid* if it is the result of A on some input x_0. We show that, if the algorithm is sufficiently well behaved, then the set of all valid sequences is accepted by a finite automaton.

1 Introduction

Simple continued fractions are finite expressions of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots + \cfrac{1}{a_n}}},$$

usually abbreviated $[a_0, a_1, \ldots, a_n]$, or infinite expressions of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}},$$

J. Shallit (✉)
School of Computer Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1
e-mail: shallit@cs.uwaterloo.ca

J.M. Borwein et al. (eds.), *Number Theory and Related Fields: In Memory of Alf van der Poorten*, Springer Proceedings in Mathematics & Statistics 43,
DOI 10.1007/978-1-4614-6642-0_17,

usually abbreviated $[a_0, a_1, \ldots]$. The latter expression is defined as the limit as $n \to \infty$, if it exists, of the corresponding finite expression ending in a_n. The a_i are called the *partial quotients* of the continued fraction.

The standard bracket notation above for continued fractions conflicts with the standard notation $[x, y]$ for closed intervals of the real line. We abuse notation by using both and trust that the reader can disambiguate if necessary.

The *simple continued fraction algorithm*, on the other hand, is the following algorithm that, given a real number x, produces a finite sequence of partial quotients $a_0, a_1, \ldots, a_n$ or infinite sequence of partial quotients $a_0, a_1, \ldots$, such that $x = [a_0, a_1, \ldots, a_n]$ or $x = [a_0, a_1, \ldots]$.

Algorithm SCF(x)**; outputs** $(a_0, a_1, \ldots)$**:**
SCF1. Set $x_0 \leftarrow x$; set $i \leftarrow 0$.
SCF2. Set $a_i \leftarrow \lfloor x_i \rfloor$.
SCF3. If $a_i = x_i$ then stop. Otherwise set $x_{i+1} \leftarrow 1/(x_i - a_i)$; set $i \leftarrow i+1$ and go to step SCF2.

For example,

$$\begin{aligned}
\mathrm{SCF}(\frac{52}{43}) &= (1,4,1,3,2) \\
\mathrm{SCF}(\pi) &= (3,7,15,1,292,\ldots) \\
\mathrm{SCF}(\sqrt{2}) &= (1,2,2,2,\ldots) \\
\mathrm{SCF}(e) &= (2,1,2,1,1,4,1,1,6,1,1,8,\ldots)
\end{aligned}$$

In the literature, these two different concepts

- The *function* mapping $(a_0, \ldots, a_n)$ to the rational number $[a_0, \ldots, a_n]$, or $(a_0, a_1, \ldots)$ to the real number $[a_0, a_1, \ldots,]$; and
- The *algorithm* taking x as input and producing the a_i

have, unfortunately, often been confused. This is probably due to two reasons: the fact that the birth of continued fractions [4] long predates the appreciation of algorithms as mathematical objects [10] and the fortunate happenstance that by imposing some simple rules on the partial quotients, we can ensure that that there is exactly one valid expansion for each real number.

The concept of "rules" that describe the set of possible outputs of a continued fraction expansion has appeared before in many places. For example, Hurwitz [7] used them to describe a nearest integer continued fraction algorithm in $\mathbb{Z}[i]$. It is our goal to formalize the concept and explain it in terms of automata theory.

As has long been known, if x is a real irrational number, then the set of outputs produced by the SCF algorithm is exactly

$$\{(a_0, a_1, \ldots) \ : \ \forall i \ a_i \in \mathbb{Z} \text{ and } a_i \geq 1 \text{ for } i \geq 1 \ \}.$$

On the other hand, if x is a rational number, then the set of outputs produced by the SCF algorithm is exactly

$$\{(a_0, a_1, \ldots, a_n) \ : \ \forall i \ a_i \in \mathbb{Z} \text{ and } a_i \geq 1 \text{ for } 1 \leq i \leq n \text{ and } a_n \geq 2 \text{ if } n \geq 1 \ \}.$$

Hence, if we insist that in any expression $[a_0, a_1, \ldots]$, we must have

- $\forall i \ a_i \in \mathbb{Z}$;
- $a_i \geq 1$ for $i \geq 1$;
- If the expansion terminates with a_n, then $a_n \geq 2$;

then the ambiguity between the function and the algorithm disappears. The question remains about how we could *discover* rules like this. This becomes important because there exist many other versions of the continued fraction algorithm, and we would like to have a similar characterization of the outputs.

For example, the *ceiling algorithm* (CCF) replaces the use of the floor function with the ceiling; that is, it replaces step SCF2 with

SCF2′. Set $a_i \leftarrow \lceil x_i \rceil$.

For example,

$$\begin{aligned}
\mathrm{CCF}(\frac{52}{43}) &= [2, -1, -3, -1, -3, -2] \\
\mathrm{CCF}(\pi) &= [4, -1, -6, -15, -1, -292, \ldots] \\
\mathrm{CCF}(\sqrt{2}) &= [2, -1, -1, -2, -2 - 2, -2, -2, \ldots] \\
\mathrm{CCF}(e) &= [3, -3, -1, -1, -4, -1, -1, -6, -1, -1, -8, \ldots]
\end{aligned}$$

The expansions produced by CCF include negative partial quotients, and obey the following rules:

- $\forall i \ a_i \in \mathbb{Z}$.
- $a_i \leq -1$ for $i \geq 1$.
- If the expansion ends with a_n and $n \geq 1$, then $a_n \neq -1$.

Indeed, it is easy to see that if

$$\mathrm{SCF}(-x) = [a_0, a_1, a_2, \ldots],$$

then

$$\mathrm{CCF}(x) = [-a_0, -a_1, -a_2, \ldots].$$

Yet another expansion is the so-called *nearest integer continued fraction* (NICF). It is generated by an algorithm similar to SCF above, except that step SCF2 is replaced by

SCF2″. Set $a_i \leftarrow \lfloor x_i + \frac{1}{2} \rfloor$.

For example,

$$\begin{aligned} \text{NICF}(\frac{52}{43}) &= (1,5,-4,-2) \\ \text{NICF}(\pi) &= (3,7,16,-294,\ldots) \\ \text{NICF}(\sqrt{2}) &= (1,2,2,2,2,\ldots) \\ \text{NICF}(e) &= (3,-4,2,5,-2,-7,2,9,-2,-11,\ldots) \end{aligned}$$

The partial quotients generated by NICF satisfy the following rules:

- $\forall i \; a_i \in \mathbb{Z}$;
- $a_i \le -2$ or $a_i \ge 2$ for $i \ge 1$;
- If $a_i = -2$ then $a_{i+1} \le -2$;
- If $a_i = 2$ then $a_{i+1} \ge 2$; and
- If the expansion terminates with a_n, then $a_n \ne 2$.

(Actually, the NICF is usually described slightly differently in the literature, but our formulation is essentially the same. See [8].)

In this paper, we are concerned with the following questions:

1. Which functions f are suitable replacements for the floor function in algorithm SCF (i.e., yield generalized continued fraction algorithms)?
2. Which of these functions correspond to generalized continued fraction algorithms that have "easily describable" outputs (i.e., accepted by a finite automaton)?

In this paper, we will answer question (1) by *fiat*, and then examine the consequences for question (2). Before we do, however, we mention a useful connection with another famous type of continued fraction.

2 Semiregular Continued Fractions

There is a close relationship between the continued fractions we study here and what is called *semiregular continued fractions* in the literature. A semiregular continued fraction is a finite or infinite expression of the form

$$b_0 + \cfrac{\varepsilon_1}{b_1 + \cfrac{\varepsilon_2}{b_2 + \cfrac{\varepsilon_3}{b_3 + \cdots}}} \tag{1}$$

where

- $b_i \in \mathbb{Z}$;
- $b_i \geq 1$ for $i \geq 1$;
- $\varepsilon_i = \pm 1$ for $i \geq 1$;
- $b_i + \varepsilon_{i+1} \geq 1$ for $i \geq 1$.

It is easily seen that (1) is equivalent to

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}, \tag{2}$$

where $a_i = \varepsilon_1 \varepsilon_2 \cdots \varepsilon_i b_i$ for $i \geq 0$. This expresses a semiregular continued fraction in the form that we study.

This connection extends to the convergents. Setting, as usual, $p_{-1} = 1$, $q_{-1} = 0$, $p_0 = a_0$, and $q_0 = 1$, and $p_n = a_n p_{n-1} + p_{n-2}$, $q_n = a_n q_{n-1} + q_{n-2}$, the theory of continued fractions (or an easy induction) gives $p_n/q_n = [a_0, \ldots, a_n]$. The convergents to a semiregular continued fraction are defined similarly: $p'_{-1} = 1$, $q'_{-1} = 0$, $p'_0 = b_0$, $q'_0 = 1$, $p'_n = b_n p'_{n-1} + \varepsilon_n p'_{n-2}$, and $q'_n = b_n q'_{n-1} + \varepsilon_n q'_{n-2}$. If the a_i and b_i are related as above, an easy induction gives

$$\begin{aligned} p'_{2n} &= \varepsilon_2 \varepsilon_4 \cdots \varepsilon_{2n} p_{2n} \\ q'_{2n} &= \varepsilon_2 \varepsilon_4 \cdots \varepsilon_{2n} q_{2n} \end{aligned}$$

and

$$\begin{aligned} p'_{2n-1} &= \varepsilon_1 \varepsilon_3 \cdots \varepsilon_{2n-1} p_{2n-1} \\ q'_{2n-1} &= \varepsilon_1 \varepsilon_3 \cdots \varepsilon_{2n-1} q_{2n-1} \end{aligned}$$

for $n \geq 1$.

We will need the following classical results on semiregular continued fractions (see, e.g., [13, 18, §37, 38, pp. 135–143]):

Theorem 2.1. *Let*

$$b_0 + \cfrac{\varepsilon_1}{b_1 + \cfrac{\varepsilon_2}{b_2 + \cfrac{\varepsilon_3}{b_3 + \cdots}}}$$

be a semiregular continued fraction obeying the rules above. Then

(a) The sequence p'_n/q'_n *converges;*
(b) $\lim_{n\to\infty} |q'_n| = +\infty$*;*
(c) For a given infinite sequence of signs $(\varepsilon_i)_{i\geq 1}$*, the expansion of a given irrational real number exists and is unique, provided* $b_i + \varepsilon_{i+1} \geq 2$ *infinitely often.*

3 Real Integer Functions and Finite Automata

Let $f : \mathbb{R} \to \mathbb{Z}$. We say f is a *real integer function* if

(a) $|f(x) - x| < 1$ for all $x \in \mathbb{R}$;
(b) $f(x+j) = f(x) + j$ for all $x \in \mathbb{R}$, $j \in \mathbb{Z}$.

Examples include the floor function $f(x) = \lfloor x \rfloor$, the ceiling function $f(x) = \lceil x \rceil$, and the round function $f(x) = \lfloor x + \frac{1}{2} \rfloor$.

Real integer functions induce generalized continued fraction algorithms by imitating algorithm SCF above:

Algorithm $\mathrm{CF}_f(x)$**;** outputs $(a_0, a_1, \ldots)$:
CF1. Set $x_0 \leftarrow x$; set $i \leftarrow 0$.
CF2. Set $a_i \leftarrow f(x_i)$.
CF3. If $a_i = x_i$ then stop. Otherwise set $x_{i+1} \leftarrow 1/(x_i - a_i)$, $i \leftarrow i+1$ and go to step CF2.

For each such expansion, we have an associated sequence of *convergents* p_n and q_n, defined as in Sect. 2. The theory of continued fractions (or an easy induction) gives $p_n/q_n = [a_0, \ldots, a_n]$ and furthermore $x = [a_0, \ldots, a_n, x_{n+1}]$, for all $n \geq 0$.

Now we examine the properties of the expansion of rational numbers.

Theorem 3.2. *Let f be an integer function and let x be a real number. The algorithm* $\mathrm{CF}_f(x)$ *terminates iff x is rational. Furthermore, if* $\mathrm{CF}_f(x)$ *terminates, with $(a_0, a_1, \ldots a_n)$ as output, then $x = [a_0, a_1, \ldots, a_n]$.*

Proof. Suppose x is rational. The algorithm successively replaces x_i by x_{i+1}. Suppose $i \geq 1$ and $x_i = p/q$ for $p, q \in \mathbb{Z}$ with $q \geq 1$. If $x_i \in \mathbb{Z}$, that is, if $q \mid p$, then $x_i = f(x_i)$ and the algorithm terminates immediately. Then either $a_i = \lfloor x_i \rfloor$ or $a_i = \lceil x_i \rceil$. Since $x_{i+1} = (x_i - a_i)^{-1}$, we have either $x_{i+1} = q/(p \bmod q)$ or $x_{i+1} = -q/((-p) \bmod q)$. In both cases we have have replaced a denominator of q with a number strictly less than q. Thus after at most $q-1$ steps, we will reach a denominator of 1, and the algorithm terminates.

For the other direction, an easy induction gives $x = [a_0, a_1, \ldots, a_{n-1}, x_n]$. If the algorithm terminates, then $x_n = a_n$ and we have $x = [a_0, a_1, \ldots, a_n]$, a rational function of the integers $a_0, \ldots, a_n$. □

Next, we prove two useful lemmas. The first concerns occurrences of partial quotients ± 1, and the second concerns convergents.

Lemma 3.3. *Suppose f is an integer function and let a_i and x_i be defined as in the algorithm* CF_f*. Then*

(a) If $a_i = 1$ for $i \geq 1$, then $x_{i+1} > 1$ and $a_{i+1} \geq 1$.
(b) If $a_i = -1$ for $i \geq 1$, then $x_{i+1} < -1$ and $a_{i+1} \leq -1$.
(c) There exists no i such that $a_{i+t} = (-1)^t 2$ for $t \geq 0$.

Proof. (a) Suppose $a_i = 1$. Then there is a corresponding x_i from the algorithm with $f(x_i) = a_i$. Since $i \geq 1$, we have $x_i > 1$ or $x_i < -1$. Since $|x_i - f(x_i)| < 1$ by the definition of integer function, we have $|x_i - 1| < 1$. It follows that $1 < x_i < 2$, so $x_{i+1} = 1/(x_i - 1)$ satisfies $x_{i+1} > 1$. Since $|x_{i+1} - a_{i+1}| < 1$, we have $a_{i+1} \geq 1$.

(b) Analogous to (a).

(c) It is easy to see that $1 = [2, -2, 2, -2, \ldots]$. If there exists i such that $a_{i+t} = (-1)^t 2$ for $t \geq 0$, then in the algorithm we have $x_i = a_i - 1$, a contradiction.

□

Lemma 3.4. *If* $|q_n| \leq |q_{n-1}|$ *then* $|q_{n-2}| < |q_{n-1}|$ *and either*

(a) $\operatorname{sgn} q_{n-1} \neq \operatorname{sgn} q_{n-2}$ *and* $a_n = 1$*; or*
(b) $\operatorname{sgn} q_{n-1} = \operatorname{sgn} q_{n-2}$ *and* $a_n = -1$.

In both cases we have $|q_{n+1}| > |q_n|$.

Proof. We verify this by induction. Consider the smallest index n such that $|q_n| \leq |q_{n-1}|$. Then necessarily $|q_{n-2}| < |q_{n-1}|$. Suppose $\operatorname{sgn} q_{n-1} \neq \operatorname{sgn} q_{n-2}$. Since $q_n = a_n q_{n-1} + q_{n-2}$, if $a_n \leq -1$ or $a_n \geq 2$ then $|q_n| = |a_n q_{n-1} + q_{n-2}| > |q_{n-1}|$, a contradiction. So $a_n = 1$, and furthermore $\operatorname{sgn} q_{n-1} = \operatorname{sgn} q_n$.

Now it follows from Lemma 3.3 that $a_{n+1} \geq 1$, so $|q_{n+1}| = |a_{n+1} q_n + q_{n-1}| \geq |q_n + q_{n-1}| > |q_n|$.

The analogous analysis works if $\operatorname{sgn} q_{n-1} = \operatorname{sgn} q_{n-2}$.

Since $|q_{n+1}| \geq |q_n|$, if we let n' be the smallest index $n' > n$ such that $|q_{n'}| \leq |q_{n'-1}|$, then $n' > n+1$ and $|q_{n'-2}| < |q_{n'-1}|$. Now we are in exactly the same situation as above, with n' replacing n, and the induction step proceeds in the same way. □

Next, we discuss the properties of the expansion of irrational numbers. We start by characterizing those real numbers whose expansion has a given prefix.

For a list $\mathbf{a} = (a_0, a_1, a_2, \ldots, a_n, \ldots)$ containing at least $n+1$ elements, we let $\operatorname{pref}_n(\mathbf{a}) = (a_0, \ldots, a_n)$ be the prefix consisting of the first $n+1$ elements.

Theorem 3.5. *Let* f *be an integer function, and* $a_0, a_1, \ldots$ *a sequence of integers. Define* $S_f(0) = f^{-1}[0]$ *and*

$$S_f(n) = (S_f(n-1)^{-1} \cap f^{-1}[a_n]) - a_n = (S_f(n-1)^{-1} - a_n) \cap f^{-1}[0]$$

for $n \geq 1$. *Then*

$$S_f(n) = \{\xi \in \mathbb{R} \;:\; \operatorname{pref}_n(\mathrm{CF}_f([a_0, a_1, \ldots, a_{n-1}, a_n + \xi])) = (a_0, a_1, \ldots, a_n)\}$$

for $n \geq 0$.

Proof. By induction on n. The base case is $n = 0$. Here $\operatorname{pref}_0(\mathrm{CF}_f([a_0 + \xi])) = (a_0)$ iff $f(a_0 + \xi) = a_0$ iff $f(\xi) = 0$, and so $\xi \in f^{-1}[0]$, as required.

Now assume the result is true for $n' \leq n$; we prove it for $n+1$. By definition $S_f(n+1) = (S_f(n)^{-1} - a_{n+1}) \cap f^{-1}[a_{n+1}]$. Then

$$\begin{aligned}
\xi \in S_f(n+1) &\iff \xi \in S_f(n)^{-1} - a_{n+1} \text{ and } \xi \in f^{-1}[0] \\
&\iff a_{n+1} + \xi \in S_f(n)^{-1} \text{ and } \xi \in f^{-1}[0] \\
&\iff (a_{n+1} + \xi)^{-1} \in S_f(n) \text{ and } \xi \in f^{-1}[0] \\
&\iff \mathrm{pref}_n(\mathrm{CF}_f([a_0, \ldots, a_{n-1}, a_n + (a_{n+1} + \xi)^{-1}])) \\
&\qquad = (a_0, \ldots, a_n) \text{ and } \xi \in f^{-1}[0] \\
&\iff \mathrm{pref}_n(\mathrm{CF}_f([a_0, \ldots, a_{n-1}, a_n, a_{n+1} + \xi])) \\
&\qquad = (a_0, \ldots, a_n) \text{ and } \xi \in f^{-1}[0] \\
&\iff \mathrm{pref}_{n+1}(\mathrm{CF}_f([a_0, \ldots, a_{n-1}, a_n, a_{n+1} + \xi])) = (a_0, \ldots, a_{n+1}),
\end{aligned}$$

which completes the proof. □

Corollary 3.6. *Let f be an integer function, and let the sets $S_f(n)$ be defined as in the previous theorem. Then for $n \geq 0$ we have*

$$\{x \in \mathbb{R} \ : \mathrm{pref}_n(\mathrm{CF}_f(x)) = (a_0, a_1, \ldots, a_n)\} = \{\frac{p_n + \xi p_{n-1}}{q_n + \xi q_{n-1}} \ : \ \xi \in S_f(n)\}.$$

Next, we prove an approximation theorem.

Theorem 3.7. *Suppose* $\mathrm{pref}_n(\mathrm{CF}_f(x)) = (a_0, a_1, \ldots, a_n)$. *Then*

$$\left| x - \frac{p_n}{q_n} \right| \leq \left| \frac{1}{q_n} \right|.$$

Proof. From Corollary 3.6 we know that $x = \frac{p_n + \xi p_{n-1}}{q_n + \xi q_{n-1}}$ for some $\xi \in S_f(n)$. But each $S_f(n)$ is a subset of $f^{-1}[0]$, so $-1 < \xi < 1$. Hence

$$\begin{aligned}
\left| x - \frac{p_n}{q_n} \right| &= \left| \frac{p_n + \xi p_{n-1}}{q_n + \xi q_{n-1}} - \frac{p_n}{q_n} \right| \\
&= \left| \frac{\xi}{q_n(q_n + \xi q_{n-1})} \right| \\
&< \left| \frac{1}{q_n(q_n + \xi q_{n-1})} \right|.
\end{aligned}$$

Now suppose $|q_n| > |q_{n-1}|$. Then since $|\xi| < 1$, and the q_i are integers, we have $|q_n + \xi q_{n-1}| \geq 1$.

Otherwise $|q_n| \leq |q_{n-1}|$. Then from Lemma 3.4, we know that $a_n = \pm 1$. Suppose $a_n = 1$ (the case $a_n = -1$ is analogous). Then also from Lemma 3.4, we know that

$\operatorname{sgn} q_{n-1} = \operatorname{sgn} q_n$. Also from Lemma 3.3, we know that $x_{n+1} > 1$ and so $\xi > 0$. Thus $|q_n + \xi q_{n-1}| > |q_n|$. It follows that $|q_n(q_n + \xi q_{n-1})| \geq |q_n|$, which completes the proof. □

Theorem 3.8. *If* $\mathrm{CF}_f(x) = (a_0, a_1, \ldots)$ *then* $\lim_{n\to\infty}[a_0, a_1, \ldots, a_n]$ *exists and equals x.*

Proof. Suppose $\mathrm{CF}_f(x) = (a_0, a_1, \ldots)$. Consider $p_n/q_n = [a_0, \ldots, a_n]$. From Theorem 3.7, we know that $|x - \frac{p_n}{q_n}| < |\frac{1}{q_n}|$. From Lemma 3.3 we know that the partial quotients fulfill the rules corresponding to a semiregular continued fraction, and hence from Theorem 2.1 we know that the sequence $[a_0, \ldots, a_n]$ converges to some limit α. Since from Theorem 2.1 we also know that $|q_n| \to +\infty$, it follows that $x = \alpha$. □

The main result of this paper is that the outputs of CF_f are easily describable in most of the interesting cases, including the examples SCF, CCF, and NICF mentioned previously. Let us define more rigorously what we mean by "easily describable."

Call a finite sequence of integers *valid* if it is the result of $\mathrm{CF}_f(x)$ for some rational number x. We envision a deterministic finite automaton which reads a purported finite expansion $\mathbf{a} = (a_0, a_1, \ldots a_n)$ and reaches a final state on the last input iff $\mathbf{a}$ is valid.

Definition 3.9. A finite automaton is a 5-tuple $(Q, \Sigma, \delta, q_0, F)$ where Q is a finite set of states, Σ is an (not necessarily finite) input alphabet, $q_0 \in Q$ is the initial state, $F \subseteq Q$ is the set of final states, and δ is the transition function mapping $Q \times \Sigma$ to Q. The transition function δ may be a partial function; i.e., $\delta(q, a)$ may be undefined for some pairs q, a.

We extend δ to a function which maps $Q \times \Sigma^*$ to Q in the obvious fashion.

The reader to whom these definitions are unfamiliar should consult [6].

The acceptance criterion for infinite expansions clearly needs to be different, since in this case there is no "last" partial quotient. We address the case of infinite expansions in Sect. 6.

One minor problem with this model is that the a_i belong to $\mathbb{Z}$, but in defining finite automata we usually insist that our alphabet Σ be finite. We can get around this in one of two ways: first, we could expand the definition of finite automata so that there can be infinitely many transitions (but still only finitely many states). However, such a model is too arbitrary, since allowing infinitely many transitions allows us to accept a set of expansions that is not even recursively enumerable. It suffices to allow only *finitely many* transitions, where each transition must be

- Either a single integer, or
- A set of the form $\{x \in \mathbb{Z} : x \geq \alpha\}$, or
- A set of the form $\{x \in \mathbb{Z} : x \leq \alpha\}$.

As an alternative, we could redefine our strings as numbers encoded in a particular base. That this is equivalent is clear, since the base-k representation of sets like $\{x \in \mathbb{Z} \ : \ x \geq \alpha\}$ forms a regular language. So either approach is satisfactory, but for simplicity we choose the first.

Notation. If $A \subseteq \mathbb{R}$ is a set, then by A^{-1} we mean the set of reciprocals $\{x \in \mathbb{R} - \{0\} : x^{-1} \in A\}$. Thus, for example, $[-\frac{1}{2}, \frac{1}{2}]^{-1} = (-\infty, -2] \cup [2, \infty)$ and $[1, \infty)^{-1} = (0, 1]$. If f is a function, then by $f^{-1}[a]$ we mean, as usual, the set $\{x \in \mathbb{R} \ : \ f(x) = a\}$. If A is a set, then by $A - a$ we mean the set $\{x \ : \ x + a \in A\}$. We will say x is *quadratic* if x is the real root of a quadratic equation with integer coefficients.

Definition 3.10. Let f be a real integer function. Then we say that the finite automaton $A = (Q, \mathbb{Z}, \delta, q_0, F)$ *accepts the outputs* of the algorithm CF_f if $\delta(q_0, a_0 a_1 a_2 \cdots a_n) \in F$ iff there exists $x \in \mathbb{Q}$ such that $\mathrm{CF}_f(x) = (a_0, a_1, \ldots, a_n)$.

The object of this paper is to prove the following theorem:

Theorem 3.11. *Let f be an integer function and suppose $f^{-1}[0]$ is the finite union of intervals. Then there exists a finite automaton accepting the outputs of* CF_f *iff all the endpoints of the intervals are rational or quadratic.*

In Sect. 4, we will prove one direction of this theorem; in Sect. 5, we prove the other.

Now we give these automata for the three continued fraction algorithms discussed so far: SCF, CCF, and NICF (Figs. 1–3).

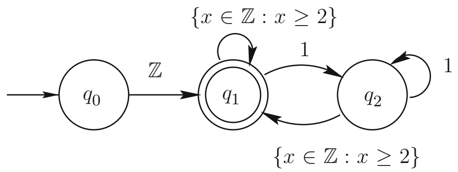

Fig. 1 Automaton for the simple continued fraction algorithm SCF

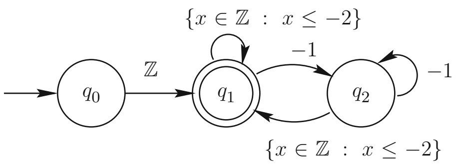

Fig. 2 Automaton for the ceiling algorithm CCF

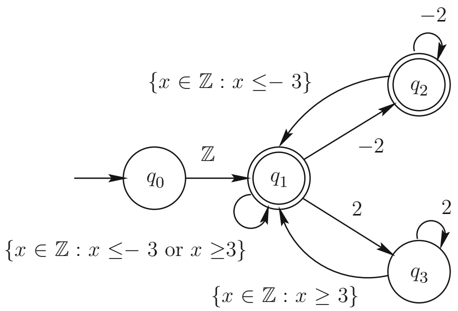

Fig. 3 Automaton for the nearest integer continued fraction algorithm NICF

Remark 12. No simple characterization seems to exist in the case where f is not the finite union of intervals. In Sect. 5, we will give an example of an f that is accepted by a finite automaton, but $f^{-1}[0]$ is not the finite union of intervals.

4 One Direction of the Theorem

We first prove the following theorem:

Theorem 4.12. *Suppose $f^{-1}[0]$ is the finite union of intervals with rational or quadratic endpoints. Consider all possible sets $S_0, S_1, \ldots,$ constructed in Theorem 3.5, corresponding to all possible input sequences. Then among these there are only finitely many distinct sets.*

Proof. If S is a finite union of intervals, then $S^{-1} - a$ is also such a union. Hence $T = (S^{-1} - a) \cap f^{-1}[0]$ is such a union. The endpoints e of intervals of T are either those of $f^{-1}[0]$, or $e = 1/d - f(1/d)$, where d is an endpoint of S. Since $f(1/d)$ equals either $\lfloor 1/d \rfloor$ or $\lceil 1/d \rceil$, it will suffice to prove the following:

Lemma 4.14. *Define $s_1 : x \to (1/x) - \lfloor 1/x \rfloor$ and $s_2 : x \to (1/x) - \lceil 1/x \rceil$. Consider the monoid u formed by the maps s_1 and s_2 under composition. Let $u(x)$ be the orbit of x under elements of u.*

Then $u(x)$ is finite iff x is rational or quadratic.

Proof. One direction is easy. Assume $u(x)$ is finite. Then in particular the set

$$x, s_1(x), s_1^{(2)}(x), \ldots$$

is finite. Hence we have $s_1^{(j)}(x) = s_1^{(k)}(x)$ for some $j \neq k$. But it is easily proved by induction that

$$x = [0, a_0, a_1, \ldots, a_{n-1} + s_1^{(n)}(x)]$$

for some sequence of integers $a_0, a_1, \ldots$; hence there exist integers such that

$$x = \frac{a_j + b_j s_1^{(j)}(x)}{c_j + d_j s_1^{(j)}(x)},$$

and similarly

$$x = \frac{a_k + b_k s_1^{(k)}(x)}{c_k + d_k s_1^{(k)}(x)}.$$

Thus we see that $s_1^{(j)}(x)$ is the root of a quadratic equation, and so is either quadratic or rational. Thus x itself is either quadratic or rational.

If x is rational, the other direction follows easily, using an argument exactly like that in the proof of Theorem 3.2. If x is the root of a quadratic equation with integer coefficients, the result follows immediately from an old theorem of Blumer [3, Satz IX, p. 50]; for a more recent proof, see [16, Theorem A, p. 225]. □

This completes the proof of Theorem 4.12. □

We can now prove one direction of Theorem 3.11. Given an integer function f such that $f^{-1}[0]$ is the finite union of intervals with rational or quadratic endpoints, we create a finite automaton A_f as follows: the states of A_f are the distinct sets $S_f(n)$ constructed in Theorem 3.5 for all possible real numbers, together with a start state q_0. For convenience, we rename the states as $q_0, q_1, \ldots, q_t$ for some $t \geq 1$. From our results above, we know that the number of states is finite. We define $\delta(q_0, a_0) = q_1 = f^{-1}[0]$ for all $a_0 \in \mathbb{Z}$ and inductively define

$$\delta(q_i, a) = q_j$$

where $q_j = (q_i^{-1} \cap f^{-1}[a]) - a$, provided this set is nonempty. We say $q_i \in F$ if $0 \in q_i$.

It remains to verify that (a) the automaton accepts CF_f; and (b) the transitions can be characterized finitely, as discussed previously.

Corollary 4.15. *$\delta(q_0, a_0 a_1 \cdots a_n) \in F$ iff there exists $x \in \mathbb{Q}$ such that $\mathrm{CF}_f(x) = (a_0, a_1, \ldots, a_n)$.*

Proof. Assume $\delta(q_0, a_0 a_1 \cdots a_n) \in F$. Then by the definition of the set of final states F, we must have $0 \in \delta(q_0, a_0 a_1 \cdots a_n)$. By Theorem 3.2 the first $n+1$ outputs of the algorithm CF_f on input $[a_0, a_1, \ldots, a_n]$ are precisely $(a_0, a_1, \ldots, a_n)$. Hence we may take $x = [a_0, a_1, \ldots, a_n]$.

Now assume that there exists $x \in \mathbb{Q}$ such that $\mathrm{CF}_f(x) = (a_0, a_1, \ldots, a_n)$. Then from the definition of CF_f, we see that $x_n = a_n$; hence

$$0 = x_n - a_n \in \delta(q_0, a_0 a_1 \cdots a_n)$$

which shows that $\delta(q_0, a_0 a_1 \cdots a_n)$ is a final state. □

The final step is to characterize the transitions. If the transition comes from q_0, then it is labeled $\mathbb{Z}$, which we can write, for example, $\{x \,:\, x \geq 0\}$ together with $\{x \,:\, x < 0\}$. Otherwise consider a transition of the form $\delta(q_i, a) = q_j$ where

$$q_j = (q_i^{-1} \cap f^{-1}[a]) - a = (q_i^{-1} - a) \cap f^{-1}[0].$$

If q_i is the finite union of intervals, then so is q_i^{-1} and $q_i^{-1} - a$. If q_i^{-1} is bounded, then as a ranges over all integers, there are only finitely many nonempty intersections of $(q_i^{-1} - a)$ with $f^{-1}[0]$. If q_i^{-1} is unbounded, say on the positive axis, then the intersection of $(q_i^{-1} - a)$ with $f^{-1}[0]$ is the same for all sufficiently large a. The same result holds when q_i^{-1} is unbounded on the negative axis. Hence there exists α, β such that the transition on each $x \geq \alpha$ is the same, and the transition on each $x \leq \beta$ is the same.

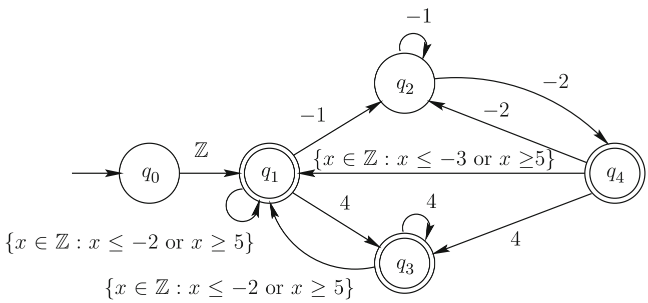

Fig. 4 Automaton corresponding to $f^{-1}[0] = [-\frac{\sqrt{2}}{2}, \frac{2-\sqrt{2}}{2})$

Combining this observation with Lemma 4.14 completes the proof of one direction of Theorem 3.11. □

We now give some examples of the construction of the finite automaton.

Example 4.16. Let us obtain the description of the outputs for CF_f for $f(x) = \lfloor x + \frac{\sqrt{2}}{2} \rfloor$. We find

$$q_1 = f^{-1}[0] = [-\frac{\sqrt{2}}{2}, \frac{2-\sqrt{2}}{2})$$

$$q_2 = [-\frac{\sqrt{2}}{2}, 1-\sqrt{2}]$$

$$q_3 = (\sqrt{2}-2, \frac{2-\sqrt{2}}{2})$$

$$q_4 = [1-\sqrt{2}, \frac{2-\sqrt{2}}{2}).$$

The full automaton is given below in Fig. 4.

Example 4.17. Our next example corresponds to the integer function defined by $f^{-1}[0] = (-1, -\frac{1}{2}] \cup \{0\} \cup (\frac{1}{2}, 1)$. We have

$$q_1 = f^{-1}[0]$$

$$q_2 = \{0\} \cup (\frac{1}{2}, 1)$$

$$q_3 = (-1, -\frac{1}{2}]$$

$$q_4 = (\frac{1}{2}, 1).$$

This gives the automaton below (Fig. 5).

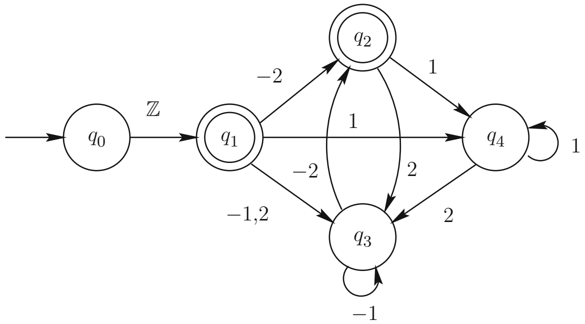

Fig. 5 Automaton generating bounded partial quotients

These expansions were introduced by Lehner [11] and further studied by Dajani and Kraaikamp [5]. An interesting feature of this expansion is that the partial quotients all lie in the set $\{-2,-1,1,2\}$. For example, the expansion of $\frac{52}{43}$ is $(2,-2,2,-1,-1,-2,2,-1,-1,-2)$. One undesirable aspect of these expansions is their slow convergence; the expansion of of $\frac{1}{2n}$ is $(1,\overbrace{-2,2,-2,2,\ldots}^{2n-1})$. Another undesirable aspect is that there exist infinite paths through the automaton (such as $(1,-2,2,-2,2,-2,2,\ldots)$) which do not correspond to the expansion of any real x. However, this is essentially the only problematic case, as we will see below in Theorem 6.14.

5 Completing the Proof of Theorem 3.11

Proof. We now wish to show that if $f^{-1}[0]$ consists of the finite union of intervals, but one of those intervals has an endpoint that is not rational or quadratic, then no finite automaton can accept CF_f.

Assume that such an automaton A exists. Then we may assume that each state is in fact reachable from q_0; otherwise this state may be discarded without affecting A. For each state q_j, construct an input sequence $a_0a_1\cdots a_i$ such that $\delta(q_0,a_0a_1\cdots a_i)=q_j$. Let us label each state q_j with a subset of $\mathbb{Q}$, $L(q_j)$, by the following rule: If $\delta(q_0,a_0a_1\cdots a_i)=q_j$, then

$$L(q_j)=\{x\in\mathbb{Q}\ :\ \mathrm{pref}_i(\mathrm{CF}_f([a_0,a_1,\ldots,a_{i-1},a_i+x]))=(a_0,a_1,\ldots,a_i)\}.$$

We need to show that this map is indeed well defined, in the sense that different paths from q_0 to q_j give the same labels $L(q_j)$. Assume that

$$\delta(q_0,a_0a_1\cdots a_i)=q_j$$

and

$$\delta(q_0, b_0 b_1 \cdots b_k) = q_j,$$

and there exists a rational number p such that

$$p \in S_1 = \{x \in \mathbb{Q} \,:\, \mathrm{CF}_f([a_0, a_1, \ldots, a_{i-1}, a_i + x]) = (a_0, a_1, \ldots, a_i, \ldots)\}$$

but

$$p \notin S_2 = \{x \in \mathbb{Q} \,:\, \mathrm{CF}_f([b_0, b_1, \ldots, b_{k-1}, b_k + x]) = (b_0, b_1, \ldots b_k, \ldots)\}.$$

Write $\mathrm{CF}_f(p) = (0, a_{i+1}, \ldots, a_n)$; by our definition of what it means to accept the output of CF_f, we know that

$$\delta(q_j, a_{i+1} \cdots a_n) = q_r \in F,$$

a final state. Let $y = [b_0, b_1, \ldots, b_k, a_{i+1}, \ldots a_n]$. Then since the automaton is in state q_j upon reading inputs $b_0 b_1 \cdots b_k$, we have

$$\delta(q_0, b_0 b_1 \cdots b_k a_{i+1} \cdots a_n) = q_r.$$

Hence $\mathrm{CF}_f(y) = (b_0, b_1, \ldots, b_k, a_{i+1}, \ldots a_n)$. But then $y = [b_0, b_1, \ldots, b_k + p]$ which shows that indeed $p \in S_2$, a contradiction.

Thus we may assume that sets $L_i = L(q_i)$ are well defined. Let $\bar{A}$ denote the *closure* of the set A in $\mathbb{R}$, and consider the sets $\bar{L_i}$. I claim that since $f^{-1}[0]$ consists of the finite union of intervals, so does each of the sets $\bar{L_i}$; this follows easily from the definition of CF_f. Suppose $\delta(q_i, a) = q_j$; then the endpoints e of intervals of $\bar{L}_j$ are those of $f^{-1}[0]$ or are related to the endpoints E of $\bar{L}_i$ by the equation

$$e = \frac{1}{E} - a.$$

Since $f^{-1}[0]$ contains an endpoint which is not rational or quadratic, so must $\bar{L}_0$. Hence there exists a transition $\delta(q_0, a) = q_i$ such that $\bar{L}_i$ contains an endpoint which is not rational or quadratic. Continuing in this fashion, and remembering that there are only a finite number of states, we eventually return to a state previously visited, which gives one of the two equations

$$e = [0, a_1, \ldots, a_k]$$

or

$$e = [0, a_1, \ldots, a_k + e]$$

which shows that e is rational or quadratic, contrary to assumption.

This completes the proof of Theorem 3.11. □

Now let us give an example of an f such that $f^{-1}[0]$ is not the finite union of intervals, but nevertheless there is a finite automaton accepting CF_f.

Let $f(x)$ be defined by

$$f(x) = \begin{cases} \lfloor x \rfloor, & \text{if } x \text{ is rational;} \\ \lceil x \rceil, & \text{if } x \text{ is irrational.} \end{cases}$$

Then

$$f^{-1}[0] = \{x \,:\, x \text{ rational},\ 0 \le x < 1\} \cup \{x \,:\, x \text{ irrational},\ -1 < x < 0\}.$$

Clearly $f^{-1}[0]$ cannot be written as the finite union of intervals. Then it is easily verified that the procedure of Sect. 4 generates a finite automaton with four states that accepts CF_f.

It may be of interest to remark that the automata accepting the result of CF_f may be arbitrarily complex. For example, it can be shown that the automaton corresponding to

$$f^{-1}[0] = [-\frac{F_{n-1}}{F_n}, \frac{F_{n-2}}{F_n})$$

has $n+1$ states. (Here F_n denotes the nth Fibonacci number.)

6 Infinite Expansions

So far we have just addressed the case of finite expansions, the ones arising from rational number. In this section we handle irrational numbers. We say that an infinite sequence $(a_0, a_1, \ldots)$ is valid for an integer function f if it is the output of the algorithm CF_f on some input x.

In the remainder of this section we assume that f is an integer function such that $f^{-1}[0]$ consists of a finite union of intervals with rational or quadratic endpoints. We construct the associated automaton A_f as in Sect. 3. We would expect that infinite paths through A_f correspond in a 1-1 fashion with outputs of CF_f. However, this is not quite true; there can be certain infinite paths that do not correspond to any output of CF_f. By ruling these out, we can get the correspondence we desire.

Theorem 6.13. *If* $\mathrm{CF}_f(x) = \mathbf{a} = (a_0, a_1, \ldots)$ *and* $A_f = (Q, \Sigma, \delta, q_0, F)$, *then* $\delta(q_0, a_0 a_1 \cdots a_n)$ *exists for all* $n \ge 0$.

Proof. By induction on n. Clearly this is true for $n = 0$, since there is a transition labeled with each $a \in \mathbb{Z}$ leaving q_0 to q_1. Now assume the claim is true for all $n' < n$; we prove it for n. Now $\mathrm{pref}_n(\mathrm{CF}_f(x)) = (a_0, a_1, \ldots, a_n)$ if and only if $\mathrm{pref}_{n-1}(\mathrm{CF}_f(x)) = (a_0, \ldots, a_{n-1})$ and $x_n = (x_{n-1} - a_{n-1})^{-1}$ and $a_n = f(x_n)$. Let $q = \delta(q_0, a_0 \ldots a_{n-1})$; then by definition we have a transition out of q labeled a_n if and only if the set $q^{-1} \cap f^{-1}[a_n]$ is nonempty. But $x_n \in q^{-1} \cap f^{-1}[a_n]$, so there is indeed such a transition. □

Finally, we characterize the infinite paths through A_f that correspond to the expansion associated with some irrational number.

Theorem 6.14. *Let* $\mathbf{a} = (a_0, a_1, \ldots)$ *be an infinite path through* A_f *that has no infinite suffix of the form* $(2, -2, 2, -2, \ldots)$*. Then there is an irrational real number* x *such that* $\mathrm{CF}_f(x) = \mathbf{a}$.

Proof. Take an infinite path through A_f labeled with $(a_0, a_1, \ldots)$. For each finite prefix $(a_0, a_1, \ldots, a_n)$, we can consider the continued fraction $[a_0, \ldots, a_n]$ and the corresponding convergents p_n, q_n. Then $\mathrm{CF}_f(p_n/q_n) = (a_0, a_1, \ldots, a_n)$. Furthermore, the p_n/q_n converge to $x = [a_0, a_1, \ldots]$ and $|x - p_n/q_n| < |1/q_n|$.

From our correspondence with semiregular continued fractions given in Sect. 2, the continued fraction $[a_0, a_1, \ldots]$ corresponds to a certain semiregular continued fraction, with pattern of signs in the numerator dependent on the pattern of signs of the a_i, as given in the equivalence between (1) and (2). However, from Theorem 2.1, for a given pattern of signs, an infinite semiregular continued fraction expansion is unique, provided it obeys the rule that $a_i + \varepsilon_{i+1} \geq 2$ infinitely often. We have already seen that $a_i + \varepsilon_{i+1} \geq 1$ for our expansions. The case where $a_i + \varepsilon_{i+1} = 1$ for all but finitely many i corresponds to (in our notation) an expansion that looks like $[\ldots, 2, -2, 2, -2, \ldots]$ with an infinite suffix of $(2, -2)$ repeating. However, by hypothesis, our path A_f has no such suffix. Therefore, it corresponds to a unique real number x. □

7 Variations

Exactly the same result holds for minor variations of our continued fraction algorithm. For example, suppose we have some finite list of integer functions $f_0, f_1, \ldots, f_{n-1}$ and apply them periodically, as follows:

Algorithm PSCF(x)**; outputs** $(a_0, a_1, \ldots)$**:**
PSCF1. Set $x_0 \leftarrow x$; set $i \leftarrow 0$.
PSCF2. Set $a_i \leftarrow f_{i \bmod n}(x)$.
PSCF3. If $a_i = x_i$ then stop. Otherwise set $x_{i+1} \leftarrow 1/(x_i - a_i)$; set $i \leftarrow i + 1$ and go to step PSCF2.

Then the analogous version of Theorem 3.11 holds. The only difference is that the automaton needs to keep track of the current value of i, taken modulo n. For example, suppose we let $n = 2$ and $f_0(x) = \lceil x \rceil$ and $f_1(x) = \lfloor x \rfloor$. The resulting algorithm gives what is often called the *reduced simple continued fraction expansion* in the literature. The corresponding automaton is given below (Fig. 6).

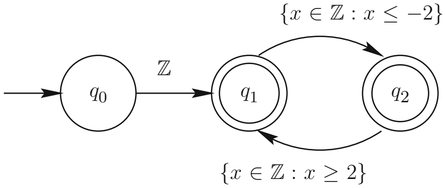

Fig. 6 Automaton for the reduced simple continued fraction algorithm

Because the signs of terms alternate, these continued fractions are often written in the form

$$a_0 - \cfrac{1}{a_1 - \cfrac{1}{a_2 - \cdots}},$$

where $a_i \in \mathbb{Z}$ and $a_i \geq 2$ for $i \geq 1$.

Another variation is to treat positive and negative numbers differently. For example, we could define

$$f(x) = \begin{cases} \lfloor x + \frac{1}{2} \rfloor, & \text{if } x \geq 0; \\ \lceil x - \frac{1}{2} \rceil, & \text{if } x < 0. \end{cases}$$

Our results, with small differences, also apply here.

8 Concluding Remarks

Our results apply to, for example, the α-continued fractions of Tanaka and Ito [17], which correspond to the integer function $f(x) = \lfloor x - \alpha + 1 \rfloor$, where $\frac{1}{2} \leq \alpha \leq 1$ is a real number.

Several other writers have noted connections between finite automata and continued fractions. One of the best-known papers is that of Raney, who showed how to obtain the simple continued fraction for

$$\beta = \frac{a\alpha + b}{c\alpha + d}$$

in terms of the continued fraction for α. See [2, 14] for more details.

Istrail considered the language consisting of all prefixes of the continued fraction for x, and observed that this language is context-free and non-regular iff x is a quadratic irrational [9].

Allouche discusses several applications of finite automata to number theory, including continued fractions [1].

In this paper, we have been concerned with a different approach; namely, describing the "set of rules" associated with a generalized continued fraction algorithm. One immediately wonders if similar theorems may be obtained for continued fraction algorithms in $\mathbb{Z}[i]$, such as those discussed by Hurwitz [7] and McDonnell [12].

In [15] the author proved that the McDonnell's complex continued fraction algorithm can be described by a finite automaton with 25 states. The corresponding result for Hurwitz's algorithm is not known.

Acknowledgements I am very grateful to the referee for many suggestions that considerably improved the chapter.

Supported by a grant from NSERC.

References

1. J.-P. Allouche, Automates finis en théorie des nombres. Exposition Math. **5**, 239–266 (1987)
2. W.M. Beynon, A formal account of some elementary continued fraction algorithms. J. Algorithms **4**, 221–240 (1983)
3. F. Blumer, Über die verschiedenen Kettenbruchentwicklungen beliebiger reeller Zahlen und die periodischen Kettenbruchentwicklungen quadratischer Irrationalitäten. Acta Arith. **3**, 3–63 (1939)
4. C. Brezinski, *History of Continued Fractions and Padé Approximants* (Springer, New York, 1991)
5. K. Dajani, C. Kraaikamp, The mother of all continued fractions. Colloq. Math. **84/85**(Pt 1), 109–123 (2000)
6. J.E. Hopcroft, J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, Boston, 1979)
7. A. Hurwitz, Über die Entwicklung complexer Grössen in Kettenbrüche. Acta Math. **11**, 187–200 (1888) [= Werke, II, pp. 72–83]
8. A. Hurwitz, Über eine besondere Art der Kettenbruch-Entwicklung reeller Grössen. Acta Math. **12**, 367–405 (1889) [= Werke, II, pp. 84–115]
9. S. Istrail, On formal construction of algebraic numbers of degree two. Rev. Roum. Math. Pures Appl. **22**, 1235–1239 (1977)
10. D.E. Knuth, *The Art of Computer Programming, V. II (Seminumerical Algorithms)*, 2nd edn (Addison-Wesley, Boston, 1981)
11. J. Lehner, Semiregular continued fractions whose partial denominators are 1 or 2, in *The Mathematical Legacy of Wilhelm Magnus: Groups, Geometry, and Special Functions*, ed. by W. Abikoff, J.S. Birman, K. Kuiken (Amer. Math. Soc., New York, 1994), pp. 407–410
12. E.E. McDonnell, Integer functions of complex numbers, with applications. IBM Philadelphia Scientific Center, Technical Report 320-3015, February 1973
13. O. Perron, Die Lehre von den Kettenbrüchen, in *Band I: Elementare Kettenbrüche* (Teubner, 1977)
14. G.N. Raney, On continued fractions and finite automata. Math. Ann. **206**, 265–283 (1973)
15. J.O. Shallit, Integer functions and continued fractions. A.B. Thesis, Princeton University, 1979. Available at http://www.cs.uwaterloo.ca/~shallit/papers.html
16. J. Steinig, On the complete quotients of semi-regular continued fractions for quadratic irrationals. Arch. Math. **43**, 224–228 (1984)
17. S. Tanaka, S. Ito, On a family of continued-fraction transformations and their ergodic properties. Tokyo J. Math. **4**, 153–175 (1981)
18. H. Tietze, Über Kriterien für Konvergenz und Irrationalität unendlichen Kettenbrüche. Math. Annalen **70**, 236–265 (1911)