HOFSTADTER'S EXTRACTION CONJECTURE

Russell Jay Hendel*
P.O. Box 1533, GMF, Boston, MA 02205

Sandra A. Monteferrante
Department of Mathematics, Dowling College, Oakdale, NY 11769
(Submitted February 1991)

Let α, $0<\alpha<1$, be irrational. For integer $n>0$, define $f(n)=[(n+1)\alpha]-[n\alpha]$. Define $g(n)=c$ if $f(n)=0$ and $g(n)=d$ if $f(n)=1$ and let $x=x(\alpha)$ be the infinite string whose n^{th} element is $g(n)$.

Both the string x, and the three functions $f(n), g(n)$, and $[n\alpha]$ have been studied extensively. Classically, an astronomical problem of Bernoulli led Markov to prove results about the structure of x. A concise summary is presented in [12]. The results use continued fraction methods and the theory of semigroups.

Recent research connects x with monoid homomorphisms (e.g., Fraenkel et al. [4]), outputs of automata (e.g., Shallit [9]), and general properties of strings (e.g., Mignosi [8]).

These functions and their related sequences have obvious recreational and experimental flavor, are noted for their exotic functional patterns (e.g., Doster [1]), and lend themselves readily to computer experiments (e.g., Engel [2], or Hofstadter [7]).

In this paper we study a problem first described by Hofstadter in an unpublished manuscript [6]:

> But now I would like to give an example par excellence of horizontal properties, a property which I call "extraction." The idea is this. To begin with, write down x. Now choose some arbitrary term in it, called the "starting point." Beginning at the starting point, try to match x term by term. Every time you find a match, circle that term. Soon you will come to a term which differs from x. When this happens, just skip over it without circling it, and look for the earliest match to the term of x you are seeking. Continue this process indefinitely. In the end you have circled a great number of terms after the starting point, and left some uncircled. We are interested in the uncircled terms, which are now "extracted" from x. The interesting fact is that the extracted sequence is the subsequence of x which begins two terms earlier than the starting point! To decrease confusion, I now show an example, where instead of circling I underline the terms which match x. In this example, $\alpha=(\sqrt{5}-1)/2$.
>
> I have chosen this "d" as the starting point:
>
> |
> |
>
> dcd dcdcddcddcdcddcdcddcddcdcddcddcdcddcdcddcddcdcddcd···
>
> The underlined sequence matches the full sequence, x, term by term. Now what is the extracted sequence? It is:
>
> $c\,d\,d\,c\,d\,c\,d\,d\,c\,d\,d\,cdc\,d\,d\,c\,d\cdots$
>
> And you will find that this matches with the sequence which begins two places earlier than the starting point. Carrying it further is tedious, and does nothing but confirm our observation. Why does this extraction-property hold? At this point, I must admit that I don't know. It is a curious property which needs further investigation.

---

* Written while Hendel was at Dowling College, which partially supported this project.

To rigorously formulate this, we present the following definition.

Definition 1: Suppose $U = u_1 \ldots u_n$, $V = v_1 \ldots v_m$, and $E = e_1 \ldots e_p$ with $u_i, v_j, e_k \in \{c, d\}$, $n, m > 0$, and $n = m + p$. We say U *aligns* (with) V with *extraction* E (notationally indicated by $U \supset V; E$), if there exist integers $j(0), j(1), j(2), \ldots, j(p)$, such that

$$U = \{v_1 \ldots v_{j(1)}\} e_1 \{v_{j(1)+1} \ldots v_{j(2)}\} e_2 \ldots e_p \{v_{j(p)+1} \ldots v_m\} \quad (\text{with } \{v_a \ldots v_b\} \text{ empty if } b < a),$$

where

(i) $0 = j(0) \le j(1) \le j(2) \le \cdots \le j(p) < m$,

(ii) $e_i \ne v_{j(i)+1}$, for $1 \le i \le p$.

For example, if $p = 0$, $U \supset V; E$ with $U = V$ and E the empty string. Throughout this paper we use the nonstandard symbol ϕ to denote the empty string. It is easy to see that $U \supset V; \phi$ if and only if $U = V$.

If $U \supset V; E$, then U, V, and E are call the *original*, *aligned*, and *extracted* strings, respectively, and the relationship itself is called an *alignment*.

Remark: Define strings $U = dcdcd$ and $V = dd$. To clarify some subtleties in Definition 1, we explore the consequences of dropping requirements (i) or (ii).

If we drop the requirement of strict inequality, $j(p) < m$, in Definition 1(i), then we allow $U \supset V; ccd$ with $j(1) = 1$, $j(2) = j(3) = m = 2$.

If we keep requirement (i) but drop requirement (ii), then we allow $U \supset V; cdc$, with $j(1) = j(2) = j(3) = 1$, $m = 2$, $e_2 = v_{j(2)+1}$ and, similarly, we allow $U \supset V; dcc$, with $j(1) = j(2) = 0$, $j(3) = 1$, $m = 3$, $e_1 = v_{j(1)+1}$.

Thus, for given original and aligned strings, without requirements (i) and (ii), the extracted string is not necessarily unique. However, with requirements (i) and (ii), we can prove the following lemma.

Lemma 1: For given strings U and V, there is at most one string E such that $U \supset V; E$.

Proof: We suppose $U \supset V; E$, $U \supset V; E'$, and $E \ne E'$ and derive a contradiction.

By Definition 1, there are sequences $j(1), \ldots, j(p)$, and $j'(1), \ldots, j'(p)$ satisfying (i) and (ii) of Definition 1 and

$$U = \{v_1 \ldots v_{j(1)}\} e_1 \{v_{j(1)+1} \ldots v_{j(2)}\} e_2 \ldots e_p \{v_{j(p)+1} \ldots v_m\}, \tag{$*$}$$

$$U = \{v_1 \ldots v_{j'(1)}\} e_1' \{v_{j'(1)+1} \ldots v_{j'(2)}\} e_2' \ldots e_p' \{v_{j'(p)+1} \ldots v_m\}. \tag{$**$}$$

Observe that, for $1 \le r \le p$, $e_r =$ the $\{j(r) + r\}^{\text{th}}$ element of U. Similarly, if t is given such that either $j(r) + r < t < j(r+1) + (r+1)$ for some r, $0 \le r \le p-1$, or $j(r) + r < t \le m$ with $r = p$, then $v_{t-r} =$ the t^{th} element of U.

Let s be the largest integer such that $j(r) = j'(r)$ for $0 \le r < s$. Then s exists and is positive because $j(0) = 0 = j'(0)$. Since we assume $E \ne E'$, $s \le p$.

If we further suppose that $j(s) < j'(s)$, then $j'(s-1) + (s-1) < j(s) + s < j'(s) + s$.

Therefore, by considering $(*)$ and $(**)$, respectively, the $\{j(s)+s\}^{\text{st}}$ element of U is, simultaneously, e_s and $v_{j(s)+1}$, contradicting Definition 1(ii). A similar argument holds if $j'(s) < j(s)$. These contradictions show that $E = E'$ and complete the proof.

Recall that u is a *prefix* (that is, *left factor*) of v if there is a string y such that $v = uy$. Similarly, u is a *suffix* (or *right factor*) of v, if $v = yu$ for some string y. We say that the string y is the *limit* of the sequence of strings $y(n)$, $n = 1, 2, 3, \ldots$, notationally indicated by $y = \lim y(n)$, if, for each positive integer m less than or equal to the length of y, the left factors of length m of $y(n)$ and y are equal for all sufficiently large n.

Definition 2: Suppose U, V, and E are (possibly infinite) strings. Suppose $U(n), V(n)$, and $E(n)$, $n \geq 1$, are sequences of finite strings such that $U(n) \supset V(n); E(n)$, with $\lim U(n) = U$, $\lim V(n) = V$, and $\lim E(n) = E$. Then we say U *aligns* V with *extraction* E and indicate this, notationally, by $U \supset V; E$ (we do not require E to be infinite).

Remark: By a proof similar to that of Lemma 1, it can be proved in the infinite case also that E is (uniquely) functionally dependent on U and V.

Let x_m denote x with the left factor of length m deleted. We can now formulate the general Hofstadter conjecture as follows:

Hofstadter's Conjecture: For any α and any $m \geq 2$

$$x_m \supset x; x_{m-2}. \tag{1}$$

Example 1: For the remainder of this paper we assume $\alpha = (\sqrt{5}-1)/2$. In this case, the sequence

$$x = dcd\ dcd\ cd\ dcd\ dc\ dcd\ dcd\ cd\ dcd\ dcd\ cd\ dcd\ dcd\ cd\ dcd\ cd\ dcd\ dcd\ cd\ dcd \cdots$$

has been described fairly thoroughly in the literature (see Tognetti et al. [11]). The sequence is referred to as the Golden sequence or, sometimes, the Fibonacci sequence. With

$$x_1 = cd\ dcd\ cddcd\ dcdcddcd\ cddcddcdcddcd \cdots$$
$$x_3 = dcd\ cddcd\ dcdcddcd\ cddcddcdcddcd \cdots,$$

Hofstadter's conjecture for $m = 3$ asserts $x_3 \supset x; x_1$.

We define $c_0 = c$, $c_1 = d$,

$$c_n = c_{n-2}c_{n-1}, \quad n \geq 2. \tag{2}$$

Then $c_2 = cd$, $c_3 = dcd$, $c_4 = cddcd$, $c_5 = dcd\ cddcd$, and $c_6 = cddcd\ dcd\ cddcd$.

The following result is well known [12].

Lemma 2: $x = c_1c_2 \ldots$.

A crucial component of the proof of Hofstadter's conjecture is a *concatenation lemma* asserting that under approprite conditions the extractions of concatenated strings are the concatenations of their extractions.

Lemma 3:

(i) Let U, V, E and U', V', E' denote arbitrary strings of finite length. If $U \supset V; E$ and $U' \supset V'; E'$, then $UU' \supset VV'; EE'$.

(ii) If U_i, V_i, and E_i, $1 \le i \le m$, are arbitrary strings of finite lengths with m some integer, and if $U_i \supset V_i; E_i$, $1 \le i \le m$, then $\prod U_i \supset \prod V_i; \prod E_i$ (with products denoting concatenation).

Proof: Part (ii) follows from part (i) by simple induction. To prove (i), we suppose, using Definition 1, that

$$U = \{v_1 \dots v_{j(1)}\} e_1 \{v_{j(1)+1} \dots v_{j(2)}\} e_2 \dots e_p \{v_{j(p)+1} \dots v_m\},$$
$$U' = \{v_1' \dots v'_{j'(1)}\} e_1' \{v'_{j'(1)+1} \dots v'_{j'(2)}\} e_2' \dots e'_{p'} \{v'_{j'(p')+1} \dots v'_{m'}\},$$

for some sequences of integers, $0 \le j(1) \le \cdots \le j(p) < m$, $0 \le j'(1) \le \cdots \le j'(p') < m'$ with $V = v_1 \dots v_m$, $V' = v_1' \dots v'_{m'}$, $E = e_1 \dots e_p$, and $E' = e_1' \dots e'_{p'}$. Then

$$UU' = \{v_1 \dots v_{j(1)}\} e_1 \{v_{j(1)+1} \dots v_{j(2)}\} e_2 \dots e_p \{v_{j(p)+1} \dots v_m\} \{v_1' \dots v'_{j'(1)}\} e_1' \dots e'_{p'} \{v'_{j'(p')+1} \dots v'_{m'}\}.$$

To prove $UU' \supset VV'; EE'$, we verify that requirements (i) and (ii) of Definition 1 are satisfied by the sequence of integers $0 \le j(1) \le j(2) \le \cdots \le j(p) < m + j'(1) \le m + j'(2) \le \cdots \le m + j'(p') < m + m'$.

The applicability of Lemma 3 will be enhanced by developing a notation for products of c_n. Formally, for integers $k, p \ge 0$, $q \ge 1$, with q dividing $(p-k)$, recursively define $P_{k,\,p;\,q} = P_{k,\,p-q;\,q} c_p$ if $p > k$, and $P_{k,\,k;\,q} = c_k$. If $p < k$, then $P_{k,\,p;\,q} = \phi$. If $q = 1$, then by abuse of notation we will drop q and let $P_{k,\,p} = P_{k,\,p;\,1}$. Similarly, we let $P_k = \lim_{p \to \infty} P_{k,\,p;\,1}$. Using this notation, Lemma 2 reads $x = P_1$.

Lemma 4:

$$P_{a+2,\,b} \supset P_{a+1,\,b-1}; P_{a,\,b-2}, \quad \text{for } a \ge 0, b \ge a+2,$$
$$P_{a+2,\,b} \supset P_{a,\,b-1}; P_{a+1,\,b-2}, \quad \text{for } a \ge 0, b \ge a+2,$$
$$P_{a,\,b} \supset P_{a,\,b}; \phi, \quad \text{for } b \ge a \ge 0.$$

Proof: First, observe that $c_2 \supset c_1; c_0$ and $c_3 \supset c_2; c_1$. If, by an induction assumption, $c_{n-2} \supset c_{n-3}; c_{n-4}$ and $c_{n-1} \supset c_{n-2}; c_{n-3}$, for some $n \ge 4$, then, by Lemma 3 and (2), $c_n \supset c_{n-1}; c_{n-2}$. Consequently, applying concatenation (that is, Lemma 3) to the $b+1-(a+2)$ alignments, $c_{a+2+i} \supset c_{a+1+i}; c_{a+i}$, $0 \le i \le b-(a+2)$, yields $P_{a+2,\,b} \supset P_{a+1,\,b-1}; P_{a,\,b-2}$.

To prove the second assertion in Lemma 4, note that $c_{a+2} \supset c_a c_{a+1}; \phi$, by (2). We then apply concatenation to this alignment and the alignments $c_{a+2+i} \supset c_{a+1+i}; c_{a+i}$, $1 \le i \le b-(a+2)$. Note that, if $b = a+2$, then $P_{a+1,\,b-2} = \phi$ and both the statement and the proof are still valid.

The last assertion in Lemma 4 is obvious.

Corollary: $P_{a+2} \supset P_{a+1}; P_a$, $P_{a+2} \supset P_a; P_{a+1}$, $P_a \supset P_a; \phi$.

Proof: Let b go to infinity in Lemma 4.

Examples: Using Lemmas 3 and 4 and the Corollary, we can explore Hofstadter's conjecture, (1), for $m = 2, 3, 4$.

$m = 2$: By applying concatenation to $d \supset d; \phi$ and $P_3 \supset P_2; P_1$, we infer $x_2 \supset x; x$.

<u>$m=3$</u>: The assertion $P_3 \supset P_1; P_2$ is equivalent to $x_3 \supset x; x_1$.

<u>$m=4$</u>: Note that $x_4 = cdP_4, x = dP_2$, and $x_3 = P_3$. Therefore, applying concatenation to the alignments $cd \supset d; c$ and $P_4 \supset P_2; P_3$ implies that $x_4 \supset x; cx_3$. Consequently, by Lemma 1, (1) cannot hold for $m = 4$, since x_2 begins with a d. Similar reasoning shows that (1) is false for $m = 9, 12, \ldots$.

To generalize the $m = 4$ case precisely, recall Zeckendorf's result that every integer m can be represented uniquely as a sum of nonconsecutive Fibonacci numbers, $m = \sum_{i\geq 2} \varepsilon(i)F_i$, with $\varepsilon(i)$ in $\{0, 1\}$, $\varepsilon(i) = 0$ if $\varepsilon(i+1) = 1$, and $\varepsilon(n) = 1$ with $\varepsilon(i) = 0$ for $i \geq n+1$, for some integer $n \geq 2$. The *ascending* set of $\varepsilon(i)$ is the *Fibonacci representation* of m [9]. We define an injective map from nonnegative integers to finite binary strings, $m^* = s$, such that s has length $n-1$ and the i^{th} component of s equals $\varepsilon(i+1)$ for $1 \leq i \leq n-1$.

We will use standard conventions about exponents and string concatenations. For example, $54^* = (01)^4$. In the sequel, in the proofs of Lemma 5 and Theorem 1, certain closed formulas will be given for $(m+1)^*$ and $(m-2)^*$. The relationships between m^* and $(m \pm j)^*$ can be "translated" easily into well-known identities. For example, the assertion that, *if* $m^* = (10)^k 1$ *for some* $k \geq 0$, *then* $(m+1)^* = (00)^k 01$ is seen to correspond to the identity $F_2 + F_4 + \cdots + F_{2k+2} = F_{2k+3} - 1$.

Therefore, in the proofs of Lemma 5 and Theorem 1, these closed formulas will simply be stated without further elaboration.

Some of the relationships between m^* and the m^{th} character of x are explored in [3]. The examples for which (1) fails, $m = 4, 9, 12, 17, 22, 25, 30, 33, \ldots$, have Fibonacci representations beginning with a one followed by an odd number of zeros. This suggests the following *modified Hofstadter's conjecture*:

For all $m \geq 2$, if $m^* = 10^{2k+1}1s$, for some integer $k \geq 0$ and some binary string s, then

$$x_m \supset x;\ c\, x_{m-1}. \tag{3}$$

Otherwise, (1) holds.

Remark: By the examples presented after Lemma 4 and its corollary, the modified Hofstadter conjecture is true for $m = 2, 3, 4$.

We now state all identities needed in the proofs of Lemma 5 and Theorem 1:

$$c_1 P_{2,2k;2} = c_{2k+1}, \quad \text{for } k \geq 1, \tag{4}$$

$$c_2 P_{3,2k-1;2} = c_{2k}, \quad \text{for } k \geq 1, \tag{5}$$

$$P_{3,2k+1;2} = P_{1,2k}, \quad \text{for } k \geq 1, \tag{6}$$

$$P_{2,2k;2} = c\, P_{1,2k-1}, \quad \text{for } k \geq 1, \tag{7}$$

$$c_1 P_{4,2k;2} = P_{1,2k-1}, \quad \text{for } k \geq 1, \tag{8}$$

$$P_{a+1,b-2} c_{b+1} = c_{a+1} P_{a+2,b} = P_{a+1,b} \text{ if } a+1 \leq b-1. \tag{9}$$

For $t \geq 2$, and integers $K(i)$, with $K(i+1) \geq K(i) + 2$, $j \leq i \leq t-1$, with j in $\{0, 1\}$,

$$P_{K(j)+1,\,K(j+1)-2} \cdots P_{K(t-1)+1,\,K(t)-2}c_{K(t)+1} = P_{K(j)+1,\,K(j+1)}\left\{P_{K(j+1)+2,\,K(j+2)} \cdots P_{K(t-1)+2,\,K(t)}\right\}, \tag{10}$$

the expression in braces being empty if $t < j+2$.

To prove (4), note that, if $k=1$, then $c_1c_2 = c_3$ while, if $k>1$, then, by (2) and an induction assumption, $c_{2k+1} = c_{2k-1}c_{2k} = c_1P_{2,\,2k-2;\,2}c_{2k} = c_1P_{2,\,2k;\,2}$. The proofs of (5)-(7) also follow from (2) and an induction assumption. Equation (8) follows from (7) by cancelling the leftmost c on both sides of the equation.

To prove (9) note that, if $a+1 \le b-2$, then, by (2), $P_{a+1,\,b-2}c_{b+1} = P_{a+1,\,b} = c_{a+1}P_{a+2,\,b}$ while, if $a+1 = b-1$, then $P_{a+1,\,b-2} = \phi$, so that (9) becomes $c_{b+1} = c_{b-1}P_{b,\,b} = P_{b-1,\,b}$, which follows from (2). Note, however, that, if $a+1 \ge b$, (9) is false. Equation (10) follows from (9) by a straightforward induction.

Definition 3: Given an integer m, a strictly increasing function f on the positive integers is said to be a *representation of* x_m if $x_m = c_{f(1)}c_{f(2)}c_{f(3)} \cdots$.

To each integer $m \ge 1$ with Fibonacci representation, $\varepsilon(i)$, $i \ge 2$, with $\varepsilon(n) = 1$, $\varepsilon(i) = 0$ for $i \ge n+1$, we associate a triple $\langle n, j, z\rangle$, where $n-j$ is the total number of ones in the Fibonacci representation ε of m, and z is a strictly increasing sequence, $z(1), z(2), \ldots, z(j)$ with $\varepsilon(z(i)+1) = 0$, $1 \le i \le j$. As an example, if $m = 54$, then $n = 9$, $j = 4$, and $z(i) = 2i-1$ for i = 1, 2, 3, 4. We now describe a canonical representation of x_m.

Lemma 5: Given an integer $m \ge 2$ and its associated triple, $\langle n, j, z\rangle$, the function f, defined by $f(i) = z(i)$, $1 \le i \le j$, $f(j+1+t) = n+t$, $t = 0, 1, 2, 3\ldots$, is a representation of x_m.

Proof: To start an induction argument, we treat the case $m = 2$. If $m = 2$, then $m^* = 01$, $n = 3$, $j = 1$, and $z(1) = 1$. Clearly, $x_m = c_1P_3$ as required. The induction step has three cases.

Case 1—$m^* = 00s$ with s a binary string: Clearly $(m+1)^* = 10s$. By induction, we may assume that a representation f of x_m exists such that $f(i) = i$, $i = 1, 2$. Thus, $x_m = c_1c_2y$ for some infinite string y and, consequently, $x_{m+1} = c_2y$ as required.

Case 2—$m^* = (01)^k00s$ with $k \ge 1$ and s a (possibly empty) binary string: Then $(m+1)^*$ = $(00)^k10s$. By induction, we may assume that there is a representation f of x_m such that, whether s is empty or not, $f(i) = 2i-1$, $1 \le i \le k$, and $f(k+i) = 2k+i$, $i = 1, 2$. Thus, $x_m = P_{1,\,2k+1;\,2}y$ for some infinite string y and, therefore, by (6), $x_{m+1} = P_{3,\,2k+1;\,2}y = P_{1,\,2k}y$ as required.

Case 3—$m^* = (10)^k0s$ with $k \ge 1$ and s a binary string: Then $(m+1)^* = (00)^{k-1}010s$. By induction, we may assume there is a representation f of x_m with $f(i) = 2i$, $1 \le i \le k$, $f(k+1) = 2k+1$. Thus, $x_m = P_{2,\,2k;\,2}y$ for some infinite string y and, consequently, by (8), $x_{m+1} = c_1P_{4,\,2k;\,2}y = P_{1,\,2k-1}y$ as required.

Clearly, for each $m \ge 2$, one of these three cases must hold and, consequently, the proof is complete.

Theorem 1: The modified Hofstadter's conjecture is true for all $m \ge 2$.

Proof: The theorem has already been verified for $m = 2, 3, 4$. If $m \geq 5$, then there exist integers $t \geq 1$, $k(1), k(2), \dots k(t)$, $k(i) \geq 1$, such that either

$$m^* = 10^{k(1)}1\dots 0^{k(t)}1 \tag{11}$$

or

$$m^* = 0^{k(1)}10^{k(2)}1\dots 0^{k(t)}1. \tag{12}$$

To prove the theorem, we need the Fibonacci representations for $(m-1)^*$ and $(m-2)^*$. There are now four cases—1A, 1B, 1C, and 1D—depending on whether m^* begins with a 1 or not and depending on whether $k(1)$ is even or odd.

Case 1A—(11) holds, with $k(1)$ odd: Then, clearly, $(m-1)^* = 0^{k(1)+1}1\{0^{k+2)}1\dots 0^{k(t)}1\}$, the expression in braces being empty if $t = 1$.

Define integers

$$K(0) = 0,\ K(i+1) = K(i)+1+k(i+1),\ i = 0, 1, \dots, t-1. \tag{13}$$

Clearly,

$$K(i+1) \geq K(i)+2,\ i = 0, 1, \dots, t-1. \tag{14}$$

By Lemma 5,

$$x_m = P_{K(0)+2,K(1)} \cdots P_{K(t-1)+2,K(t)} P_{K(t)+2} \tag{15}$$

and

$$x_{m-1} = P_{1,K(1)}\left\{P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2,K(t)}\right\} P_{K(t)+2}. \tag{16}$$

The expression in braces is empty if $t = 1$.

Using Lemma 4 and its corollary, we apply concatenation to the alignments

$$P_{2,K(1)} \supset P_{1,K(1)-1};\ P_{0,K(1)-2},$$

$$P_{K(i)+2,K(i+1)} \supset P_{K(i),K(i+1)-1};\ P_{K(i)+1,K(i+1)-2},\quad 1 \leq i \leq t-1,$$

and

$$P_{K(t)+2} \supset P_{K(t)};\ P_{K(t)+1},$$

to obtain

$$x_m \supset x; y \tag{17}$$

with

$$y = cP_{K(0)+1,K(1)-2} \cdots P_{K(t-1)+1,K(t)-2} P_{K(t)+1}. \tag{18}$$

Since $k(1)$ is odd, we must prove (3). By (17), to prove (3), it suffices to prove $y = cx_{m-1}$. Therefore, by (16) and (18), it suffices to prove

$$P_{K(0)+1,K(1)-2} \cdots P_{K(t-1)+1,K(t)-2} c_{K(t)+1} = P_{K(0)+1,K(1)}\left\{P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2,K(t)}\right\},$$

which follows from (14) and (10).

Case 1B—(11) holds with

$$\mathbf{k(1) = 2k,\ k \geq 1:} \tag{19}$$

For notational reasons, it will be clearer in cases 1B, 1C, and 1D to first assume that $t \geq 2$. The $t = 1$ case can then be treated separately. If $t \geq 2$, then $(m-2)^* = (10)^k 100^{k(2)}1\dots 0^{k(t)}1$. Define $K(i)$ as in (13). Then (14) and (15) still hold. By Lemma 5, we have

$$x_{m-2} = P_{2,2k;2}c_{2k+2}P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2,K(t)}P_{K(t)+2}. \tag{20}$$

Proceeding as in case 1A, we apply Lemmas 3 and 4. Equations (17) and (18) still hold.

Since $k(1)$ is even, we must prove (1) instead of (3). By (17), to prove (1) it suffices to prove $y = x_{m-2}$. Therefore, by (18) and (20), it suffices to prove

$$cP_{K(0)+1,K(1)-2} \cdots P_{K(t-1)+1,K(t)-2}c_{K(t)+1} = P_{2,2k;2}c_{2k+2}P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2,K(t)}. \tag{21}$$

By (19) and (13), $K(0)+1=1$ and $K(1)-2=\{k(1)+1\}-2=2k-1$. Hence, by (7), proving (21) is equivalent to proving

$$P_{2,2k;2}P_{K(1)+1,K(2)-2} \cdots P_{K(t-1)+1,K(t)-2}c_{K(t)+1} = P_{2,2k;2}c_{K(1)+1}P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2,K(t)},$$

which follows from (14) and (10).

To complete the proof of case 1B, we treat the $t=1$ case: If $t=1$, then m^*, $(m-2)^*$, x_m, and x_{m-2} are $10^{2k}1$, $(10)^k1$, $P_{2,2k+1}P_{2k+3}$, and $P_{2,2k;2}P_{2k+2}$, respectively. Using Lemma 4, we apply concatenation to the alignments $P_{2,2k+1} \supset P_{1,2k}; cP_{1,2k-1}$ and $P_{2k+3} \supset P_{2k+1}; P_{2k+2}$ to obtain (17) with $y = cP_{1,2k-1}P_{2k+2}$ To prove (1), it suffices to prove $x_{m-2} = y$, which follows from (7).

Case 1C—(12) holds with (19): For $t \geq 2$, we have $(m-2)^* = 10(01)^{k-1}00^{k(2)}1\ldots0^{k(t)}1$. Define

$$K(0)=0,\ K(1)=k(1),\ K(i+1)=K(i)+1+k(i+1),\ 1 \leq i \leq t-1. \tag{22}$$

Note that, by (19), (14) still holds. By Lemma 5,

$$x_m = P_{1,K(1)}P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2,K(t)}P_{K(t)+2} \tag{23}$$

and

$$x_{m-2} = c_2P_{3,2k-1;2}c_{2k+1}P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2,K(t)}P_{K(t)+2}. \tag{24}$$

Using Lemma 4 and its corollary, we apply concatenation to the alignments

$$P_{1,K(1)} \supset P_{1,K(1)}; \phi,$$

$$P_{K(1)+2,K(2)} \supset P_{K(1)+1,K(2)-1}; P_{K(1),K(2)-2},$$

$$P_{K(i)+2,K(i+1)} \supset P_{K(i),K(i+1)-1}; P_{K(i)+1,K(i+1)-2},\ 2 \leq i \leq t-1,$$

and

$$P_{K(t)+2} \supset P_{K(t)}; P_{K(t)+1},$$

to obtain (17) with

$$y = P_{K(1),K(2)-2}\left\{P_{K(2)+1,K(3)-2} \cdots P_{K(t-1)+1,K(t)-2}\right\}P_{K(t)+1}, \tag{25}$$

the expression in braces being empty if $t=2$.

By (17), to prove (1) it suffices to prove $y = x_{m-2}$. Therefore, by (25) and (24), it suffices to prove

$$P_{K(1),K(2)-2}\left\{P_{K(2)+1,K(3)-2} \cdots P_{K(t-1)+1,K(t)-2}\right\}c_{K(t)+1} = c_2P_{3,2k-1;2}c_{2k+1}P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2,K(t)}. \tag{26}$$

By (19) and (22), $K(1)=2k$ so that, by (5), proof of (26) is reduced to proof of

$$P_{K(1)+1,K(2)-2} \cdots P_{K(t-1)+1,K(t)-2}c_{K(t)+1} = P_{K(1)+1,K(2)}\left\{P_{K(2)+2,K(2)} \cdots P_{K(t-1)+2,K(t)}\right\},$$

which follows from (10) and (14).

It remains to treat the case $t=1$. If $k=1$ also, then case 1C reduces to (1) with $m=3$, which has already been treated. If $k>1$, then $m^*, (m-2)^*, x_m$, and x_{m-2} are $0^{2k}1, 10(01)^{k-1}$, $P_{1,2k}P_{2k+2}$, and $c_2P_{3,2k-1;2}P_{2k+1}$, respectively. By concatenating the alignments, $P_{1,2k} \supset P_{1,2k}; \phi$ and $P_{2k+2} \supset P_{2k+1}; P_{2k}$, we derive (17) with $y = P_{2k} = c_{2k}P_{2k+1}$. To prove (1), we must prove that $y = x_{m-2}$, which follows from (5).

Case 1D—(12) holds with

$$\mathbf{k(1) = 2k+1,\ k \geq 0:} \tag{27}$$

For $t \geq 2$, we have $(m-2)^* = 0(01)^k 00^{k(2)}1 \ldots 0^{k(t)}1$. Define $K(i)$ by (22). Then (14) and (23) still hold. By Lemma 5,

$$x_{m-2} = c_1P_{2,2k;2}c_{2k+2}P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2,K(t)}P_{K(t)+2}. \tag{28}$$

Proceeding as in case 1C, we have (17) with (25). By (17), to prove (1) it suffices to show that $y = x_{m-2}$. Therefore, by (25) and (28), it suffices to show

$$c_1P_{2,2k;2}c_{2k+2}P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2}, K(t) = c_{K(1)}P_{K(1)+1,K(2)-2} \cdots P_{K(t-1)+1,K(t)-2}c_{K(t)+1}. \tag{29}$$

By (27) and (22), $K(1) = 2k+1$; therefore, by (4), proof of (29) reduces to proof of

$$c_{K(1)}c_{K(1)+1}P_{K(1)+2,K(2)} \cdots P_{K(t-1)+2,K(t)} = c_{K(1)}P_{K(1)+1,K(2)-2}\left\{P_{K(2)+1,K(3)-2} \cdots P_{K(t-1)+1,K(t)-2}\right\}c_{K(t)+1},$$

which follows from (10) and (14).

The $t=1$ case is treated in a manner similar to the $t=1$ case in 1B and 1C. This completes the proof of Theorem 1.

The proof and formulation of a modified Hofstadter's conjecture for other irrationals remains an open and difficult problem. To generalize (3), it seems reasonable to conjecture that, for every irrational, there exists a finite set of strings and a finite set of integers such that, for every m, $x_m \supset x;\ Qx_{m-n}$ with Q and n belonging to these finite sets. The authors announced a proof of the deceptively simple case $\alpha = \sqrt{2}-1$ with m equal to sums of Pell numbers [5]. This proof required considerable alteration of Definition 1 and Lemma 3, as well as a more developed form of Lemma 4.

ACKNOWLEDGMENTS

The authors wish to thank the referees and the editor for their patience and suggestions that significantly improved the appearance of this manuscript.

REFERENCES

1. D. Doster. "Problem E3391." *Amer. Math. Monthly* **97** (1990):528.
2. A. Engel. *Mathematical Experimentation and Simulation with the Computer*. The New Mathematical Library Series #35. Washington D.C.: The Mathematical Association of America, 1993.
3. A. S. Fraenkel, J. Levitt, & M. Shimshoni. "Characterization of the Set of Values of $f(n) = [n\theta]$, $n = 1, 2, \ldots$." *Discrete Math.* **2** (1972):335-45.
4. A. S. Fraenkel, M. Mushkin, & U. Tassa. "Determination of $[n\theta]$ by Its Sequences of Differences." *Can. Math. Bull.* **21** (1978):441-46.

5. R. J. Hendel & S. A. Monteferrante. "Hofstadter's Conjecture." Paper presented at the Mathematical National Meeting, San Francisco, 1991.
6. D. R. Hofstadter. *Eta-Lore*, p. 13. First presented at the Stanford Math Club, Stanford, California, 1963.
7. D. R. Hofstadter. *Godel, Escher, and Bach*, p. 137. New York: Vantage Books, 1980.
8. F. Mignosi. "Infinite Words with Linear Subword Complexity." *Theoretical Computer Science* **66** (1989):221-42.
9. J. O. Shallit. "A Generalization of Automatic Sequences." *Theoretical Computer Science* **61** (1988):1-16.
10. K. B. Stolarsky. "Beatty Sequences, Continued Fractions, and Certain Shift Operators." *Can. Math. Bull.* **19** (1976):473-82.
11. K. P. Tognetti, G. Winley, & T. van Ravenstein. "The Fibonacci Tree, Hofstadter and the Golden String." In *Applications of Fibonacci Numbers*, **3**:325-34. Ed. G. E. Bergum, A. N. Philippou, and A. F. Horadam. Netherlands: Kluwer, 1990.
12. B. A. Venkoff. *Elementary Number Theory*, pp. 65-68. Trans. and ed. H. Alderson. Gronigen: Wolters-Noordhof, 1970.

AMS Classification Numbers: 68R15, 20M35, 20M05

❖❖❖