Journal of Integer Sequences, Vol. 29 (2026),
Article 26.3.3

Pointwise Order of Generalized Hofstadter Functions G, H, and Beyond

Pierre Letouzey
IRIF
Université Paris Cité, CNRS, Inria
F-75013 Paris
France
letouzey@irif.fr

Shuo Li
Hangzhou International Innovation Institute
Beihang University
166 Shuanghongqiao Street
Pingyao Town, Yuhang District, Hangzhou, 311115
China
shuoli@buaa.edu.cn

Wolfgang Steiner
IRIF
Université Paris Cité, CNRS
F-75013 Paris
France
steiner@irif.fr

Abstract

Hofstadter's G function is recursively defined via $G(0) = 0$ and then $G(n) = n - G(G(n-1))$. Following Hofstadter, we vary the number k of nested recursive calls in this equation and obtain a family of functions (F_k). Here we establish that this family is ordered pointwise: for all k and n, we have $F_k(n) \leq F_{k+1}(n)$. To achieve this, we make a detour via infinite morphic words generalizing the Fibonacci word. We prove various properties of these words, concerning the lengths of substituted prefixes of these words and the number of occurrences of specific letters in these prefixes. We also relate the limits of $\frac{1}{n}F_k(n)$ to the frequencies of letters in the considered words. We provide a certified formalization of all these results in the Rocq proof assistant.

1 Introduction

1.1 The functions

For each integer $k \geq 1$, we recursively define the function[1]

$$F_k : \mathbb{N} \to \mathbb{N}, \quad n \mapsto \begin{cases} 0, & \text{if } n = 0; \\ n - F_k^k(n-1), & \text{otherwise.} \end{cases}$$

The function F_k is well defined, since one may prove that $0 \leq F_k(n) \leq n$ for all $n \geq 0$. This family of functions is due to Hofstadter [13, Chap. 5]. In particular, F_2 is Hofstadter's function G (see OEIS entry A005206 [22, 13, 5, 10]). Note that $G(n) = \lfloor (n+1)/\varphi \rfloor$ where φ is the golden ratio $(1+\sqrt{5})/2$. Similarly, F_3 is Hofstadter's function H (see OEIS A005374), and the generalization to higher degrees of recursive nesting in the definition of F_k was already suggested by Hofstadter [13]. To be complete, the OEIS database already includes F_4 as A005375, F_5 as A005376, and F_6 as A100721. On the other hand, we choose to start this sequence with F_1, where only one recursive call is done, leading to a function that can easily be shown to verify $F_1(n) = \lfloor (n+1)/2 \rfloor = \lceil n/2 \rceil$. Throughout this paper, we never consider the case $k = 0$: although the previous definition could be extended and gives a non-recursive function F_0, this F_0 has too little in common with the other F_k functions to be of much interest.

1.2 A monotonicity problem

Small values of the functions F_1 to F_5 are displayed in Figure 1.1. From this figure, one may easily guess that F_1 is everywhere below or equal to F_2, similarly that F_2 is below or equal to F_3, and so forth. Indeed, our main result is the following theorem.

Theorem 1.1. *For all $k \geq 1$, $n \geq 0$, we have $F_k(n) \leq F_{k+1}(n)$.*

This property seems to defy any attempt to prove it directly via induction on the functions F_k. Therefore, a different approach is used. We relate each function F_k to a morphic word x_k and turn the function comparison into an equivalent statement about these words. This gives Theorem 7.4 below, which contains Theorem 1.1 as a special case.

[1] In all this work, the k-th exponent of a function denotes its k-th iterate. For instance, $F_2^2(n)$ is $F_2(F_2(n))$, not the square of $F_2(n)$.

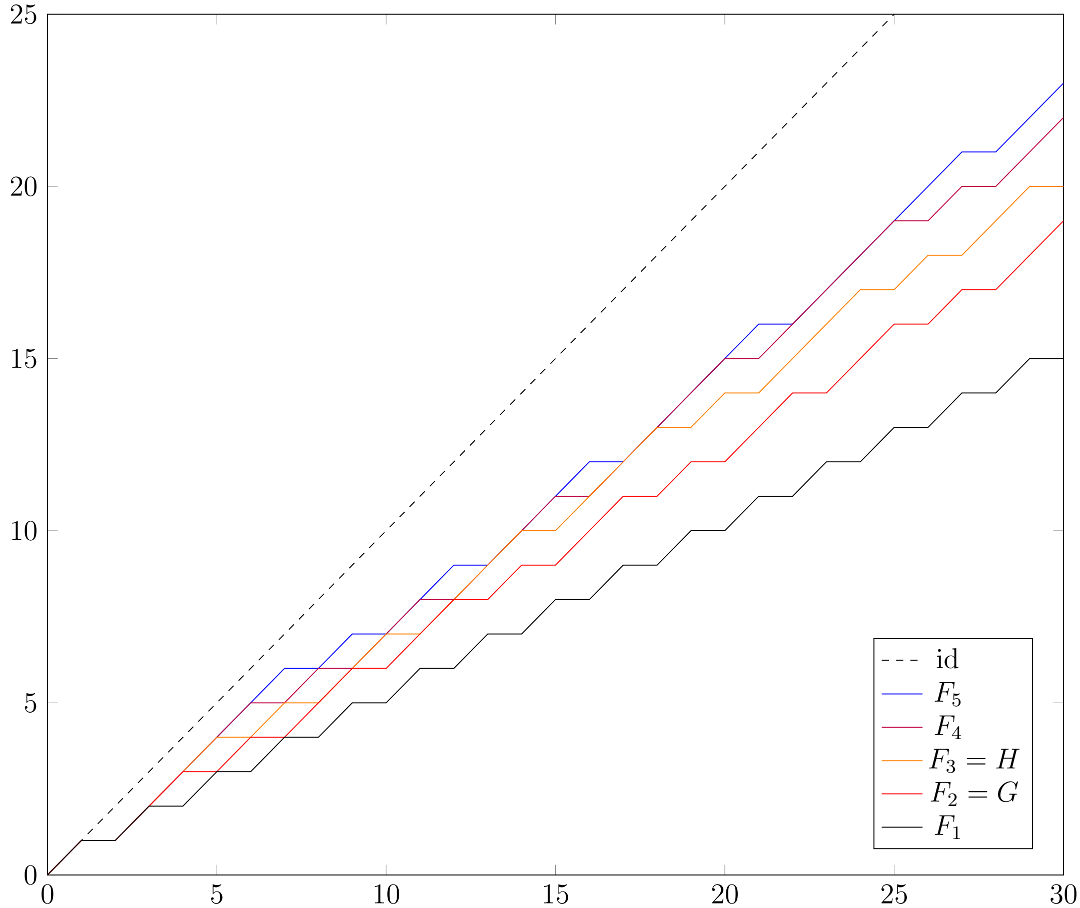

Figure 1.1: Plotting $F_1, F_2, \ldots, F_5$

1.3 Substitutions and morphic words

For $k \geq 1$, let τ_k be the substitution on the alphabet $\{1, 2, \ldots, k\}$ defined by

$$\begin{aligned} \tau_k :\; & k \mapsto k\,1, \\ & i \mapsto i+1, \quad \text{for } 1 \leq i < k, \end{aligned}$$

and let $x_k = x_k[0]x_k[1]\cdots \in \{1, 2, \ldots, k\}^\infty$ be the fixed point of τ_k. Here, a substitution (or morphism) on an alphabet A is a map $\tau : A^* \to A^*$ satisfying $\tau(uv) = \tau(u)\tau(v)$ for all $u, v \in A^*$, where A^* denotes the set of finite words with letters in A. The map τ is therefore defined by its value on the letters of A, and it is extended in a natural way to infinite words (or sequences) $w = w[0]w[1]\cdots \in A^\infty$ by setting $\tau(x) = \tau(x[0])\tau(x[1])\cdots$. For more on substitutions, we refer to Pytheas Fogg [8].

These substitutions and words are not novel. For instance, τ_1 operates on the singleton alphabet $\{1\}$ in such a way that $\tau_1(1) = 11$, and hence $x_1 = 1^\infty$. It is the only case where

$$x_2 = 2122121221221212212122122121221221212212\cdots$$
$$x_3 = 3123313123123312331312331312312331312312\cdots$$
$$x_4 = 4123441412412341234412344141234414124123\cdots$$
$$x_5 = 5123455151251235123451234551234551512345\cdots$$

Figure 1.2: Infinite words $x_2, \dots, x_5$ (with largest letter in red)

x_k is ultimately periodic. For $k = 2$, we recover the well-known Fibonacci substitution and word.[2] The substitution τ_3 already appears (up to letter renaming) as $\sigma(1, 0)$ in Pytheas Fogg [8, Ex. 8.1.2], as an example of a *modified Jacobi–Perron substitution*. More generally, the substitutions τ_k and words x_k can be associated with the Rényi expansion of 1 in base β_k, where β_k is the positive zero of $X^k - X^{k-1} - 1$, and were studied by Frougny, Masáková and Pelantová [9]. When $k \geq 2$, this expansion can be written as $1(0)^{k-2}1$. As a consequence, the factor complexity of x_k is $n \mapsto (k-1)n + 1$; in particular, x_2 is Sturmian [9].

1.4 New results

In this article, in addition to the aforementioned monotonicity of F_k (Theorem 1.1), we exhibit several relations between the functions F_k and the words x_k. For instance, for $k \geq 1$ and a position $n \geq 0$, the value of $x_k[n]$ is either the first index $j \leq k$ such that $F_k^j(n+1) = F_k^j(n)$, or k if there is no such small index; see Proposition 4.4 below. In particular, $x_k[n]$ is the letter 1 exactly when $F_k(n+1) = F_k(n)$, and this also leads to the fact that $F_k(n)$ is equal to the number of non-1 letters in the first n letters of x_k; see Proposition 4.3. Such properties were already known in the case $k = 2$, where $F_2 = G = \lfloor (n+1)/\varphi \rfloor$ and x_2 amounts to the Fibonacci word; see, for instance, Allouche and Shallit [2, Ex. 8.1.3]. The extended results to an arbitrary parameter k seem new.

All these relations between the functions F_k and the words x_k stem from the key Theorem 3.1 and its Corollary 3.2: F_k and their iterates admit Galois connections [7] (i.e., almost inverses) that can be simply expressed in terms of the length of τ_k-substituted prefixes of x_k. These lengths are called L_k here; see Proposition 2.3.

The words x_k are primarily considered here in order to study the monotonicity of the functions F_k. Nonetheless, several results on x_k are also of interest on their own. For example, for $k \geq 1$ and $n \geq 0$, we prove that there are more occurrences of the letter k among the first n letters of x_k than occurrences of the letter $k+1$ among the first n letters of x_{k+1} (Theorem 7.2). Similarly, the letter 1 occurs more often among the first n letters of x_k than among the first n letters of x_{k+1} (Proposition 8.1). We denote the number of letters i among the prefix of x_k of length n by $C_k^{(=i)}(n)$; see Section 4.

[2] Usually, the letter k is replaced by the letter 0 in the definition of τ_k and x_k. We choose the letter k in order to simplify some formulas and to extend the definition without modification to the case $k = 1$.

1.5 The Rocq formalization

All the results presented in this article have been formally certified using the Rocq proof assistant [19] (previously known as Coq). This implementation in Rocq is freely available [15]. Thanks to this Rocq formalization, we ensure precise definitions and statements while preventing any error during the corresponding proofs. It can hence serve as a reference for the interested reader. In particular, the recommended entry point to read alongside this article is

https://github.com/letouzey/hofstadter_g/blob/main/Article1.v

The current article tries to be faithful to this formal work while staying readable by a large audience, at the cost of possible remaining mistakes during the transcription.

As a quick illustration, here is one possible Rocq encoding of the F_k function:[3]

```
Fixpoint recf k p n : nat := match p with
| 0 => 0
| S p' => n - Nat.iter k (recf k p') (n-1)
end.
Definition f k n := recf k n n.
```

Note that `Nat.iter k` iterates a given function `k` times. This encoding departs slightly from the previous definition of F_k: in order to satisfy the Rocq constraints on recursive definitions, an extra parameter `p` is initialized to `n` and decreased in recursive calls. Then we proved that this Rocq definition fulfills the desired equation:[4]

```
Lemma f_eqn : forall k n : nat, f k n = n - Nat.iter k (f k) (n-1).
```

In this setting, our main Theorem 1.1 has a very simple Rocq statement:[5]

```
Theorem f_grows : forall k n, f k n <= f (k+1) n.
```

The provided Rocq files can be machine-checked again by any recent installation of Rocq; see the joint `README` file in the repository [15]. Since Sections 5 and 6 of this article involve real numbers, their corresponding Rocq files rely on the Rocq library `Reals` and its four logical axioms, in particular the axiom of excluded middle. All the rest of the formalization, including our key Theorem 1.1, has been performed within Rocq core logic, with no axiom. This fact can be checked via the Rocq command `Print Assumptions` on our theorems.

[3] See `Definition f` in the file `GenG.v` [15]. Faster implementations of F_k are also provided in the file `Fast.v` and proved equal with the version shown here.

[4] See `Lemma f_eqn` and its proof in the file `GenG.v` [15]. Note that the Rocq subtraction on natural numbers of type `nat` is bounded by `0`, for example `0-1 = 0`. Therefore, the statement of `f_eqn` holds both when $\mathtt{n} = 0$ and when $\mathtt{n} \neq 0$.

[5] See `Theorem f_grows` and its proof in the file `WordGrowth.v` [15].

1.6 Summary

After introducing some notation and basic properties in Section 2, we establish a key link between the functions F_k and words x_k in Section 3. In Section 4, we count some letters in prefixes of x_k. Section 5 presents some zeros of polynomials that are used in Section 6 when stating the infinitary behavior of our objects, i.e., when $n \to \infty$. In particular, we present linear equivalents of F_k and a weak form of monotonicity for the family (F_k). Section 7 contains the promised full monotonicity proof for all $n \geq 0$, at first for words, then functions, and various extensions are made or conjectured. Section 8 revisits the previous results, this time in terms of the number of occurrences of letters.

1.7 Related works

Various aspects of the functions F_k can be studied. We mention some of them, even if they do not seem to be connected to the monotonicity, which is the main focus of this article.

First, the case $k = 2$ has been well explored [5, 10]. However, the fact that F_2 has an exact expression $\lfloor (n+1)/\varphi \rfloor$ is quite particular and too specific to draw general conclusions about all the other F_k.

A nice general fact about the functions F_k is that they can be presented as right shifts of digits when numbers n are written with the appropriate numerical representations. In particular, for $k = 2$, the corresponding numerical representation is the Zeckendorf decomposition, writing numbers as sums of distinct Fibonacci numbers. This can be extended to other values of k by changing the sequence used as the base for the decomposition, from Fibonacci numbers to a similar linear recurrence [17, 5, 14, 6, 3, 21]. In particular, Meek and Van Rees [17] focus on a slightly different family of functions, which can be written here as $F_k(n + 1) - 1$. If we denote these "shifted" functions by $\tilde{F}_k$, they verify the equations $\tilde{F}_k(0) = 0$ and $\tilde{F}_k(n) = n - 1 - \tilde{F}_k^k(n-1)$; note the extra -1 when compared to the equation of F_k. In any case, having a description of $F_k(n)$ (or $\tilde{F}_k(n)$) in terms of a k-decomposition of n does not directly help for our monotonicity problem, since the k- and $(k+1)$-decompositions of n differ too much, just as a number tends to have very different base-2 and base-3 digits. The concept of k-decompositions is also related to the notions of substitutions and words that we exploit here directly.

As a final remark about F_k, note that F_k can also be presented as a companion infinite tree where the same pattern is continuously repeated: a binary node is followed on its right by $k - 1$ unary nodes. When the nodes of this tree are labelled by increasing numbers, in a breadth-first manner, left to right, then the node number n has $F_k(n)$ as parent node. This was already described by Hofstadter [13]; also see the general introduction to the problem by the first author [16]. Here again, these alternative presentations of F_k do not seem to help concerning the monotonicity, with no direct relations between trees for k and for $k + 1$.

The sequences x_k defined in Section 1.3 can be considered as generalizations of the Fibonacci word. However, one can find many different "generalized Fibonacci" words in the literature, for example the k-bonacci words, which can be defined as fixed points of the morphism satisfying $1 \mapsto k$, $i \mapsto k(i-1)$, for $2 \le i \le k$ [23], the fixed points of the morphisms $0 \mapsto 0^m 1^n$, $1 \mapsto 0$, for positive integers m, n [2, Ex. 10.11.18], and other examples [18, 11, 12].

2 Notation and basic properties

We let $\partial F_k^j(n)$ denote the difference $F_k^j(n+1) - F_k^j(n)$.

Proposition 2.1. *For all $j, k \ge 1$, the function F_k^j satisfies the following basic properties:*

(a) $F_k^j(1) = F_k^j(2) = 1$,

(b) $F_k^j(n) \ge 1$ *whenever* $n \ge 1$,

(c) $F_k^j(n) < n$ *whenever* $n \ge 2$,

(d) we have $\partial F_k(n) = 1 - \partial F_k^k(n-1)$ *for all* $n \ge 1$,

(e) hence $\partial F_k^j(n) \in \{0, 1\}$ *for all* $n \ge 0$,

(f) the function F_k^j is monotonically increasing and onto (but not one-to-one).

Proof. Direct use of the definition or easy induction on n. ☐

For $n \ge 0$, let $x_k[0{:}n) = x_k[0]x_k[1]\cdots x_k[n-1]$ denote the prefix of x_k of length n. In particular, $x_k[0{:}0)$ is the empty word, and $\left|x_k[0{:}n)\right| = n$, where $|\cdot|$ denotes the length of a word.

Note that x_k can also be seen as the fixed point of the substitution τ_k^{k-1}. Since $\tau_k^{k-1}(i) = k12\cdots(i-1)$ for all $1 \le i \le k$, this provides a decomposition of x_k into blocks delimited by the letter k (in red in Figure 1.2). The lengths of these blocks are the successive letters of x_k, and this gives us a convenient way to compute x_k. For instance, $\tau_3^2(3) = 312$, then $\tau_3^2(312) = 312\,3\,31$, etc.

Similarly, x_k is also the fixed point of τ_k^k. Since $\tau_k^k(i) = k12\cdots i$ for all $1 \le i \le k$, this other decomposition is interesting for counting occurrences of the letter 1 in x_k, because 1 occurs only in the second position of each such block.

As an alternative way of computing x_k, note that its prefixes of the form $\tau_k^j(k)$ follow the following base cases and recursive rule.

Proposition 2.2. *Let $k \ge 1$. Then*

$$\tau_k^j(k) = \begin{cases} k12\cdots j, & \text{if } 0 \le j \le k; \\ \tau_k^{j-1}(k)\,\tau_k^{j-k}(k), & \text{if } j \ge k. \end{cases}$$

Proof. We use induction on j. For $j = 0$, we have $\tau_k^0(k) = k$. Otherwise, $\tau_k^j(k) = \tau_k(\tau_k^{j-1}(k))$, and we use the induction hypothesis on $j-1$. If $j \le k$, then

$$\tau_k^j(k) = \tau_k(\tau_k^{j-1}(k)) = \tau_k(k12\cdots(j-1)) = k12\cdots j,$$

which also gives the recursive rule for $j = k$ since $k12\cdots(k-1)k = \tau_k^{k-1}(k)\tau_k^0(k)$. If $j > k$, then $j - 1 \geq k$, and

$$\tau_k^j(k) = \tau_k(\tau_k^{j-1}(k)) = \tau_k(\tau_k^{j-1-1}(k)\tau_k^{j-1-k}(k)) = \tau_k^{j-1}(k)\tau_k^{j-k}(k). \qquad \square$$

For $k \geq 1$, an important counterpart to F_k is the function

$$L_k : \mathbb{N} \to \mathbb{N}, \quad n \mapsto \left|\tau_k(x_k[0{:}n))\right|.$$

Proposition 2.3. *The lengths $L_k(n) = \left|\tau_k(x_k[0{:}n))\right|$ satisfy the following basic properties for all $k \geq 1$ and $j, n \geq 0$:*

(a) the j-th iterate of L_k satisfies $L_k^j(n) = \left|\tau_k^j(x_k[0{:}n))\right|$,

(b) $L_k^j(0) = 0$ and $L_k^j(n+1) = L_k^j(n) + \left|\tau_k^j(x_k[n])\right|$,

(c) $L_k^j(1) = j + 1$ when $j \leq k$ and $L_k^j(1) = L_k^{j-1}(1) + L_k^{j-k}(1)$ when $j \geq k$,

(d) the function L_k^j is strictly monotonically increasing over n,

(e) $L_k^j(n) \geq n$, with equality only when $j = 0$ or $n = 0$,

(f) for $n \geq 1$, the function $j \mapsto L_k^j(n)$ is strictly monotonically increasing over j.

Proof. First, $\tau_k(x_k[0{:}n))$ is the prefix of x_k of length $L_k(n)$. It is hence equal to $x_k[0{:}L_k(n))$. We claim that, more generally, $\tau_k^j(x_k[0{:}n))$ is the prefix of x_k of length $L_k^j(n)$ for all $j \geq 0$. Since $\tau_k^j(x_k[0{:}n))$ is a prefix of x_k, we only have to prove that its length is equal to $L_k^j(n)$, and this follows inductively for $j \geq 1$ from

$$\left|\tau_k^j(x_k[0{:}n))\right| = \left|\tau_k(\tau_k^{j-1}(x_k[0{:}n)))\right| = \left|\tau_k(x_k[0{:}L_k^{j-1}(n)))\right| = L_k(L_k^{j-1}(n)) = L_k^j(n).$$

Point (b) is a direct consequence of this more general claim. Point (c) directly comes from Proposition 2.2.

Now a key fact: for all finite words w, we have $|w| \leq |\tau_k(w)|$. This inequality is even strict when w is a non-empty prefix of x_k, since $x_k[0] = k$ and $|\tau_k(k)| = 2$. The remaining points are direct consequences of this key fact. $\square$

Proposition 2.4. *For $j \geq 0$ and $k, m \geq 1$, there exists a unique $n \geq 1$ such that $L_k^j(n-1) < m \leq L_k^j(n)$.*

Proof. Standard consequence of $L_k^j(0) = 0$ and the strict monotonicity of L_k^j. $\square$

3 Relating functions and word lengths

We now establish that L_k^j allows expressing preimages of the function F_k^j.

Theorem 3.1. *Let $j \geq 0$ and $k \geq 1$.*

(a) For all $n \geq 1$, we have

$$F_k^{-j}(\{n\}) = \big(L_k^j(n-1), L_k^j(n)\big] \cap \mathbb{N}.$$

(b) Equivalently, for all $m \geq 1$, we have

$$L_k^j\big(F_k^j(m)-1\big) < m \leq L_k^j\big(F_k^j(m)\big). \tag{3.1}$$

Proof. We first prove the equivalence of (a) and (b). The case (a) directly implies (b); just replace n by $F_k^j(m)$. Now we prove the set equation in (a) assuming (b): If $m \in F_k^{-j}(\{n\})$, then $F_k^j(m) = n$, and (3.1) becomes $L_k^j(n-1) < m \leq L_k^j(n)$. Hence m is a natural number in the right interval. On the other hand, if $m \in \big(L_k^j(n-1), L_k^j(n)\big] \cap \mathbb{N}$, then Proposition 2.4 and (3.1) imply that $F_k^j(m) = n$, so $m \in F_k^{-j}(\{n\})$. The two sets in (a) are hence equal.

We now prove the statement (b). The case $k = 1$ must be separately handled, since our general proof below requires the letters 1 and k to differ. Fortunately, when $k = 1$ we have $L_1^j(n) = 2^j n$ and $F_1^j(m) = \lceil m/2^j \rceil$, which easily resolves this case. We can also separately handle the case $j = 0$, which is obvious since both L_k^0 and F_k^0 are the identity functions.

From now on, we assume $k \geq 2$ and prove the inequalities (3.1) for all $j \geq 1$ by strong induction on $m \geq 1$. For $m \in \{1, 2\}$ and $j \geq 1$, we have $F_k^j(m) = 1$ by Proposition 2.1, while $L_k^j(0) = 0$ and $L_k^j(1) > L_k^0(1) = 1$ by Proposition 2.3. Hence the desired inequalities hold in this case.

Let $n \geq 2$, and assume that (3.1) holds for all $1 \leq m \leq n$, $j \geq 1$. We now prove the inequalities in (3.1) for $m = n+1$, first in the case $j = 1$, and then for $j \geq 2$.

Case $j = 1$. We use the following instances of the induction hypothesis:

$$L_k(F_k(n)-1) < n \leq L_k(F_k(n)), \tag{3.2}$$

$$L_k^k(F_k^k(n)-1) < n \leq L_k^k(F_k^k(n)), \tag{3.3}$$

$$L_k^k(F_k^k(n-1)-1) < n-1 \leq L_k^k(F_k^k(n-1)). \tag{3.4}$$

Also recall from Proposition 2.1 that

$$F_k(n+1) - F_k(n) = 1 - \big(F_k^k(n) - F_k^k(n-1)\big)$$

and $F_k^k(n) - F_k^k(n-1)$ is either 0 or 1. We separately handle these two possible differences.

Case $j = 1$, $F_k^k(n) = F_k^k(n-1)$. If $F_k^k(n) = F_k^k(n-1)$, then $F_k(n+1) = F_k(n) + 1$. From (3.2) and the strict monotonicity of L_k (Proposition 2.3), we get

$$n \leq L_k(F_k(n)) < L_k(F_k(n)+1) = L_k(F_k(n+1)).$$

Hence the right inequality in (3.1) holds for $m = n+1$, $j = 1$. We claim that $L_k(F_k(n)) = n$. If this were not the case, then (3.2) would give $L_k(F_k(n)-1) \leq n-1 < n < L_k(F_k(n))$. For the words corresponding to these lengths, this would imply that $x_k[n-1]x_k[n]$ is a subword of $\tau_k(x_k[F_k(n)-1])$, and in particular that $|\tau_k(x_k[F_k(n)-1])| \geq 2$. This could only happen when $x_k[F_k(n)-1] = k$, and thus $x_k[n] = 1$. From $F_k^k(n) = F_k^k(n-1)$, (3.3) and (3.4), we get $L_k^k(F_k^k(n)-1) \leq n-2 < n-1 < L_k^k(F_k^k(n))$, and hence $x_k[n-2]x_k[n-1]$ is a subword of the block $\tau_k^k(x_k[F_k^k(n)-1])$. Recall that the letter 1 occurs in this kind of block only at the second position. But here, either $x_k[n]$ is still in the block, but at least in third position, or it starts the next block, and hence $x_k[n] = k$. Both cases are incompatible with $x_k[n] = 1$. Therefore, we have $L_k(F_k(n)) = n$. Hence

$$L_k(F_k(n+1)-1) = L_k(F_k(n)) = n < n+1,$$

and also the left inequality in (3.1) holds here for $m = n+1$, $j = 1$.

Case $j = 1$, $F_k^k(n) = F_k^k(n-1) + 1$**.** If $F_k^k(n) - F_k^k(n-1) = 1$, then $F_k(n+1) = F_k(n)$. In this case, the left inequality in (3.1) is clear from (3.2):

$$L_k(F_k(n+1)-1) = L_k(F_k(n)-1) < n < n+1.$$

From (3.3) and (3.4), we deduce that $L_k^k(F_k^k(n-1)) = n-1$. Hence, $x_k[n-1]x_k[n]$ starts a new τ_k^k block, and thus $x_k[n-1]x_k[n] = k1$. This means that the left inequality of (3.2) is actually strict here, because the letter 1 cannot start the word $\tau_k(x_k[F_k(n)])$. So finally

$$n < L_k(F_k(n)) = L_k(F_k(n+1)),$$

and we have finished the proof of (3.1) for $m = n+1$, $j = 1$, in all possible cases.

Case $j \geq 2$**.** It remains to prove (3.1) for $m = n+1$ and $j \geq 2$. Since $1 \leq F_k(n+1) \leq n$ by Proposition 2.1, the induction hypothesis holds for $m = F_k(n+1)$ and $j-1$. Hence

$$L_k^{j-1}(F_k^j(n+1)-1) \leq F_k(n+1) - 1 \leq F_k(n+1) \leq L_k^{j-1}(F_k^j(n+1)).$$

Applying L_k, which preserves these inequalities by monotonicity (Proposition 2.3), gives

$$L_k^j(F_k^j(n+1)-1) \leq L_k(F_k(n+1)-1) \leq L_k(F_k(n+1)) \leq L_k^j(F_k^j(n+1)).$$

Finally, we can use (3.1), which we proved for $m = n+1$ and $j = 1$ above, in the middle:

$$L_k^j(F_k^j(n+1)-1) \leq L_k(F_k(n+1)-1) < n+1 \leq L_k(F_k(n+1)) \leq L_k^j(F_k^j(n+1)).$$

Therefore, (3.1) indeed holds for $m = n+1$ and all $j \geq 1$. Thus, by induction, (3.1) holds for all $m \geq 1$, which concludes the proof of the theorem. $\square$

In particular, we have $F_k^j(L_k^j(n)) = n$ for all $k \geq 1$ and $j, n \geq 0$. Moreover, $L_k^j(n)$ is the largest element of $F_k^{-j}(\{n\})$ while $L_k^j(n-1)+1$ is the smallest one (for $n > 0$), and these extrema may coincide. In particular, this is always the case when $j = 0$ and quite frequently when $j = 1$ and $k > 1$; see Proposition 6.5 for a study of this ratio.

The relationship between F_k^j and L_k^j can also be formulated as a Galois connection [7].

Corollary 3.2. *For all $j \geq 0$ and $k \geq 1$, the functions F_k^j and L_k^j form a* Galois connection *between $\mathbb{N}$ and itself (with F_k^j as* left adjoint *and L_k^j as* right adjoint*). Indeed, for all $m, n \geq 0$ we have $F_k^j(n) \leq m$ if and only if $n \leq L_k^j(m)$. Moreover, this Galois connection is said to be a* Galois insertion *because $F_k^j \circ L_k^j = \mathrm{id}$.*

Proof. The case $n = 0$ is obvious. For $n \geq 1$, if $F_k^j(n) \leq m$, then $n \leq L_k^j(F_k^j(n)) \leq L_k^j(m)$ by Theorem 3.1 and the monotonicity of L_k^j. Conversely, if $n \leq L_k^j(m)$, then $F_k^j(n) \leq F_k^j(L_k^j(m))$ by monotonicity of L_k^j and $F_k^j(L_k^j(m)) = m$ as seen above. □

Note that Proposition 4.2 below gives a nice expression of $L_k(n)$ as $n + F_k^{k-1}(n)$. Hence

$$L_k(F_k(n)) = F_k(n) + F_k^k(n) = n + 1 - \partial F_k(n) \in \{n, n+1\}.$$

4 Counting letters

We now express the number of occurrences for the letters $1, \dots, k$ in prefixes of x_k. Thanks to Theorem 3.1, this relates them to functions F_k in various ways.

Let $C_k^{(P)}(n)$ denote the number of letters satisfying the predicate P in the prefix $x_k[0{:}n)$. More formally,

$$C_k^{(P)}(n) = \#\{0 \leq j < n \,:\, P(x_k[j])\}.$$

In particular, we use the expressions

- $C_k^{(=i)}(n)$ for counting the occurrences of a specific letter i,
- $C_k^{(>i)}(n) = C_k^{(=i+1)}(n) + \cdots + C_k^{(=k)}(n)$ for counting all letters strictly greater than i.

Proposition 4.1. *For all $k \geq 1$ and $n \geq 0$, we have*

$$F_k^{k-1}(n) = C_k^{(=k)}(n), \tag{4.1}$$

$$F_k^j(n) = C_k^{(>j)}(n) \qquad \text{for all } 0 \leq j < k, \tag{4.2}$$

$$F_k^{k+i-1}(n) = C_k^{(=i)}(n+i) \qquad \text{for all } 1 \leq i < k. \tag{4.3}$$

Proof. We can use the same counting technique to prove each of these three equations.

For Equation (4.1), we already mentioned that x_k can be seen as a succession of "blocks" $\tau_k^{k-1}(i) = k1 \cdots (i-1)$, each one containing k only as first letter. For a given $m \geq 0$, the letter $x_k[m]$ belongs to one of these blocks, say the p-th one (counting from $p = 1$ for the first block). We hence have p occurrences of the letter k in $x_k[0] \cdots x_k[m]$, so $C_k^{(=k)}(m+1) = p$. Also note that the first p blocks have a total length of $\left|\tau_k^{k-1}(x_k[0{:}p))\right| = L_k^{k-1}(p)$. So $L_k^{k-1}(p)$ is also the first index in the $(p+1)$-st block, and hence $L_k^{k-1}(p) > m$. Similarly $L_k^{k-1}(p-1)$ is the first index of the p-th block, so all in all

$$L_k^{k-1}(p-1) \leq m < L_k^{k-1}(p).$$

After substituting p and setting $n = m + 1$, we obtain that for all $n > 0$ we have

$$L_k^{k-1}(C_k^{(=k)}(n) - 1) < n \leq L_k^{k-1}(C_k^{(=k)}(n)).$$

By Theorem 3.1 and Proposition 2.4, this implies $F_k^{k-1}(n) = C_k^{(=k)}(n)$ for all $k, n \geq 1$. Moreover, this identity trivially holds for $n = 0$ as well.

For Equation (4.2), we generalize the previous counting technique. For $0 \leq j < k$ and a letter $1 \leq i \leq k$, the word $\tau_k^j(i)$ starts with exactly one letter strictly greater than j, while the rest of this word consists of letters less than or equal to j. Indeed, either $i + j \leq k$ and $\tau_k^j(i) = i + j > j$, or $i + j \geq k$ and $\tau_k^j(i) = \tau_k^{i+j-k}(k) = k1 \cdots (i + j - k)$ with $i + j - k \leq j$. Just as before, we deduce that

$$L_k^j(C_k^{(>j)}(n) - 1) < n \leq L_k^j(C_k^{(>j)}(n))$$

for all $0 \leq j < k$ and $n > 0$. As earlier, this allows establishing $F_k^j(n) = C_k^{(>j)}(n)$ for all $0 \leq j < k$ and $n > 0$. Once again, this identity also holds for $n \geq 0$.

Finally, we use yet another instance of the same technique for proving Equation (4.3). Consider $1 \leq i < k$. For all letters $1 \leq j \leq k$ the words $\tau_k^{k+i-1}(j)$ contain the letter i only at the $(i+1)$-st position. Indeed, these words can also be written $\tau_k^{j+i-1}(k)$ (since $\tau_k^{k-j}(j) = k$), so they all admit $\tau_k^i(k) = k1 \cdots i$ as a common prefix, possibly followed first by letters greater than i and then by new blocks no larger than $k1 \cdots (i-1)$. For similar reasons as before, we hence have

$$L_k^{k+i-1}(C_k^{(=i)}(n) - 1) < n - i \leq L_k^{k+i-1}(C_k^{(=i)}(n))$$

for all $n \geq i + 1$. So for all $1 \leq i < k$ and $n \geq 0$, we have $F_k^{k+i-1}(n) = C_k^{(=i)}(n + i)$. $\square$

Actually, Equation (4.1) can also be deduced from Equation (4.2) in the particular case $j = k - 1$, since $C_k^{(>k-1)} = C_k^{(=k)}$. Also note that Equation (4.3) could be extended to the case $i = 0$ if we replace the letter k by 0 in the substitution τ_k and its fixed point x_k.

We can now give an interesting alternative expression for $L_k(n)$ (which is the largest element of $F_k^{-1}(\{n\})$ by Theorem 3.1).

Proposition 4.2. *For all $k \geq 1$ and $n \geq 0$,*

$$L_k(n) = n + C_k^{(=k)}(n) = n + F_k^{k-1}(n).$$

Proof. We have $|\tau_k(k)| = 2$ while $|\tau_k(i)| = 1$ for the other letters $i \neq k$. Hence

$$L_k(n) = 2\,C_k^{(=k)}(n) + C_k^{(\neq k)}(n) = n + C_k^{(=k)}.$$

Finally, Equation (4.1) leads to the last equality. $\square$

By Equation (4.2), we can also express the number of occurrences of a specific letter $1 \leq j < k$ via the difference between $C_k^{(>j-1)}$ and $C_k^{(>j)}$. Hence for $1 \leq j < k$ and $n \geq 0$

$$F_k^{j-1}(n) - F_k^j(n) = C_k^{(>j-1)}(n) - C_k^{(>j)}(n) = C_k^{(=j)}(n). \tag{4.4}$$

In particular, we obtain the following proposition.

Proposition 4.3. *For $k > 1$ and $n \geq 0$,*

$$F_k(n) = n - C_k^{(=1)}(n) = C_k^{(\neq 1)}(n).$$

Hence $\partial F_k(n)$ is 0 if and only if $x_k[n] = 1$, and 1 otherwise.

Proof. Direct use of the previous equation in the particular case $j = 1$. Alternatively, one may use Equation (4.3) for $i = 1$, and then the recursive definition of F_k.

Afterwards, we get $F_k(n+1) - F_k(n) = 1 - (C_k^{(=1)}(n+1) - C_k^{(=1)}(n))$, which is 0 if and only if $x_k[n] = 1$ and 1 otherwise. □

This important link between ∂F_k and x_k helps us to transfer many properties from one to the other. In particular, ∂F_k cannot have two consecutive zeros, and it admits up to k consecutive ones but not $k + 1$.

Now let us describe the letters of x_k in terms of differences $\partial F_k^j(n)$. Recall from Proposition 2.1 that these differences are always either 0 or 1. Of course, $\partial F_k^0(n) = 1$ and $\partial F_k^j(0) = 1$. Now, for a non-zero n, we have $F_k^j(n) = 1$ when j is large enough, in particular for $j \geq n-1$. Indeed, on a non-zero argument, F_k either returns 1 or removes at least one from its argument, and this is iterated here j times. As a consequence, we always have $\partial F_k^n(n) = 0$ when $n > 0$. Moreover, if for some j, we have $\partial F_k^j(n) = 0$, then $\partial F_k^{j+1}(n) = 0$ as well. Conversely, if $\partial F_k^j(n) = 1$, then $\partial F_k^{j-1}(n) = 1$. Hence for all $k, n \geq 1$, the sequence $(\partial F_k^j(n))_{j \in \mathbb{N}}$ consists of a block of ones followed by an infinity of zeros. Actually, the letters of x_k indicate how deep to dive in these differences to find a first zero (or give up after $k - 1$ attempts).

Proposition 4.4. *Let $1 \leq j < k$, $n \geq 0$. We have $x_k[n] = j$ if and only if both $\partial F_k^{j-1}(n) = 1$ and $\partial F_k^j(n) = 0$. Moreover, for $k \geq 1$, we have $x_k[n] = k$ if and only if $\partial F_k^{k-1}(n) = 1$. (In the latter case, $\partial F_k^k(n)$ could be either 0 or 1.)*

Proof. First, this statement is obvious for $k = 1$. We now assume that $k > 1$. By subtracting Equation (4.4) for $n + 1$ and n, we obtain for each $1 \leq j < k$ that

$$\partial F_k^{j-1}(n) - \partial F_k^j(n) = C_k^{(=j)}(n+1) - C_k^{(=j)}(n).$$

This amounts to 1 if and only if $x_k[n] = j$, and 0 otherwise. So in particular for each $i < x_k[n]$ we have $\partial F_k^{i-1}(n) = \partial F_k^i(n)$. Meanwhile, we just noticed in Proposition 4.3 that $\partial F_k(n) = 0$ if and only if $x_k[n] = 1$. Write $\ell = x_k[n]$. Three situations may occur:

- Either $\ell = 1$, and we directly have $\partial F_k(n) = 0$ and $\partial F_k^0(n) = 1$.
- Or $1 < \ell < k$ and $\partial F_k(n) = 1$. We can propagate $\partial F_k^i(n) = 1$ for all $i < \ell$ and finish with $\partial F_k^{\ell-1}(n) - \partial F_k^\ell(n) = 1$. Hence $\partial F_k^\ell(n) = 0$.
- Lastly, if $\ell = k$, then the propagation $\partial F_k^i(n) = 1$ goes up to $i \leq k - 1$. Hence the desired statement follows. □

5 Related polynomials and algebraic integers

We now introduce two families of polynomials whose positive zeros appear as average slopes for F_k and L_k and letter frequencies for x_k (Section 6).

Definition 5.1. For $k \geq 1$, we define $P_k(X) = X^k + X - 1$ and $Q_k(X) = X^k - X^{k-1} - 1$. We let α_k (and β_k respectively) be the unique positive zero of P_k (and Q_k respectively). Note that $\beta_k = 1/\alpha_k$.

A more complete investigation of these polynomials can be found in Selmer [20] (where the polynomials $X^k \pm (X+1)$ lead to $P_k(-X)$) and Dilcher [4]. We now state the elementary results needed in the rest of our work. In particular, for $k \geq 1$, the polynomial P_k is strictly increasing on $\mathbb{R}_+$ and admits exactly one zero there. This zero α_k clearly satisfies $\frac{1}{2} \leq \alpha_k < 1$. Moreover the polynomial P_k is the negation of the reciprocal polynomial of Q_k, i.e., $P_k(X) = -X^k Q_k(1/X)$ (and vice versa). As a consequence, the zeros of P_k are the inverse of the zeros of Q_k and vice versa. In particular $\beta_k = 1/\alpha_k$ is indeed the unique zero of P_k in $\mathbb{R}_+$ and furthermore $1 < \beta_k \leq 2$.

Figure 5.1 gives approximate values for the first α_k and β_k. In particular, note that $\beta_1 = 2$ and β_2 is the golden ratio $\varphi = \frac{1+\sqrt{5}}{2}$. Thanks to the rational zero theorem, one can easily show that α_k and β_k are irrational for $k \geq 2$.

$$\begin{aligned}
\alpha_1 &= 0.5 & \beta_1 &= 2\\
\alpha_2 &= 0.6180339887498948\dots & \beta_2 &= 1.618033988749895\dots\\
\alpha_3 &= 0.6823278038280193\dots & \beta_3 &= 1.465571231876768\dots\\
\alpha_4 &= 0.7244919590005157\dots & \beta_4 &= 1.380277569097614\dots\\
\alpha_5 &= 0.7548776662466925\dots & \beta_5 &= 1.324717957244746\dots\\
\alpha_6 &= 0.7780895986786012\dots & \beta_6 &= 1.285199033245349\dots
\end{aligned}$$

Figure 5.1: The first α_k and β_k, positive zeros of $X^k + X - 1$ and $X^k - X^{k-1} - 1$.

Proposition 5.2. *The sequence $(\alpha_k)_{k\in\mathbb{N}_+}$ is strictly increasing while $(\beta_k)_{k\in\mathbb{N}_+}$ is strictly decreasing. Moreover, for $k \geq 1$, we have $1+\frac{1}{k} \leq \beta_k \leq 1+\frac{1}{\sqrt{k}}$ or equivalently $\sqrt{k} \leq \beta_k^{k-1} \leq k$. Hence the (β_k) sequence converges to 1, as well as the (α_k) sequence. Actually, the following asymptotic expressions hold for $k \to \infty$:*

$$\alpha_k = 1 - \frac{\ln k}{k} + o\left(\frac{\ln k}{k}\right), \tag{5.1}$$

$$\beta_k = 1 + \frac{\ln k}{k} + o\left(\frac{\ln k}{k}\right). \tag{5.2}$$

Proof. Let $k \geq 1$. Suppose $\alpha_{k+1} \leq \alpha_k$. Since α_k and α_{k+1} are in $(0,1)$, we would have $\alpha_{k+1}^{k+1} = \alpha_{k+1}^k \alpha_{k+1} < \alpha_{k+1}^k \leq \alpha_k^k$, leading to $0 = P_{k+1}(\alpha_{k+1}) < P_k(\alpha_k) = 0$, a contradiction. Hence (α_k) is strictly increasing and $(\beta_k) = (\alpha_k^{-1})$ is strictly decreasing.

Now $Q_k(\beta_k) = 0$ can be reformulated as $\beta_k^{k-1}(\beta_k - 1) = 1$ and hence

$$\alpha_k^{k-1} = \beta_k^{-(k-1)} = \beta_k - 1. \tag{5.3}$$

This provides the equivalence between $\sqrt{k} \leq \beta_k^{k-1} \leq k$ and $1 + \frac{1}{k} \leq \beta_k \leq 1 + \frac{1}{\sqrt{k}}$.

For proving the lower bound $\sqrt{k} \leq \beta_k^{k-1}$, it is sufficient to note that

$$1 + \beta_k + \cdots + \beta_k^{k-1} = \frac{\beta_k^k - 1}{\beta_k - 1} = \frac{\beta_k^{k-1}}{\beta_k^{-(k-1)}} = \beta_k^{2(k-1)}. \tag{5.4}$$

Each term on the left of Equation (5.4) is 1 or larger, so $k \leq \beta_k^{2(k-1)}$, hence the desired lower bound. For the upper bound $\beta_k^{k-1} \leq k$, we divide by β_k^{k-1} in Equation (5.4) and obtain

$$\beta_k^{k-1} = \frac{\beta_k^{2(k-1)}}{\beta_k^{k-1}} = \frac{1}{\beta_k^{k-1}} + \cdots + \frac{\beta_k^{k-1}}{\beta_k^{k-1}} = \alpha_k^{k-1} + \cdots + \alpha_k + 1 \leq k.$$

The two asymptotic expressions (5.1) and (5.2) are equivalent: considering the multiplicative inverse of any of these expressions leads to the other one. An approximation related to (5.1) appears in Selmer [20, Eq. 4.4] but without further explanation, while (5.2) is a direct consequence of Dilcher [4, Lem. 3]. Alternatively, we give here a simple direct proof of (5.1). For $k \geq 1$, we consider the function $h_k(x) = \sqrt[k]{1-x}$ and note that α_k is a fixed point of h_k. It could be proved that iterating h_k on any initial point in $(0, 1)$ converges to α_k, but here it suffices to use a few initial terms of such a sequence, namely $u_0 = 1 - 1/e$ and $u_1 = h_k(u_0)$ and $u_2 = h_k(u_1)$ and $u_3 = h_k(u_2)$. For these values, it is straightforward to compute

$$\begin{aligned} u_1 &= e^{-1/k} = 1 - \frac{1}{k} + o\left(\frac{1}{k}\right), \\ u_2 &= 1 - \frac{\ln k}{k} + o\left(\frac{1}{k}\right), \\ u_3 &= 1 - \frac{\ln k}{k} + \frac{\ln \ln k}{k} + o\left(\frac{1}{k}\right). \end{aligned}$$

Meanwhile, a basic study of the function h_k ensures that $u_0 < u_2 < \alpha_k < u_3 < u_1$ when $k \geq 3$. The desired expression (5.1) follows by weakening the previous asymptotic expressions for both u_2 and u_3 to $1 - \ln k/k + o(\ln k/k)$.

$\square$

6 Infinitary behavior

For each $k \geq 1$, the substitution τ_k is a *primitive* morphism [1]; i.e., there exists an exponent $p \geq 1$ such that for all letters $1 \leq i, j \leq k$, the letter i occurs in $\tau_k^p(j)$. Here, the first adequate exponent is $p = 2k-2$. Indeed, $\tau_k^{2k-2}(j)$ admits $\tau_k^{k-1}(k) = k1\cdots(k-1)$ as a prefix because $\tau_k^{k-1}(j)$ has k as first letter. Since x_k is the fixed point of a primitive morphism, each letter i, $1 \le i \le k$, has a *frequency*, which we denote by $\mathrm{freq}_k(i)$. More precisely, the limit $\mathrm{freq}_k(i) = \lim_{n\to\infty} \frac{1}{n} C_k^{(=i)}(n)$ exists, we have $0 < \mathrm{freq}_k(i) \le 1$, and there exists a formula for computing it [2, Thm. 8.4.7]. Here, an easier approach for computing $\mathrm{freq}_k(i)$ is to consider F_k.

Theorem 6.1. *For $j \ge 0$ and $k \ge 1$, the following limits exist and have the given values, where α_k and β_k come from Definition 5.1:*

$$\begin{aligned}
\lim_{n\to\infty} \tfrac{1}{n} F_k^j(n) &= \alpha_k^j, \\
\lim_{n\to\infty} \tfrac{1}{n} L_k^j(n) &= \beta_k^j, \\
\mathrm{freq}_k(i) &= \alpha_k^{k+i-1}, && \text{for } 1 \le i < k, \\
\mathrm{freq}_k(k) &= \alpha_k^{k-1} = \beta_k - 1.
\end{aligned}$$

Proof. Recall from Proposition 4.3 that for $k > 1$ and $n \ge 0$, we have $F_k(n) = n - C_k^{(=1)}(n)$. Hence the limit $\lim_{n\to\infty} \frac{1}{n} F_k(n)$ exists as well[6] and is $1 - \mathrm{freq}_k(1)$. In the case $k = 1$, we have $F_1(n) = \lceil n/2 \rceil$. Hence $\lim_{n\to\infty} \frac{1}{n} F_1(n) = \frac{1}{2}$.

Now that $\frac{1}{n} F_k(n)$ is known to converge to some finite limit ℓ, computing this limit is quite straightforward. Indeed, the recursive equation of F_k can be reformulated as

$$\frac{F_k(n)}{n} = 1 - \frac{F_k(F_k^{k-1}(n-1))}{F_k^{k-1}(n-1)} \cdots \frac{F_k(n-1)}{n-1}\,\frac{n-1}{n}$$

for $n > 1$. Each fraction can be shown to converge to ℓ, except $\frac{n-1}{n}$, which tends to 1. Hence $\ell = 1 - \ell^k$ and obviously ℓ is a nonnegative real number. Hence $\ell = \alpha_k$. As a consequence, for $k \ge 2$, the frequency $\mathrm{freq}_k(1)$ is $1 - \alpha_k = \alpha_k^k$ (and 1 when $k = 1$). The same telescoping technique gives $\lim_{n\to\infty} \frac{1}{n} F_k^j(n) = \alpha_k^j$.

Concerning L_k^j, a consequence of Theorem 3.1 is $F_k^j(L_k^j(n)) = n$. Hence

$$\frac{L_k^j(n)}{n} = \left(\frac{F_k^j(L_k^j(n))}{L_k^j(n)}\right)^{-1}.$$

Since $L_k^j(n) \ge n$, this implies that $\frac{1}{n} L_k^j(n)$ converges, and its limit is $\alpha_k^{-j} = \beta_k^j$.

For the frequency of the letter k, Equation (4.1) implies that $\mathrm{freq}_k(k) = \alpha_k^{k-1}$, which is also $\beta_k - 1$ by Equation (5.3). And for the frequencies of the other letters $1 \le i < k$, one may exploit either Equation (4.4) or Equation (4.3). For instance, the former leads to

$$\mathrm{freq}_k(i) = \alpha_k^{i-1} - \alpha_k^i = \alpha_k^{i-1}(1 - \alpha_k) = \alpha_k^{k+i-1}.$$

In particular, this subsumes the case $i = 1$ seen earlier. Finally, one may check that the sum of all these frequencies, from α_k^{k-1} (letter k) to α_k^{2k-2} (letter $k-1$), is of course 1. $\square$

[6]Surprisingly, we have not found any obvious methods for proving this convergence of $\frac{1}{n} F_k(n)$ directly from the recursive definition of F_k, without using this "detour" via words.

Definition 6.2. For two functions $f, g : \mathbb{N} \to \mathbb{N}$, we say that f is *ultimately smaller* than g and write $f <_\infty g$ when there exists N such that $f(n) < g(n)$ for all $n \geq N$.

Corollary 6.3. *When $\alpha_k^j < \alpha_{k'}^{j'}$ for some $k, k' \geq 1$ and $j, j' \geq 0$, then $F_k^j <_\infty F_{k'}^{j'}$.*

Proof. Theorem 6.1 gives $\lim_{n\to\infty} \frac{1}{n} F_k^j(n) = \alpha_k^j < \alpha_{k'}^{j'} = \lim_{n\to\infty} \frac{1}{n} F_{k'}^{j'}(n)$. For n large enough, both sides become close enough to their limits, so there must exists $N \geq 1$ such that $n \geq N$ implies $\frac{1}{n} F_k^j(n) < \frac{1}{n} F_{k'}^{j'}(n)$ and hence $F_k^j(n) < F_{k'}^{j'}(n)$. ◻

Corollary 6.4. *For $k \geq 1$, we have $F_k <_\infty F_{k+1}$, $F_k^k >_\infty F_{k+1}^{k+1}$ and $F_k^{k+1} >_\infty F_{k+1}^{k+2}$. More generally, for $j \geq 0$ and $k \geq 1$:*

(a) we have $F_k^{k+j} >_\infty F_{k+1}^{k+j+1}$ when $\alpha_k \geq \frac{j}{j+1}$ (hence in particular when $k \geq j^2$);

(b) we have $F_k^{k+j} <_\infty F_{k+1}^{k+j+1}$ when $\alpha_{k+1} \leq \frac{j}{j+1}$ (hence in particular when $1 \leq k < j$).

Proof. All these facts are obtained by the previous corollary; we just have to compare the corresponding average slopes. For $F_k <_\infty F_{k+1}$, Proposition 5.2 directly gives $\alpha_k < \alpha_{k+1}$. For $F_k^k >_\infty F_{k+1}^{k+1}$, we have $\alpha_k^k = 1 - \alpha_k > 1 - \alpha_{k+1} = \alpha_{k+1}^{k+1}$. Now, for $F_k^{k+1} >_\infty F_{k+1}^{k+2}$, we have

$$\alpha_k^{k+1} - \alpha_{k+1}^{k+2} = (1-\alpha_k)\alpha_k - (1-\alpha_{k+1})\alpha_{k+1} = (\alpha_{k+1} - \alpha_k)(\alpha_k + \alpha_{k+1} - 1) > 0$$

since $\frac{1}{2} \leq \alpha_k < \alpha_{k+1}$. More generally, let $j \geq 0$. As before,

$$\alpha_k^{k+j} - \alpha_{k+1}^{k+j+1} = (1-\alpha_k)\alpha_k^j - (1-\alpha_{k+1})\alpha_{k+1}^j.$$

The function $(1-X)X^j$ is strictly increasing between 0 and $\frac{j}{j+1}$ and strictly decreasing afterwards. Since we always have $\alpha_k < \alpha_{k+1}$, then $\alpha_k^{k+j} - \alpha_{k+1}^{k+j+1} > 0$ at least when $\alpha_k \geq \frac{j}{j+1}$ or equivalently when $\beta_k \leq 1 + \frac{1}{j}$. Thanks to the bounds in Proposition 5.2, this happens at least when $\sqrt{k} \geq j$, i.e., $k \geq j^2$. Conversely, $\alpha_k^{k+j} - \alpha_{k+1}^{k+j+1} < 0$ at least when $\alpha_{k+1} \leq \frac{j}{j+1}$, for which a sufficient condition is $1 \leq k < j$, still thanks to the bounds in Proposition 5.2. ◻

Corollary 6.4 is to be compared with the results of the next section, for instance Theorem 7.4: The latter gives only non-strict inequalities, but for all $n \geq 0$, while here we have proved strict inequalities, but only when n is larger than some bounds. Moreover, we do not have explicit estimates for these bounds for the moment.

As a related matter, we can estimate the ratio of numbers having unique preimages by F_k.

Proposition 6.5. *For $k \geq 1$, $n \geq 0$, let us call $U_k(n) = \#\{0 \leq j < n : |F_k^{-1}(\{j\})| = 1\}$. Then, for $n > 0$, we have*

$$U_k(n) = 2n - 1 - L_k(n-1) = n - F_k^{k-1}(n-1),$$

and hence

$$\lim_{n\to\infty} \tfrac{1}{n} U_k(n) = 1 - \alpha_k^{k-1} = 2 - \beta_k.$$

Proof. Let $n > 0$. We already know from Theorem 3.1 and Proposition 4.2 that the largest element of $F_k^{-1}(\{n-1\})$ is $L_k(n-1) = n-1+F_k^{k-1}(n-1)$. There are $U_k(n)$ numbers between 0 and $n-1$ with a unique preimage, and the other $n-U_k(n)$ numbers have exactly two preimages; see the discussion after Proposition 4.3. By counting the cardinality of $F_k^{-1}(\{0,\dots,n-1\})$, we obtain

$$U_k(n) + 2(n - U_k(n)) = 1 + L_k(n-1) = n + F_k^{k-1}(n-1).$$

Combined with Theorem 6.1, this leads to the desired equations and limits. □

In particular, for $k = 1$ and $n > 0$ we get $U_1(n) = 1$. Indeed, since $F_1(n) = \lceil \frac{n}{2} \rceil$, the only number with a unique preimage by F_1 is 0. Then $\frac{1}{n}U_k(n)$ tends to $0.3819\dots$ for $k = 2$, to $0.5344\dots$ for $k = 3$, and to $0.6197\dots$ for $k = 4$, and these limits tend to 1 as k grows.

7 Monotonicity over the parameter k

This section studies the monotonicity of L_k and F_k when the parameter k varies. In all this section, we compare functions via *pointwise order*: an inequality such as $F_k \le F_{k+1}$ means that $F_k(n) \le F_{k+1}(n)$ *for all* points $n \ge 0$. As such, the results here are quite stronger than the ones of the previous section about the infinitary behavior, i.e., when n is large enough.

First, we state a nice duality between functions F_k^j and L_k^j with respect to this pointwise order, extending the results of Section 3. Thanks to this, all the following technical lemmas about L_k have immediate counterparts for F_k.

Proposition 7.1. *For $j, j' \ge 0$ and $k, k' \ge 1$, we have $L_k^j \le L_{k'}^{j'}$ if and only if $F_{k'}^{j'} \le F_k^j$. Furthermore, we can be more precise concerning the relative positions where these inequalities occur. For $n \ge 0$, we have*

(a) $L_k^j(n) \le L_{k'}^{j'}(n)$ if and only if $F_{k'}^{j'}(m) \le F_k^j(m)$, with $m = L_k^j(n)$,

(b) $L_k^j(n) < L_{k'}^{j'}(n)$ if and only if $F_{k'}^{j'}(m) < F_k^j(m)$, with $m = L_{k'}^{j'}(n)$,

(c) $L_k^j(m) \le L_{k'}^{j'}(m)$ for $m = F_k^j(n)$ implies that $F_{k'}^{j'}(n) \le F_k^j(n)$,

(d) $F_k^j(n) < F_{k'}^{j'}(n)$ implies that $L_{k'}^{j'}(m) < L_k^j(m)$ for $m = F_{k'}^{j'}(n)$.

Proof. First, the fact that $L_k^j \le L_{k'}^{j'}$ implies $F_{k'}^{j'} \le F_k^j$ is a consequence of point (c), while point (a) implies the other direction. Also note that (b) is a contrapositive version of (a), and the same for (d) and (c).

For proving point (a), let $n \ge 0$. Then $L_k^j(n) \le L_{k'}^{j'}(n)$ is equivalent to $F_{k'}^{j'}(L_k^j(n)) \le n$ by Corollary 3.2, and we have $n = F_k^j(L_k^j(n))$ by Theorem 3.1.

For point (c), we use that $F_{k'}^{j'}(n) \le F_k^j(n)$ is equivalent to $n \le L_{k'}^{j'}(F_k^j(n))$ by Corollary 3.2. By Theorem 3.1 (or trivially when $n = 0$), we have $n \le L_k^j(F_k^j(n))$, and thus $n \le L_{k'}^{j'}(F_k^j(n))$ when $L_k^j(F_k^j(n)) \le L_{k'}^{j'}(F_k^j(n))$. □

Let us now study the ordering of the functions L_k^j. Note first that

$$|\tau_k^k(i)| = i + 1 = |\tau_k^{k-1}(i)| + 1 \quad \text{for all } 1 \leq i \leq k.$$

For all $k \geq 1$, $n \geq 0$, this implies that $L_k^k(n) = L_k^{k-1}(n) + n$ and thus

$$L_{k+1}^{k+1}(n) - L_k^k(n) = L_{k+1}^k(n) - L_k^{k-1}(n). \tag{7.1}$$

One of our main results is that the sequence of functions (L_k) is monotonic, which we prove here by mutual induction with another property comparing iterations L_k^j and L_{k+1}^{j+1}.

Theorem 7.2. *For all $k, n \geq 1$, $0 \leq j \leq k$, we have*

$$L_k(n) \geq L_{k+1}(n), \quad \text{i.e., } C_k^{(=k)}(n) \geq C_{k+1}^{(=k+1)}(n), \tag{7.2}$$

$$L_k^j(n) < L_{k+1}^{j+1}(n). \tag{7.3}$$

Proof. Let $k \geq 1$. Since $x_k[0] = k$, $x_{k+1}[0] = k+1$, and $|\tau_k^j(k)| = j+1 < j+2 = |\tau_{k+1}^{j+1}(k+1)|$ for $0 \leq j \leq k$, the inequalities (7.2) and (7.3) hold for $n = 1$, $0 \leq j \leq k$.

Let $m \geq 2$ and assume that (7.2) and (7.3) hold for all $1 \leq n < m$, $0 \leq j \leq k$.

We first prove (7.2) at m, i.e., $L_k(m) \geq L_{k+1}(m)$, or equivalently $F_k^{k-1}(m) \geq F_{k+1}^k(m)$; see Proposition 4.2. Let us abbreviate $F_k^{k-1}(m)$ as c and $F_{k+1}^k(m)$ as c' and prove $c' \leq c$ i.e., $c' - 1 < c$. By Proposition 2.3, L_{k+1}^k is strictly increasing, and hence it is sufficient to prove $L_{k+1}^k(c'-1) < L_{k+1}^k(c)$. Indeed, we have

$$L_{k+1}^k(c'-1) < m \leq L_k^{k-1}(c) \leq L_{k+1}^k(c), \tag{7.4}$$

where the left and middle inequalities come from Theorem 3.1. To obtain the right inequality in (7.4), we distinguish between the cases $k = 1$ and $k \geq 2$. When $k = 1$, then $c = m$ and hence $L_k^{k-1}(c) = m < L_2(m) = L_{k+1}^k(c)$ by Proposition 2.3 (f). When $k \geq 2$, we can use (7.3) for $n = c$, $j = k-1$ because in this case $1 \leq c < m$ by Proposition 2.1.

Now let $0 \leq h \leq k$ and let us prove (7.3) for $n = m$ and $j = h$. Thanks to Equation (7.1), the case $j = k$ is implied by the case $j = k-1$, so now we can freely assume $h < k$. If $x_{k+1}[m-1] = k+1$, then (7.3) holds for $n = m$ because it holds for $n = m-1$ and $|\tau_k^h(i)| \leq |\tau_k^h(k)| < |\tau_{k+1}^{h+1}(k+1)|$ for all $1 \leq i \leq k$. If $x_{k+1}[m-1] \neq k+1$, then $x_{k+1}[0{:}m) = \tau_{k+1}(x_{k+1}[0{:}\ell))$ for some $\ell \geq 1$. Proposition 2.3 indicates that $\ell < L_{k+1}(\ell) = m$. Hence (7.2) holds for $n = \ell$, and thus

$$m = L_{k+1}(\ell) \leq L_k(\ell).$$

We apply L_k^h (which is monotonic by Proposition 2.3) on this inequality, and then (7.3) for $n = \ell$ and $j = h+1$, to obtain

$$L_k^h(m) \leq L_k^h(L_k(\ell)) = L_k^{h+1}(\ell) < L_{k+1}^{h+2}(\ell) = L_{k+1}^{h+1}(L_{k+1}(\ell)) = L_{k+1}^{h+1}(m).$$

Therefore, (7.3) holds for $n = m$, $0 \leq j \leq k$. By induction, (7.2) and (7.3) hold for all $n \geq 1$, $0 \leq j \leq k$. $\square$

Corollary 7.3. *For all $k \geq 1$ and $j \geq 0$, we have $L_k^j \geq L_{k+1}^j$.*

Proof. We proceed by induction on j. The case $j = 0$ is obvious. The case $j = 1$ is given by Equation (7.2), trivially extended to $n = 0$. Now assume $L_k^j \geq L_{k+1}^j$ for some $j \geq 0$. For $n \geq 0$, we hence have

$$L_k^{j+1}(n) = L_k^j(L_k(n)) \geq L_{k+1}^j(L_k(n)) \geq L_{k+1}^j(L_{k+1}(n)) = L_{k+1}^{j+2}(n),$$

thanks to the induction hypothesis for j and then the monotonicity of L_{k+1}^j, combined with the statement for $j = 1$. We can hence conclude by induction. □

At last, we obtain the following monotonicity of the function sequence (F_k) over the parameter k, where the special case $j = 1$ is our main result Theorem 1.1.

Theorem 7.4. *For all $j \geq 0$, $k \geq 1$, we have $F_k^j \leq F_{k+1}^j$.*

Proof. Consequence of Corollary 7.3 and Proposition 7.1. □

Comparing F_k^j with F_{k+1}^{j+1} instead of F_{k+1}^j, we obtain the opposite inequality.

Theorem 7.5. *For all $k \geq 1$ and $0 \leq j \leq k$, we have $F_k^j \geq F_{k+1}^{j+1}$.*

Proof. Let $k \geq 1$ and $0 \leq j \leq k$. By Proposition 7.1, the statement to prove is equivalent to $L_{k+1}^{j+1} \leq L_k^j$, which is a direct consequence of Equation (7.3) and of $L_{k+1}^{j+1}(0) = 0 = L_k^j(0)$. □

To sum up, the functions $(F_k^j, \leq)$ with their pointwise ordering form (at least) a nice lattice generated by the following basic cells for $1 \leq j \leq k$:

$$\begin{tikzcd}
F_{k-1}^{j-1} \arrow[r, "\leq"] & F_k^{j-1} & \\
& F_k^j \arrow[ul, "\leq"'] \arrow[u, "\leq"'] \arrow[r, "\leq"'] & F_{k+1}^j \arrow[ul, "\leq"']
\end{tikzcd}$$

In such a cell, the vertical edge $F_k^j \leq F_k^{j-1}$ is obvious for sub-linear functions such as F_k. Moreover it is also a double consequence of $F_k^j \leq F_{k-1}^{j-1} \leq F_k^{j-1}$ and $F_k^j \leq F_{k+1}^j \leq F_k^{j-1}$. Also note that the remaining unrelated functions F_{k-1}^{j-1} and F_{k+1}^j may actually be incomparable. For instance $F_3^3(5) = 2 > 1 = F_5^4(5)$, while $F_3^3(9) = 3 < 4 = F_5^4(5)$. Even if we only retain the infinitary behavior (as in Section 6), the ordering of these functions may vary. The currently known situation is presented in Figure 7.1, where the edges such as $G \longrightarrow H$ mean $G \leq H$. The row $j = 0$ of identity functions has been omitted here, id being trivially above all other F_k^j. The numbers displayed in blue alongside the nodes F_k^j are approximations of their average slopes $\lim \frac{1}{n} F_k^j(n) = \alpha_k^j$. If one of the functions is pointwise below another, their average slopes are ordered accordingly.[7] Also note that the diagonal $j = k$ and first

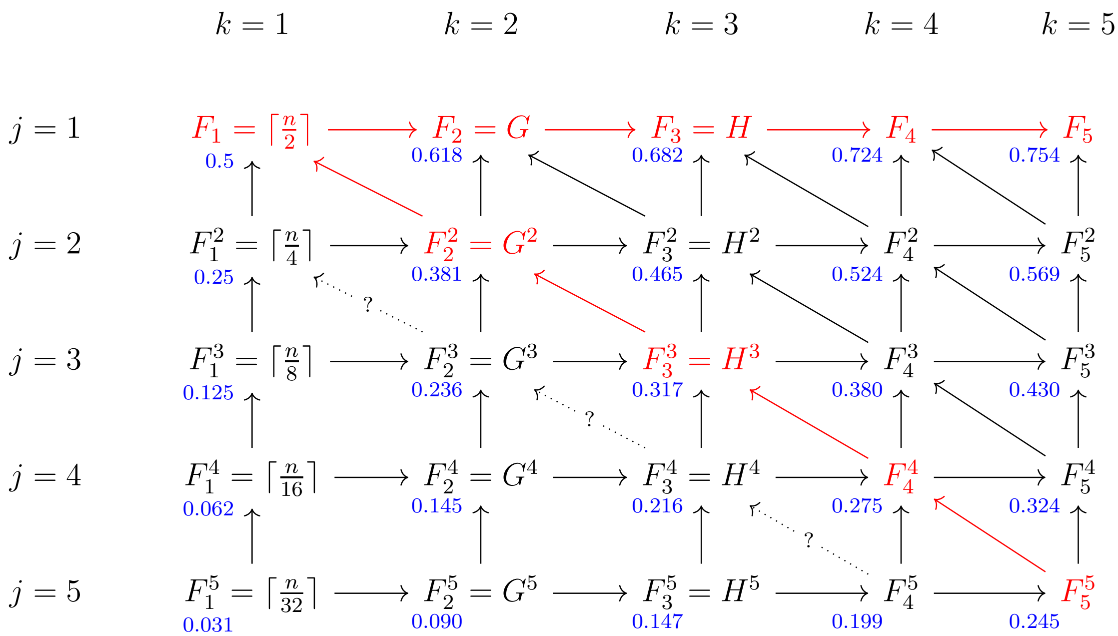

Figure 7.1: The known $(F_k^j, \leq)$ lattice, displayed here for $1 \leq j, k \leq 5$.

row $j = 1$ (in red in the figure) form an interesting chain of inequalities, with slopes ranging symmetrically between 0 and 1 (since $1 - \alpha_k^k = \alpha_k$).

In Figure 7.1, some dotted edges with question marks indicate conjectured inequalities. Indeed, we conjecture that $F_k^{k+1} \geq F_{k+1}^{k+2}$ for all $k \geq 1$, i.e., that Theorem 7.5 may be extended to the case $j = k+1$. Actually, Corollary 6.4 already proved that $F_k^{k+1}(n) > F_{k+1}^{k+2}(n)$ for sufficiently large n. Thanks to Proposition 7.1, this conjecture can also be equivalently formulated as $L_{k+1}^{k+2} \geq L_k^{k+1}$. Said otherwise, the inequality (7.3) appears to still hold as a non-strict inequality in the case $j = k+1$. Note in this case that we may indeed reach equality: we can prove $L_k^{k+1}(k+1) = L_{k+1}^{k+2}(k+1)$.

After that, for $j \geq k+2$, we prove below that $F_k^j \not\geq F_{k+1}^{j+1}$. More precisely, when $j \geq k+2$, two behaviors seem possible: either $k+2 \leq j \leq 2k$, and we prove below that F_k^j and F_{k+1}^{j+1} are incomparable in this case; or $j > 2k$, in which case we conjecture that $F_k^j \leq F_{k+1}^{j+1}$. For studying these questions, we now focus on $L_{k+1}^{j+1}(1) - L_k^j(1)$.

Lemma 7.6. *For all $k \geq 1$, we have*

$$L_{k+1}^{j+1}(1) - L_k^j(1) = \begin{cases} 1, & \textit{if } 0 \leq j \leq 2k; \\ -\frac{(j-2k-1)(j-2k+2)}{2}, & \textit{if } 2k \leq j \leq 3k. \end{cases}$$

In particular $L_{k+1}^{2k+2}(1) = L_k^{2k+1}(1)$. Moreover for all $j \geq 2k+2$, we have $L_{k+1}^{j+1}(1) < L_k^j(1)$.

[7]Conversely, strictly ordered average slopes only give clues about infinitary behavior, the functions may well be pointwise incomparable due to early values.

Proof. From Proposition 2.3, we have

$$L_k^j(1) = \begin{cases} j+1, & \text{if } 0 \le j \le k; \\ L_k^{j-1}(1) + L_k^{j-k}(1), & \text{if } j \ge k. \end{cases}$$

For $0 \le j \le k$, we have $L_{k+1}^j(1) = j+1 = L_k^j(1)$ and $L_{k+1}^{j+1}(1) = j+2 = L_k^j(1)+1$. For $k < j \le 2k$, we have

$$\begin{aligned} L_{k+1}^{j+1}(1) - L_k^j(1) &= L_{k+1}^j(1) - L_k^{j-1}(1) = 1, \\ L_{k+1}^j(1) - L_k^j(1) &= L_{k+1}^{j-1}(1) - L_k^{j-1}(1) - 1 = k - j, \end{aligned}$$

by induction on j. Note that $-\frac{(j-2k-1)(j-2k+2)}{2} = 1$ for $j = 2k$. For $2k < j \le 3k$, we have thus

$$L_{k+1}^{j+1}(1) - L_k^j(1) = -\frac{(j-2k-2)(j-2k+1)}{2} + 2k - j = -\frac{(j-2k-1)(j-2k+2)}{2},$$

by induction on j.

Still for $k \ge 1$, we now prove by induction on j that $L_{k+1}^{j+1}(1) < L_k^j(1)$ for all $j \ge 2k+2$. This is true for $j = 2k+2$: indeed, either $k = 1$ and we directly compute $L_2^5(1) = 13 < 16 = L_1^4(1)$, or $k \ge 2$ and hence $2k+2 = j \le 3k$ so $L_{k+1}^{j+1}(1) - L_k^j(1)$ is given by the previous formula, which is strictly negative here. Now, for the step case of the induction, let $j > 2k+2$. Then

$$L_{k+1}^{j+1}(1) - L_k^j(1) = (L_{k+1}^j(1) - L_k^{j-1}(1)) + (L_{k+1}^{j-k}(1) - L_k^{j-k}(1)).$$

The induction hypothesis on $j-1$ indicates that the central difference above is strictly negative, while the rightmost difference is nonpositive thanks to Corollary 7.3, allowing us to conclude this induction. ◻

Lemma 7.7. *Let $k \ge 1$. If $j \ge k+2$, the following value n_j satisfies $L_k^j(n_j) > L_{k+1}^{j+1}(n_j)$:*

$$n_j = \begin{cases} 2k+3-j, & \text{if } \; k+2 \le j \le 2k+2; \\ 1, & \text{if } \; 2k+2 \le j. \end{cases}$$

Proof. The case $j \ge 2k+2$ where $n_j = 1$ is a direct use of Lemma 7.6. Now suppose $k+2 \le j \le 2k+2$. By Proposition 2.3, we know that $L_k^p(1) = p+1$ for all $0 \le p \le k$. By considering $p = n_j - 1$, we obtain $n_j = L_k^{n_j-1}(1)$ and hence

$$L_k^j(n_j) = L_k^j(L_k^{n_j-1}(1)) = L_k^{j+n_j-1}(1) = L_k^{2k+2}(1).$$

Similarly, $n_j = L_{k+1}^{n_j-1}(1)$ and $L_{k+1}^{j+1}(n_j) = L_{k+1}^{2k+3}(1)$, which is strictly less than $L_k^{2k+2}(1)$ by Lemma 7.6. ◻

In particular, this last lemma implies that $L_k^j \not\leq L_{k+1}^{j+1}$ when $j \geq k+2$, with n_j as counterexample. Thanks to Proposition 7.1, this means equivalently that $F_k^j \not\geq F_{k+1}^{j+1}$ when $j \geq k+2$, with $n = L_k^j(n_j)$ as counterexample. Moreover, when $k+2 \leq j \leq 2k$, Lemma 7.6 implies that $L_k^j \not\geq L_{k+1}^{j+1}$, with $n = 1$ as counterexample, and hence that L_k^j and L_{k+1}^{j+1} are incomparable, and equivalently that F_k^j and F_{k+1}^{j+1} are also incomparable.

We conclude this section with a last conjecture about F_k, as always for $k \geq 1$. From Corollary 6.4, we know that the inequality $F_k(n) \leq F_{k+1}(n)$ becomes strict when n is large enough. Actually, we conjecture an explicit bound $N_k = \frac{1}{2}(k+1)(k+6)$, for which $F_k(n) < F_{k+1}(n)$ as soon as $n > N_k$. At least, it can be proved that $F_k(N_k) = F_{k+1}(N_k)$, so the bound cannot be less than N_k, but it remains to be confirmed that no equality occurs after N_k. These constants N_k can also be expressed as $\frac{1}{2}(k+3)(k+4) - 3$ and satisfy $N_{k+1} = N_k + (k+4)$. In particular $N_1 = 7$ and $N_2 = 12$ and $N_3 = 18$. Interestingly, we also have $L_{k+1}(N_k) = L_{k+2}(N_k)$. Finally, this conjecture implies two other interesting statements:

- $F_k(n) < F_{k+1}(n+1)$ for all $n \geq 2$;
- $L_{k+1}(n) > L_{k+2}(n)$ for all $n > N_k$.

8 More letter counting

Several results and conjectures of the previous section can be rephrased into statements about the number of occurrences of certain letters in words x_k. In particular, let us consider again the letter 1 and study $C_k^{(=1)}$.

Proposition 8.1. *For all $k \geq 1$ and $n \geq 0$, we have $C_k^{(=1)}(n) \geq C_{k+1}^{(=1)}(n)$.*

Proof. When $k \geq 2$, this is a consequence of Proposition 4.3 and Theorem 1.1. For $k = 1$, we have $C_1^{(=1)}(n) = n \geq C_2^{(=1)}(n)$. □

Considering the letter 2, we conjecture that $C_k^{(=2)}(n) \geq C_{k+1}^{(=2)}(n)$ for all $k \geq 3$, $n \geq 0$. In particular, this is a consequence of the conjecture $F_k^{k+1} \geq F_{k+1}^{k+2}$ mentioned in the previous section: when combining it with Equation (4.3), we get

$$C_k^{(=2)}(n) \geq C_{k+1}^{(=2)}(n) \quad \text{for all } k > 2 \text{ and } n \geq 2;$$

we can then relax the condition on n because all these quantities are zero when n is 0 or 1. (In the latter case, $x_k[0] = k \neq 2$ and similarly for x_{k+1}.)

Actually, the property $C_k^{(=2)} \geq C_{k+1}^{(=2)}$ also extends to $k = 2$, and is easy to prove in this case, since for all $n \geq 0$, we have $C_2^{(=2)}(n) \geq C_3^{(=3)}(n)$ (by Equation 7.2, trivially extended to $n = 0$) as well as $C_3^{(=3)}(n) \geq C_3^{(=2)}(n)$; indeed, in x_3, any occurrence of the letter 2 is in a subword 312. We cannot extend further: for $k = 1$, there is no letter 2 in the word x_1, hence $C_1^{(=2)}(n) = 0$ while $C_2^{(=2)}(n) = F_2(n)$ by Equation 4.1, which differs from 0 as soon as $n \geq 1$.

For the letters $3 \leq i < k$, there is no pointwise monotonicity between $C_k^{(=i)}$ and $C_{k+1}^{(=i)}$:

Proposition 8.2. *For $3 \leq i < k$, we have*

- $C_k^{(=i)}(n) < C_{k+1}^{(=i)}(n)$ *when* $n = i + L_k^{2k+2}(1)$,
- $C_k^{(=i)}(n) > C_{k+1}^{(=i)}(n)$ *when* $n = i + L_{k+1}^{k+i}(1)$.

Proof. After Lemma 7.7, we noticed that F_k^j and F_{k+1}^{j+1} are incomparable for all $j \geq k+2$. Some counterexamples are $L_k^j(n_j)$ in one direction and $L_{k+1}^{j+1}(1)$ in the other (by Proposition 7.1). Now we use Equation (4.3) to express this in terms of the number of occurrences of letters, by choosing $j = k+i-1$. When $3 \leq i < k$, we indeed have $j \geq k+2$ (and also $j \leq 2k+2$). Due to the shape of Equation (4.3), the previous counterexamples are now shifted by i. Moreover, here

$$n_j = 2k+3-j = k+4-i = L_k^{k+3-i}(1),$$

and hence

$$L_k^j(n_j) = L_k^{k+i-1}(L_k^{k+3-i}(1)) = L_k^{2k+2}(1). \qquad \square$$

For instance, for $k = 5$ and $i = 4$, we have $C_5^{(=4)}(49) = 5 < 6 = C_6^{(=4)}(49)$ while $C_5^{(=4)}(20) = 2 > 1 = C_6^{(=4)}(20)$.

Finally, we compare these quantities for sufficiently large n. Let $k \geq 1$ and $1 \leq i < k$. By Corollary 6.4 and Equation (4.3), we obtain that $C_k^{(=i)} >_\infty C_{k+1}^{(=i)}$ at least when $\alpha_k \geq 1 - \frac{1}{i}$, which in particular happens when $(i-1)^2 \leq k$. Note that this condition is always satisfied when $i = 1$ or $i = 2$. Otherwise, for $i \geq 3$, small values of k may exhibit the opposite infinitary behavior. For example $C_6^{(=5)} <_\infty C_7^{(=5)}$ since

$$\lim_{n\to\infty} \tfrac{1}{n} C_6^{(=5)}(n) = \alpha_6^{10} \approx 0.0813, \quad \lim_{n\to\infty} \tfrac{1}{n} C_7^{(=5)}(n) = \alpha_7^{11} \approx 0.0819.$$

9 Acknowledgments

The authors are deeply grateful to Yining Hu, who made this joint work possible. We also thank the anonymous referees for their helpful remarks.

The third author was supported by the ERC grant DynAMiCs (101167561) of the European Research Council, the bilateral grant SYMDYNAR (ANR-23-CE40-0024 and FWF I 6750) of the Agence Nationale de la Recherche and the Austrian Science Fund, and by the ANR project IZES (ANR-22-CE40-0011).

References

[1] J.-P. Allouche, J. Cassaigne, J. Shallit, and L. Q. Zamboni, A taxonomy of morphic sequences, 2017. Arxiv preprint arXiv:1711.10807 [cs.FL], November 29 2017. Available at arXiv:1711.10807.

[2] J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, 2003.

[3] M. Dekking, On Hofstadter's G-Sequence, *J. Integer Sequences* **26** (2023), Article 23.9.2.

[4] K. Dilcher, On a class of iterative recurrence relations, In G. E. Bergum, A. N. Philippou, and A. F. Horadam, editors, *Applications of Fibonacci Numbers*, Vol. 5, pp. 143–158. Springer, 1993.

[5] P. J. Downey and R. E. Griswold, On a family of nested recursions, *Fibonacci Quart.* **22** (1984), 310–317.

[6] L. Ericksen and P. G. Anderson, Patterns in differences between rows in k-Zeckendorf arrays, *Fibonacci Quart.* **50** (2012), 11–18.

[7] M. Erné, J. Koslowski, A. Melton, and G. E. Strecker, A primer on Galois connections, *Ann. N. Y. Acad. Sci.* **704** (1993), 103–125.

[8] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, Vol. 1794 of *Lecture Notes in Mathematics*, Springer, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[9] C. Frougny, Z. Masáková, and E. Pelantová, Complexity of infinite words associated with beta-expansions, *RAIRO Theor. Informatics Appl.* **38** (2004), 163–185. Corrigendum, **38** (2004), 269–271.

[10] D. Gault and M. Clint, "Curiouser and curiouser" said Alice. Further reflections on an interesting recursive function, *Int. J. Comput. Math.* **26** (1988), 35–43.

[11] L. Goldberg and A. S. Fraenkel, Patterns in the generalized Fibonacci word, applied to games, *Discrete Math.* **341** (2018), 1675–1687.

[12] K. Hare and J. C. Saunders, Generalised Fibonacci sequences constructed from balanced words, *J. Number Theory* **231** (2022), 349–377.

[13] D. R. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books, 1979.

[14] C. Kimberling, The Zeckendorf array equals the Wythoff array, *Fibonacci Quart.* **33** (1995), 3–8.

[15] P. Letouzey, Coq/Rocq proofs about Hofstadter's function G. 2015–2025. `https://github.com/letouzey/hofstadter_g`.

[16] P. Letouzey, Hofstadter's problem for curious readers, 2015. Research Report, Université Paris Diderot and INRIA Paris. hal-01195587v4.

[17] D. S. Meek and G. H. J. Van Rees, The solution of an iterated recurrence, *Fibonacci Quart.* **22** (1984), 101–104.

[18] J. L. Ramírez, G. N. Rubiano, and R. De Castro, A generalization of the Fibonacci word fractal and the Fibonacci snowflake, *Theoret. Comput. Sci.* **528** (2014), 40–56.

[19] The Rocq Core Team, The Rocq proof assistant. 2025–2026 (and 1985–2025 under the name Coq). `https://rocq-prover.org`.

[20] E. Selmer, On the irreducibility of certain trinomials, *Math. Scand.* **4** (1956), 287–302.

[21] J. Shallit, Proving properties of some greedily-defined integer recurrences via automata theory, *Theoret. Comput. Sci.* **988** (2024), 114363.

[22] N. J. A. Sloane and The OEIS Foundation Inc., The on-line encyclopedia of integer sequences, 2026. Published electronically at `https://oeis.org`.

[23] B. Tan and Z.-Y. Wen, Some properties of the Tribonacci sequence, *European J. Combin.* **28** (2007), 1703–1719.

2020 *Mathematics Subject Classification*: Primary 11B37; Secondary 11B39, 68R15, 68V15.
Keywords: Hofstadter sequence, Fibonacci word, theorem prover, proof assistant, Rocq.

(Concerned with sequences A005206, A005374, A005375, A005376, and A100721.)

Received September 27 2024; revised versions received April 17 2026; May 15 2026. Published in *Journal of Integer Sequences*, May 19 2026.

Return to Journal of Integer Sequences home page.