Walking on Words

Ian Pratt-Hartmann
Department of Computer Science, University of Manchester, United Kingdom
Instytut Informatyki, Uniwersytet Opolski, Opole, Poland

Abstract

Any function f with domain $\{1, \ldots, m\}$ and co-domain $\{1, \ldots, n\}$ induces a natural map from words of length n to those of length m: the ith letter of the output word ($1 \leq i \leq m$) is given by the $f(i)$th letter of the input word. We study this map in the case where f is a surjection satisfying the condition $|f(i+1) - f(i)| \leq 1$ for $1 \leq i < m$. Intuitively, we think of f as describing a "walk" on a word u, visiting every position, and yielding a word w as the sequence of letters encountered *en route*. If such an f exists, we say that u *generates* w. Call a word *primitive* if it is not generated by any word shorter than itself. We show that every word has, up to reversal, a unique primitive generator. Observing that, if a word contains a non-trivial palindrome, it can generate the same word via essentially different walks, we obtain conditions under which, for a chosen pair of walks f and g, those walks yield the same word when applied to a given primitive word. Although the original impulse for studying primitive generators comes from their application to decision procedures in logic, we end, by way of further motivation, with an analysis of the primitive generators for certain word sequences defined via morphisms.

2012 ACM Subject Classification Mathematics of computing → Combinatorics on words

Keywords and phrases word combinatorics, palindrome, Rauzy morphism

Digital Object Identifier 10.4230/LIPIcs.CPM.2024.25

Related Version *Full Version*: `https://arxiv.org/abs/2208.08913`

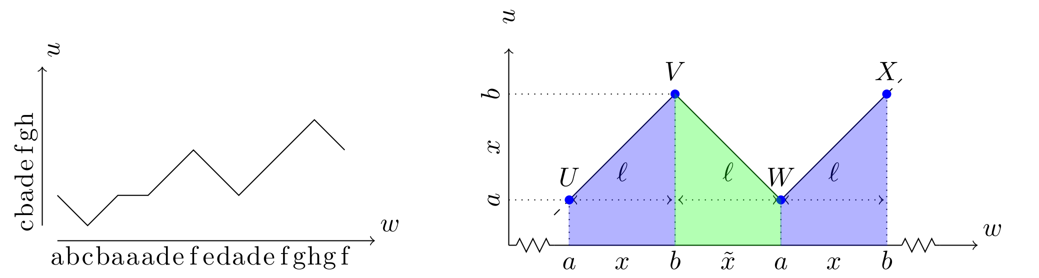
Funding This work was supported by the Polish NCN, grant number 2018/31/B/ST6/03662.

Acknowledgements The author wishes to thank Prof. V. Berthé and Prof. L. Tendera for their valuable help, and Mr. D. Kojelis for his many suggestions, in particular the much-improved reformulation of Theorem 4.

1 Introduction

Take any word over some alphabet, and, if it is non-empty, go to any letter in that word. Now repeat the following any number of times (possibly zero): scan the current letter, and print it out; then either remain at the current letter, or move one letter to the left (if possible) or move one letter to the right (if possible). In effect, we are *going for a walk on* the input word. Since any unvisited prefix or suffix of the input word cannot influence the word we print out, they may as well be ablated; letting u be the factor of the input word comprising the scanned letters, and w the word printed out, we say that u *generates* w. It is obvious that every word generates itself and its reversal, and that all other words it generates are strictly longer than itself. We ask about the converse of generation. Given a word w, what words u generate it? Call a word *primitive* if it is not generated by any word shorter than itself. It is easy to see that every word must have a generator that is itself primitive. We show that this *primitive generator* is in fact unique up to reversal. On the other hand, while primitive generators are unique, the generating walks need not be, and this leads us to ask, for a given pair of walks, whether we can characterize those primitive words u for which they output the same word w. We answer this question in terms of the palindromes contained in u. Specifically, for a primitive word u, the locations and lengths of the palindromes it

© Ian Pratt-Hartmann;
licensed under Creative Commons License CC-BY 4.0
35th Annual Symposium on Combinatorial Pattern Matching (CPM 2024).
Editors: Shunsuke Inenaga and Simon J. Puglisi; Article No. 25; pp. 25:1–25:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



(a) Example of generation. **(b)** The minimal leg $J = [V, W]$ (of length ℓ) of some walk.

Figure 1 Generation and minimal legs.

contains determine which pairs of walks yield identical outputs on u. As an illustration of the naturalness of the notion of primitive generator, we consider word sequences over the alphabet $\{1, \ldots, k\}$ generated by the generalized Rauzy morphism σ, which maps the letter k to the word 1, and any other letter i $(1 \le i < k)$ to the two-letter word $1 \cdot (i{+}1)$. Setting $\alpha_1^{(k)} = 1$ and $\alpha_{n+1}^{(k)} = \sigma(\alpha_n^{(k)})$ for all $n \ge 1$, we obtain the word sequence $\{\alpha_n^{(k)}\}_{n \ge 1}$. We show that every word in this sequence from the kth onwards has the same primitive generator.

2 Principal results

Fix some alphabet Σ. We use $a, b, c \ldots$ to stand for letters of Σ, and $u, v, w, \ldots$ to stand for words over Σ. The concatenation of two words u and v is denoted uv, or sometimes, for clarity, $u \cdot v$. For integers i, k we write $[i, k]$ to mean the set $\{j \in \mathbb{Z} \mid i \le j \le k\}$. If $u = a_1 \cdots a_n$ is a (possibly empty) word over Σ, and $f \colon [1, m] \to [1, n]$ is a function, we write u^f to denote the word $a_{f(1)} \cdots a_{f(m)}$ of length m. We think of f as telling us where in the word u we should be at each time point in the interval $[1, m]$. Define a *walk* to be a surjection $f \colon [1, m] \to [1, n]$ satisfying $|f(i{+}1) - f(i)| \le 1$ for all i $(1 \le i < m)$. These conditions ensure that, as we move through the letters $a_{f(1)} \cdots a_{f(m)}$, we never change our position in u by more than one letter at a time, and we visit every position of u at least once. If $w = u^f$ for f a walk, we say that u *generates* w. We may picture a walk as a piecewise linear function, with the generat*ed* word superimposed on the abscissa and the generat*ing* word on the ordinate. Fig. 1a shows how u = cbadefgh generates w = abcbaaadefedadefghgf.

If $u = a_1 \cdots a_n$ is a word, denote the length of u by $|u| = n$, and the reversal of u by $\tilde{u} = a_n \cdots a_1$. Generation is evidently reflexive and reverse-reflexive: every word generates both itself and its reversal. Moreover, if u generates w, then $|u| \le |w|$; in fact, u and $\tilde{u}$ are the only words of length $|u|$ generated by u. We call u *primitive* if it is not generated by any word shorter than itself – equivalently, if it is generated only by itself and its reversal. For example, *babcd* and *abcbcd* are not primitive, because they are generated by *abcd*; but *abcbda* is primitive. Since the composition of two walks is a walk, generation is transitive: if u generates v and v generates w, then u generates w. Define a *primitive generator* of w to be a primitive word that generates w. It follows easily from the above remarks that every word w has some primitive generator u, and indeed, $\tilde{u}$ as well, since the reversal of a primitive generator of w is obviously also a primitive generator of w. The principal result of this paper is that there are no others:

▶ **Theorem 1.** *The primitive generator of any word is unique up to reversal.*

As an immediate consequence, if u is the primitive generator of w, and v generates w, then u generates v. Theorem 1 is relatively surprising: let u and v be primitive words. Now suppose we go for a walk on u and, independently, go for a walk on v; recalling the stipulation that the two walks visit every position in the words they apply to, the theorem says that, provided only that $u \neq v$ and $u \neq \tilde{v}$, there is no possibility of coordinating these walks so that the sequences of visited letters are the same.

A *palindrome* is a word u such that $u = \tilde{u}$; a *non-trivial* palindrome is one of length at least 2. If u is a non-trivial palindrome, then it is not primitive. Indeed, if $|u|$ is even, then u has a double letter in the middle, and so is certainly not primitive (it is generated by the word in which one of the occurrences of the doubled letter is deleted); if $|u|$ is odd, then it is generated by its prefix of length $(|u|+1)/2 < |u|$ (start at the beginning, go just over half way, then return to the start). Call a word *uniliteral* if it is of the form a^n for some letter a and some $n \geq 0$. Note that the empty word ϵ counts as uniliteral.

▶ **Corollary 2.** *Every uniliteral word has precisely one primitive generator; all others have precisely two.*

Proof. By Theorem 1, if w is any word, its primitive generators are of the form u and $\tilde{u}$ for some word u. The first statement of the corollary is obvious: if $w = \epsilon$ then $u = \tilde{u} = \epsilon$; and if $u = a^n$ for some n ($n \geq 1$), then $u = \tilde{u} = a$. If w is not uniliteral, then $|u| > 1$. But since non-trivial palindromes cannot be primitive, $u \neq \tilde{u}$. ◀

Yet another way of stating Theorem 1 is to say that, for any fixed word w, the equation $u^f = w$ has exactly one primitive solution for u, up to reversal. The same is not true, however, of solutions for f, even if we fix the choice of primitive generator (either u or $\tilde{u}$). Indeed, $u = abcbd$ is one of the two primitive generators of $w = abcbcbd$, but we have $u^f = w$ for $f\colon [1,7] \to [1,5]$ given by either of the courses of values $[1,2,3,4,3,4,5]$ or $[1,2,3,2,3,4,5]$. Let u be a primitive word. Say that u is *perfect* if $u^f = u^g$ implies $f = g$ for any walks f and g on u. Thus, *abcbd* is primitive but not perfect. On the other hand, it is easy to characterize those primitive words that are perfect:

▶ **Theorem 3.** *Let u be a word. Then u is perfect if and only if it contains no non-trivial palindrome as a factor.*

Theorem 3 tells us that generating walks are uniquely determined as long as the primitive generator u does not contain a non-trivial palindrome, but gives us little information if u does contain a non-trivial palindrome. In that case, we would like a characterization of *which* pairs of walks on u yield identical words. We answer this question in terms of the *positions* of the palindromes contained in u. Let $u = a_1 \cdots a_n$ be a word. We denote the ith letter of u by $u[i] = a_i$, and the factor of u from the ith to jth letters by $u[i,j] = a_i \cdots a_j$. If $u[i,j]$ is a non-trivial palindrome, call the ordered pair $\langle i, j \rangle$ a *defect* of u, and denote the set of defects of u by Δ_u. Regarding Δ_u as a binary relation on the set $[1,n]$, we write Δ_u^* for its equivalence closure, the smallest reflexive, symmetric and transitive relation including Δ_u. The interplay between defects and walks is then summed up in the following theorem.

▶ **Theorem 4.** *Let u be a primitive word of length n, and f, g walks with domain $[1,m]$ and co-domain $[1,n]$. Then $u^f = u^g$ if and only if $\langle f(i), g(i) \rangle \in \Delta_u^*$ for all $i \in [1,m]$.*

The motivation for the study of primitive generators comes from the study of the decision problem in (fragments of) first-order logic, in presentations where the logical variables are taken to be $x_1, x_2, \ldots$, and all signatures are assumed to be purely relational. Call a first-order formula φ *index-normal* if, for any quantified sub-formula $Qx_k\psi$ of φ, ψ is a Boolean combination of formulas that are either atomic with free variables among $x_1, \ldots, x_k$, or have as their major connective a quantifier binding x_{k+1}. By re-indexing variables, any first-order formula can easily be written as a logically equivalent index-normal formula. We call an index-normal formula *adjacent* if, in any atomic sub-formula, the indices of neighbouring arguments never differ by more than 1. For example, an atomic sub-formula $p(x_4, x_4, x_5, x_4, x_3)$ is allowed, but an atomic sub-formula $q(x_1, x_3)$ is not. It was shown in [1] that the problem of determining validity for adjacent formulas is decidable. A key notion in analysing this fragment is that of an *adjacent type*. Let $\mathfrak{A}$ be a structure interpreting some relational signature, and $\bar{a}$ a tuple of elements from its domain, A. Define the *adjacent type of* $\bar{a}$ (*in* $\mathfrak{A}$) to be the set of all adjacent atomic formulas $q(\bar{x})$ satisfied by $\bar{a}$ in $\mathfrak{A}$. If we think now of $\bar{a}$ as a word over the (possibly infinite) alphabet A, it can easily be shown that the adjacent type of $\bar{a}$ is determined by the adjacent type of its primitive generator. Thus, models of formulas can be unambiguously constructed by specifying only the adjacent types of *primitive* tuples, a crucial technique in establishing decidability of satisfiability.

Notwithstanding its logical genealogy, the concept of primitive generator may be of interest in its own right within the field of string combinatorics. For illustration, consider the sequences of words $\{\alpha_n^{(k)}\}_{n\geq 1}$ over the alphabet $\{1, \ldots, k\}$, defined by setting $\alpha_1^{(k)} = 1$ and $\alpha_{n+1}^{(k)} = \sigma(\alpha_n^{(k)})$, where $\sigma\colon \{1, \ldots, k\}^* \to \{1, \ldots, k\}^*$ is the monoid endomorphism given by

$$\sigma(i) = \begin{cases} 1 \cdot (i+1) & \text{if } i < k \\ 1 & \text{if } i = k. \end{cases}$$

(Here, the operator $\cdot$ represents string concatenation, not integer multiplication!) For $k = 2$, we obtain the so-called *Fibonacci word sequence* 1, 12, 121, 12112, $\ldots$; for $k = 3$, we obtain the *tribonacci word sequence* 1, 12, 1213, 1213121, $\ldots$; and so on. A simple induction shows that, for all $k \geq 2$ and all $n > k$, $\alpha_n^{(k)} = \alpha_{n-1}^{(k)}\alpha_{n-2}^{(k)} \cdots \alpha_{n-k}^{(k)}$. In other words, each element of the sequence $\{\alpha_n^{(k)}\}_{n\geq 1}$ after the kth is the concatenation, in reverse order, of the previous k elements; for this reason, the word sequence obtained is referred to as the k*-bonacci word sequence*. A simple proof also shows that $\alpha_n^{(k)}$ is always a left-prefix of $\alpha_{n+1}^{(k)}$, so that we may speak of the infinite word $\omega^{(k)}$ defined by taking the limit $\lim_{n\to\infty} \alpha_n^{(k)}$ in the obvious sense. Thus, the infinite word $\omega^{(2)} = 12112\cdots$ is the (infinite) *Fibonacci word*, and $\omega^{(3)} = 1213121\cdots$ the (infinite) *tribonacci word*. The Fibonacci word is an example of a *Sturmian word* (see, e.g. [3, Ch. 6] for an extensive treatment). The morphism yielding the tribonacci word is sometimes called the *Rauzy morphism* [6, p. 149] (see also [4, Secs. 10.7 and 10.8]). Intriguingly, for a fixed k, all but the first k elements of $\{\alpha_n^{(k)}\}_{n\geq 1}$ share the same primitive generator:

▶ **Theorem 5.** *For all $k \geq 2$, there exists a word γ_k such that, for all $n \geq k$, γ_k is the primitive generator of $\alpha_n^{(k)}$.*

The proof of Theorem 5 exploits a feature of the words $\alpha_n^{(k)}$ that is obvious when one computes a few examples: they are riddled with palindromes. As one might then expect in view of Theorem 4, for all k and all $n \geq k$, the primitive generator γ_k generates $\alpha_n^{(k)}$ via many different walks – in fact via walks beginning at any position of γ_k occupied by the letter 1.

3 Uniqueness of primitive generators

The following terminology will be useful. (Refer to Fig. 1a for motivation.) Let $f\colon [1,m] \to [1,n]$ be a walk, with $m > 1$. By a *leg* of f, we mean a maximal interval $[i,j] \subseteq [1,m]$ such that, for h in the range $i \leq h < j$, the difference $d = f(h{+}1){-}f(h)$ is constant. We speak of a *descending*, *flat* or *ascending* leg, depending on whether d is -1, 0 or 1. The *length* of the leg is $j{-}i$. A leg $[i,j]$ is *initial* if $i = 1$, *final* if $j = m$, *terminal* if it is either initial or final, and *internal* if it is not terminal. A number h which forms the boundary between two consecutive legs will be called a *waypoint*. We count the numbers 1 and m as waypoints by courtesy, and refer to them as *terminal waypoints*; all other waypoints are *internal*. Thus, a walk consists of a sequence of legs from one waypoint to another. If h is an internal waypoint where the change is from an increasing to a decreasing leg, we call h a *peak*; if the change is from a decreasing to an increasing leg, we call it a *trough*. Not all waypoints need be peaks or troughs, because some legs may be flat; however, it is these waypoints that will chiefly concern us in the sequel.

▶ **Lemma 6.** *A word u is not primitive if and only if it is of any of the following forms, where a, b are letters and x, y, z are words:* (i) $xaay$*,* (ii) $b\tilde{x}axby$*,* (iii) $yb\tilde{x}axb$ *or* (iv) $yaxb\tilde{x}axbz$*.*

Proof. Straightforward: see full version [5]. ◀

In the sequel, we shall primarily employ the if-direction of Lemma 6. It easily follows from Cases (i) and (ii) of Lemma 6 that, over the alphabet $\{1,2\}$, there are exactly five primitive words: ϵ, 1, 2, 12, and 21. However, over any larger alphabet, there are infinitely many. For example, over the alphabet $\{1,2,3\}$, the set of primitive words is easily seen to be given by the regular expression $[(\epsilon+3+23)(123)^*(\epsilon+1+12)]+[(\epsilon+2+32)(132)^*(\epsilon+1+13)]$. Over alphabets of any finite size, the set of primitive words is context-sensitive. This follows from the fact that the four patterns of Lemma 6 define context-sensitive languages, together with the standard Boolean closure properties of context-sensitive languages.

We shall occasionally need to consider a broader class of functions than walks. Define a *stroll* to be a function $f\colon [1,m] \to [1,n]$ satisfying $|f(i{+}1){-}f(i)| \leq 1$ for all i $(1 \leq i < m)$. Thus, a walk is a stroll which is surjective. Let $f\colon [1,m] \to [1,n]$ be a stroll. If $f(i) = f(j)$ for some i, j $(1 < i < j < m)$ define the function $f'\colon [1, m{-}j{+}i] \to [1,n]$ by setting $f'(h) = f(h)$ if $1 \leq h \leq i$, and $f'(h) = f(h{+}j{-}i)$ otherwise. Intuitively, f' is just like f, but with the interval $[i, j{-}1]$ – equivalently, the interval $[i{+}1, j]$ – removed. Evidently, f' is a also a stroll, and we denote it by $f/[i,j]$. For the cases $i = 1$ or $j = m$, we change the definition slightly, as no analogue of the condition $f(i) = f(j)$ is required. Specifically if $1 \leq i < j \leq m$, define the functions $f'\colon [1, m{-}j+1] \to [1,n]$ and $f''\colon [1,i] \to [1,n]$ by $f'(h) = f(j{+}h{-}1)$ and $f''(h) = f(h)$. Intuitively, f' is just like f, but with the interval $[1, j{-}1]$ removed, and f'' is just like f, but with the interval $[i{+}1, m]$ removed. Again f' and f'' are also strolls, and we denote them by $f/[1,j]$ and $f/[i,m]$, respectively.

With these preliminaries behind us, we give an outline sketch of the proof Theorem 1. The proof proceeds by contradiction, supposing that u and v are primitive words such that neither $u = v$ nor $u = \tilde{v}$, and w is a word generated from u by some walk f and from v by some walk g. Write $|w| = m$. Crucially, we may assume without loss of generality that w is a *shortest counterexample* – that is, a shortest word for which such u, v, f and g exist. Observe that, since u and v are primitive, they feature no immediately repeated letter. So suppose w does – i.e. is of the form $w = xaay$ for some words x, y and letter a. Letting $i = |x|{+}1$, we must therefore have $f(i) = f(i{+}1)$ and $g(i) = g(i{+}1)$. Now let $f' = f/[i,i{+}1]$, $g' = g/[i,i{+}1]$ and $w' = w[1,i] \cdot w[i{+}2,m]$. We see that f' is surjective if f is, and similarly for g', and moreover that $w' = u^{f'} = v^{g'}$, contrary to the assumption that w is shortest. Hence w contains no immediately repeated letters, whence all legs of f and g are either increasing or decreasing, and all internal waypoints are either peaks or troughs.

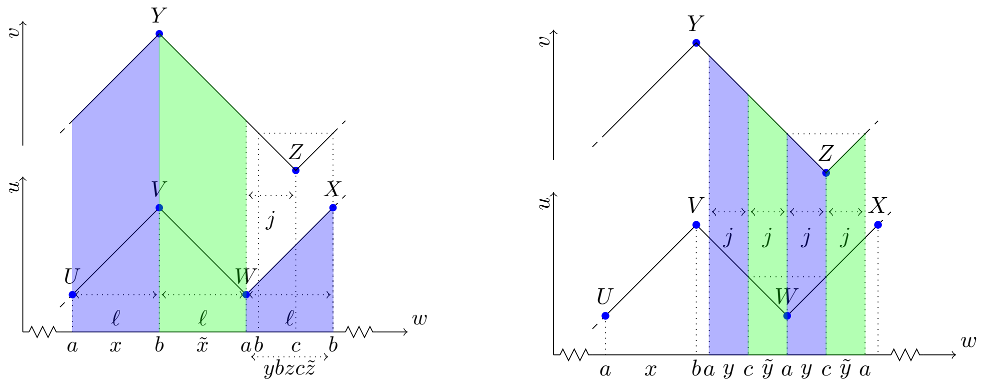
We claim first that at least one of f or g must have an internal waypoint. For if not, we have $w = u$ or $w = \tilde{u}$ and $w = v$ or $w = \tilde{v}$, whence $u = v$ or $u = \tilde{v}$, contrary to assumption. It then follows that *both f and g* have an internal waypoint. For suppose f has an internal waypoint (either a peak or a trough); then w is not primitive. But if g does not have an internal waypoint, $w = v$ or $w = \tilde{v}$, contrary to the assumption that v is primitive.

We use upper case letters in the sequel to denote integers in the range $[1, n]$ which are somehow significant for the walks f or g: note that these need not be waypoints. Let ℓ denote the minimal length of a leg on either of the walks f or g. Without loss of generality, we may take this minimum to be achieved on a leg of f, say $[V, W]$.

We suppose for the present that this leg is *internal*. Fig. 1b illustrates this situation where V is a peak and W a trough; but nothing essential would change if it were the other way around. Write $U = V - \ell$ and $X = W + \ell$. By the minimality of $[V, W]$ (assumed internal), $U \geq 1$ and $X \leq m$; moreover, f is monotone on $[U, V]$, $[V, W]$ and $[W, X]$. Now let $w[U] = a$, $w[V] = b$ and $w[U+1, V-1] = x$. Since V is a waypoint on f, $w[W] = a$ and $w[V+1, W-1] = \tilde{x}$. Similarly, $w[X] = b$ and $w[W+1, X-1] = \tilde{\tilde{x}} = x$. We see immediately that g must have a waypoint in the interval $[U+1, X-1]$, for otherwise, v (or $\tilde{v}$) contains a factor $axb\tilde{x}axb$, contrary to the assumption that v is primitive (Lemma 6, case (iv)). Let Y be the waypoint on g which is closest to either of V or W. Replacing w by its reversal if necessary, assume that $|Y-V| \leq |Y-W|$, and write $k = |Y-V|$. We consider possible values of $k \in [0, \ell-1]$ in turn, deriving a contradiction in each case.

Case (i): $k = 0$ (i.e. $Y = V$). For definiteness, let us suppose that Y is a peak, rather than a trough, but the reasoning is entirely unaffected by this determination. By the minimality of the leg $[V, W]$, g has no other waypoints in the interval $[U+1, W-1]$, and $g(U) = g(W)$. By inspection of Fig. 1b, it is also clear from the minimality of the leg $[V, W]$ that $f' = f/[U, W]$ is surjective (and hence a walk). We see immediately that the stroll $g' = g/[U, W]$ is not surjective. Indeed, if it were, writing $w' = w[1, U] \cdot w[W+1, n]$, we would have $w' = u^{f'} = v^{g'}$, contrary to the assumption that w is a shortest counterexample. In other words, there are positions of v which g reaches over the range $[U+1, W-1]$) that it does not reach outside this range. It follows that the position $g(V) = g(Y)$ in the string v is actually terminal. (Since we are assuming that Y is a peak, $g(Y) = |v|$; but the following reasoning is unaffected if Y is a trough and $g(Y) = 1$.) It also follows that W itself cannot be a waypoint of g. For otherwise, the leg following W, which is of length at least ℓ, covers all values in $g([U, W])$, thus ensuring that g' is surjective, which we have just shown to be false. However, g must have some waypoint in $[V+1, X-1]$. For if not, then g is decreasing between V and X (remember that $g(Y) = g(V) = |v|$), and thus v has a suffix $b\tilde{x}axb$, contrary to the assumption that v is primitive (Lemma 6 case (iii)). By the minimality of the leg $[V, W]$, we see that there is exactly one such waypoint, say Z. Since we have already shown that Y is the only waypoint on g in $[U+1, W-1]$, and that W is not a waypoint on g, it follows that $Z \in [W+1, X-1]$.

Now let $j = Z - W$. (Thus, $1 \leq j < \ell$.) If $j > \frac{1}{2}\ell$, we obtain the situation depicted in Fig. 2a. Since g has a waypoint at Z and remembering that $w[W+1, X-1] = x$ and $w[X] = b$, we see that x has the form $ybzc\tilde{z}$ for some strings y and z and some letter $c = w[Z]$. But we also know that $g(V) = g(Y) = |v|$, the final position of v, so that v has a suffix $xb = (ybzc\tilde{z})b$, and hence the suffix $bzc\tilde{z}b$, contrary to the assumption that v is primitive (Lemma 6 case (iii)). Furthermore, if $j = \frac{1}{2}\ell$, then, by the same reasoning, x has



(a) The condition $j = Z - W > \frac{1}{2}\ell$.

(b) The condition $j = Z - W < \frac{1}{2}\ell$.

Figure 2 The walk g has waypoints at $Y = V$ and at Z.

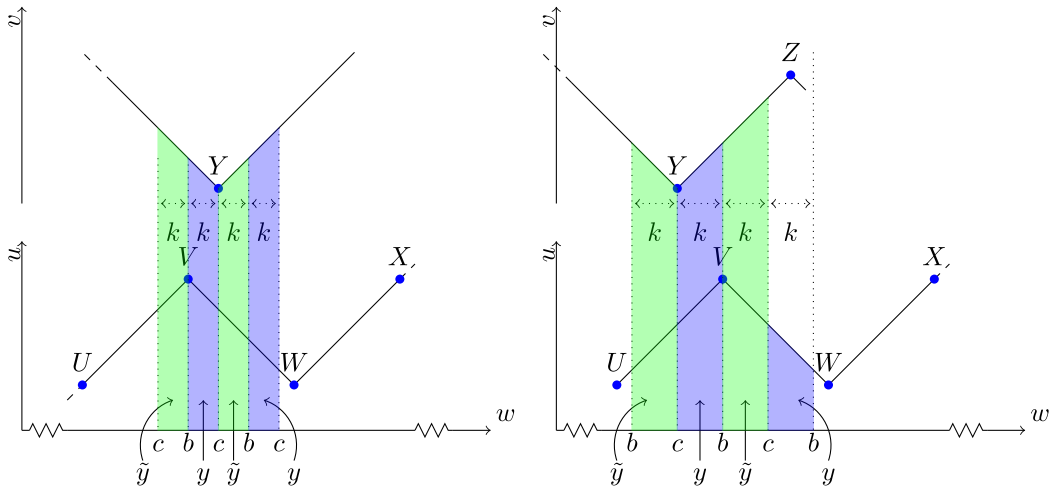
the form $zc\tilde{z}$ and $a = b$. Again then, v has a suffix $bzc\tilde{z}b$, contrary to the assumption that v is primitive. We conclude that $j < \frac{1}{2}\ell$, and we obtain the situation depicted in Fig. 2b. Now let $c = w[Z]$ and $y = w[W{+}1, Z{-}1]$. By considering the waypoint Z on g, we see that $w[Z, Z{+}j] = c\tilde{y}a$, whence $w[W, W{+}2j] = ayc\tilde{y}a$. By considering the waypoint W on f, we see that also $w[W{-}2j, W] = ayc\tilde{y}a$, whence $w[W{-}2j, W{+}j] = ayc\tilde{y}ayc$. But there are no waypoints of g strictly between $V = Y < W{-}2j$ and Z, whence $\tilde{v}$ contains the factor $ayc\tilde{y}ayc$, contrary to the supposition that v is primitive (Lemma 6 case (iv)).

Case (ii): $1 \leq k \leq \frac{1}{3}\ell$. We may have either $Y > V$ or $Y < V$: Fig. 3a shows the former case; however, the reasoning in the latter is almost identical. Let $w[V] = b$ and $w[Y] = c$. Furthermore, let $w[V{+}1, Y - 1] = y$. Since V is a waypoint of f, we have $w[V{-}k] = c$ and $w[V{-}k{+}1, V{-}1] = \tilde{y}$, whence $w[Y{-}2k, Y] = w[V{-}k, Y] = c\tilde{y}byc$. Since Y is a waypoint of g, we have $w[Y, Y{+}2k] = c\tilde{y}byc$, whence $w[V, V{+}3k] = byc\tilde{y}byc$. And since $\ell \geq 3k$, there is no waypoint on f in the interval $w[V{+}1, V{+}3k - 1]$, whence $\tilde{u}$ contains the factor $byc\tilde{y}byc$, contrary to the assumption that u is primitive (Lemma 6 case (iv)).

Case (iii): $\frac{1}{3}\ell < k < \frac{1}{2}\ell$. Again, in this case, we may have either $Y > V$ or $Y < V$. This time (for variety) assume the latter; however, the reasoning in the former case is almost identical. Thus, we have the situation depicted in Fig. 3b. Let $w[V] = b$, $w[Y] = c$ and $w[Y{+}1, V{-}1] = y$. Since Y is a waypoint on g, we see that $w[Y - k] = b$ and $w[Y{-}k{+}1, Y{-}1] = \tilde{y}$, whence $w[V{-}2k, V] = w[Y{-}k, V] = b\tilde{y}cyb$. Since V is a waypoint on f, we see that also $w[V, V{+}2k] = b\tilde{y}cyb$. Thus, u contains the factor $b\tilde{y}cyb$ and v contains the factor $cyb\tilde{y}c$; moreover $w[Y, Y{+}3k] = cyb\tilde{y}cyb$.

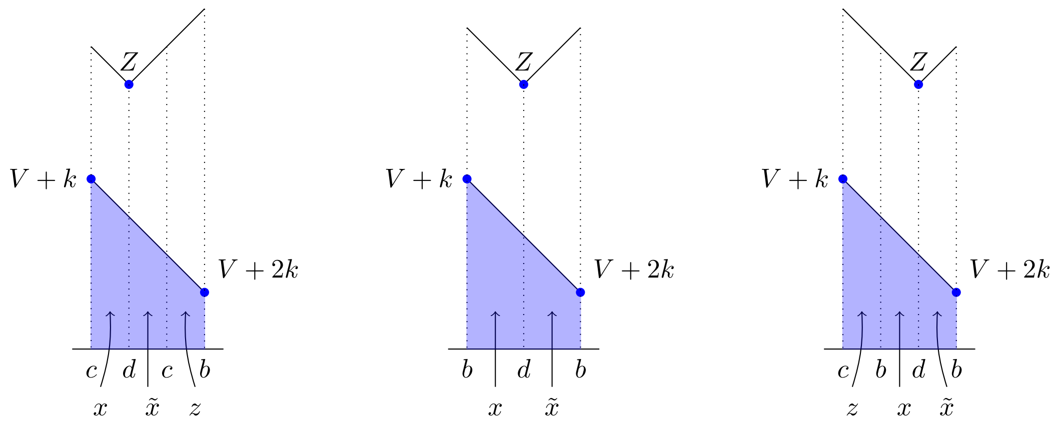
Now let Z be the next waypoint on g after Y. It is immediate that $Z{-}Y < 3k$, since otherwise, v contains the factor $cyb\tilde{y}cyc$, contrary to the assumption that v is primitive (Lemma 6 case (iv)). We consider three possibilities for the point Z, depending on where, exactly, Z is positioned in $[V{+}k, V{+}2k] = [Y{+}2k, Y{+}3k]$. The three possibilities are indicated in Fig. 4, which shows the detail of Fig. 3b in that interval. Suppose (a) that $V{+}k < Z < V{+}\frac{3}{2}k$. Then, by inspection of Fig. 4a, y must be of the form $xd\tilde{x}cz$ for some letter d and strings x and z. But we have already argued that u contains the factor

$$b\tilde{y}cyb = b(xd\tilde{x}cz)^{-1}c(xd\tilde{x}cz)b = b(\tilde{z}cxd\tilde{x})c(xd\tilde{x}cz)b$$



(a) Condition $k \leq \frac{1}{3}\ell$; for illustration, $Y > V$. **(b)** Condition $\frac{1}{3}\ell < k < \frac{1}{2}\ell$; for illustration, $Y < V$.

Figure 3 The walk g has a waypoint at Y with $k = |V - Y| \geq 1$.



(a) $Z < V + \frac{3}{2}k$. **(b)** $Z = V + \frac{3}{2}k$. **(c)** $Z > V+\frac{3}{2}k$.

Figure 4 The location of Z with respect to $V+\frac{3}{2}k$ in Case (iii).

and hence the factor $cxd\tilde{x}cxd$ contrary to the assumption that u is primitive (Lemma 6 case (iv)). Suppose (b) that $Z = V+\frac{3}{2}k$. Then, by inspection of Fig. 4b, y must be of the form $xd\tilde{x}$ for some letter d and string x, and furthermore, $b = c$. But in that case u contains the factor

$$b\tilde{y}cyb = c(xd\tilde{x})^{-1}c(xd\tilde{x})c = c(xd\tilde{x})c(xd\tilde{x})c$$

and hence the factor $cxd\tilde{x}cxd$ again. Suppose (c) that $V+\frac{3}{2}k < Z < V+2k$. Then by inspection of Fig. 4c, y must be of the form $zbxd\tilde{x}$ for some letter d and strings x and z. But we have already argued that v contains the factor

$$cyb\tilde{y}c = c(zbxd\tilde{x})b(zbxd\tilde{x})^{-1}c = c(zbxd\tilde{x})b(xd\tilde{x}b\tilde{z})c$$

and hence the factor $bxd\tilde{x}bxd$, again contrary to the assumption that u is primitive. This eliminates all possibilities for the position of Z, and thus yields the desired contradiction.

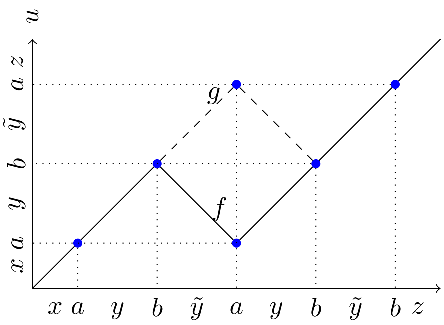


Figure 5 Distinct walks f (solid) and g (dashed and solid) on $u = xayb\tilde{y}az$ such that $u^f = u^g$.

The remaining cases, where $k > \ell/2$, or where the shortest leg is initial or final, are omitted because of space restrictions. See full version [5] for a complete proof.

4 Uniqueness of walks

In this short section, we prove Theorem 3, which states that a word is perfect if and only if it contains no non-trivial palindrome as a factor.

For the only-if direction, suppose that u contains a non-trivial palindrome. If that palindrome is odd, so that u has the form $xayb\tilde{y}az$, then the word $xayb\tilde{y}ayb\tilde{y}az$ is generated via the distinct walks f and g illustrated in Fig. 5. If the contained palindrome is even, so that u has the form $xaaz$, then the word $xaaaz$ is generated via distinct walks, one of which pauses for one step on the first a, and the other on the second.

For the converse, suppose for contradiction that u is a word of length n containing no non-trivial palindromes, for which there exist walks f and g such that $u^f = u^g$ but $f \neq g$. Let u, f and g be chosen so that $m = |u^f| = |u^g|$ is minimal. If $f(i) = f(i+1)$ for some i, we have $g(i) = g(i+1)$, since otherwise, u contains a repeated letter, and therefore a palindrome of length 2, contrary to assumption. But if both $f(i) = f(i+1)$ and $g(i) = g(i+1)$, then the functions $f' = f/[i, i+1]$ and $g' = g/[i, i+1]$ are defined, and are obviously walks, and moreover we have $u^{f'} = u^{g'}$ and $f' \neq g'$, contradicting the minimality of m. Hence, we may assume that neither f nor g is ever stationary. We claim that f and g have the same waypoints. For if i is an internal waypoint for f but not for g, we have $f(i-1) = f(i+1)$, $u[g(i-1)] = u[f(i-1)]$ and $u[g(i+1)] = u[f(i+1)]$, whence $u[g(i-1)] = u[g(i+1)]$, so that u contains an odd, non-trivial palindrome centred at $g(i)$, contrary to assumption. This establishes the claim that f and g have the same waypoints. Since u is certainly not itself a non-trivial palindrome and $f \neq g$, the walks f and g must have at least one internal waypoint between them. Now take a shortest leg of f (which must also be a shortest leg of g), say $[j, j+\ell]$. Suppose first that $[j, j+\ell]$ is an internal leg (i.e. $j < 1$ and $j+\ell < m$). To visualize the situation suppose $V = j$ and $W = j+\ell$ in Fig. 1b. Taking into account the legs either side, we see that $f(j) = f(j+2\ell)$ and $g(j) = g(j+2\ell)$, and moreover that $f' = f/[j, j+2\ell]$ and $g' = g/[j, j+2\ell]$ map $[1, m-2\ell]$ surjectively onto $[1, n]$. Clearly, $u^{f'} = u^{g'}$. But f and g have the same waypoints over the interval $[j, j+2\ell]$, whence $f \neq g$ implies $f' \neq g'$, contradicting the minimality of m. The cases where the shortest leg is terminal are handled similarly.

5 Words yielding the same results on distinct walks

In this section, we sketch the ideas behind the proof of Theorem 4, allowing us to characterize those primitive words which are solutions of a given equation $u^f = u^g$, for walks f and g.

Let $f' \colon [1,m] \to [1,n]$ be a walk. If $1 \le j \le m$, then the function $f \colon [1,m+1] \to [1,n]$ given by

$$f(i) = \begin{cases} f'(i) & \text{if } i \le j \\ f'(i{-}1) & \text{otherwise} \end{cases}$$

is also a walk, longer by one step. We call f the *hesitation on f' at j*, as it arises by executing f' up to and including the jth step, then pausing for one step, before continuing as normal. We next proceed to define an operation of *vacillation on f'*, also producing a strictly longer walk. This operation has three forms, depending on whether it occurs at the start, in the middle, or at the end of the walk. For any k $(1 \le k < m)$, we define the *initial vacillation on f' over* $[1,k{+}1]$ to be the walk $f \colon [1,m{+}k] \to [1,n]$ given by

$$f(i) = \begin{cases} f'(k{+}1{-}(i{-}1)) & \text{if } i \le k+1 \\ f'(i{-}k) & \text{otherwise.} \end{cases}$$

Thus f arises by executing the first $k+1$ steps of f' in reverse order and then continuing to execute f' from the second step as normal. Likewise, we define the *final vacillation on f' over* $[m{-}k,m]$ to be the walk $f \colon [1,m{+}k] \to [1,n]$ given by

$$f(i) = \begin{cases} f'(i) & \text{if } i \le m \\ f'(m{-}(i{-}m)) & \text{otherwise.} \end{cases}$$

Thus f arises by executing f' as normal and then repeating the k steps preceding the last in reverse order. Finally, for any j $(1 < j < m)$, and any k $(1 \le k < j)$, we define the *internal vacillation on f' over* $[j{-}k,j]$ to be the walk $f \colon [1,m{+}2k] \to [1,n]$ given by

$$f(i) = \begin{cases} f'(i) & \text{if } i \le j \\ f'(j{-}(i{-}j)) & \text{if } j < i \le j+k \\ f'(i-2k) & \text{otherwise.} \end{cases}$$

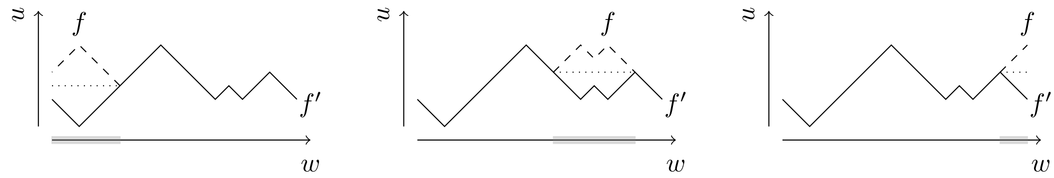
Thus f arises by executing f' up to the jth step, reversing the previous k steps back to the $(j{-}k)$th step and then continuing from the $(j-k+1)$th step as normal. A *vacillation* on f' is an initial, internal or final vacillation on f'.

Let $f' \colon [1,m] \to [1,n]$ again be a walk. We proceed to define an operation of *reflection* on f', producing a stroll (not necessarily surjective) of the same length. For any k $(1 \le k < m)$, we take the *initial reflection on f' over* $[1,k{+}1]$ to be the function f defined on the domain $[1,m]$ by setting

$$f(i) = \begin{cases} f'(k{+}1){-}(f'(i){-}f'(k{+}1)) & \text{if } i \le k+1 \\ f'(i) & \text{otherwise.} \end{cases}$$

Thus f arises by reflecting the segment of f' over the interval $[1,k{+}1]$ in the horizontal axis positioned at height $f'(k+1)$, and then continuing as normal (Fig. 6a). Likewise, we take the *final reflection on f' over* $[m{-}k,m]$ to be the function f defined on $[1,m]$ by setting

$$f(i) = \begin{cases} f'(m{-}k){-}(f'(i){-}f'(m{-}k)) & \text{if } i \ge m{-}k \\ f'(i) & \text{otherwise.} \end{cases}$$



(a) Initial. **(b)** Internal. **(c)** Final.

Figure 6 The stroll f (dashed and solid) is a reflection on the walk f' (solid) over I (shaded).

Thus f arises by executing f' as normal up to the $(m{-}k)$th step, and then thereafter reflecting the remaining segment of f' in the horizontal axis positioned at height $f'(m{-}k)$ (Fig. 6c). Finally, for integers j, k ($1 < j < m$, $1 \leq k \leq \min(j{-}1, m{-}j)$) such that $f'(j{-}k) = f'(j{+}k)$, the *internal reflection on* f' over $[j{-}k, j{+}k]$ is the function f defined on $[1, m]$ by setting

$$f(i) = \begin{cases} f'(j{-}k){-}(f'(i){-}f'(j{-}k)) & \text{if } j{-}k \leq i \leq j{+}k \\ f'(i) & \text{otherwise.} \end{cases}$$

Thus f arises by executing f' up to the point $j{-}k$, then reflecting the segment of f' over the interval $[j{-}k, j+k]$ in the horizontal axis positioned at height $f'(j{-}k) = f'(j{+}k)$, thereafter executing f' as normal (Fig. 6b). A *reflection* on f' is an initial, internal or final reflection on f'. As defined above, reflections can take values in the range $[-n{+}1, 2n{-}1]$; accordingly, we call a reflection *proper* if all its values are within the interval $[1, n]$, and in that case we take the resulting function to have co-domain $[1, n]$. We shall only ever be concerned with proper reflections in the sequel; and a proper reflection on a walk (more generally, on a stroll) is evidently a stroll; there is no *a priori* requirement for it to be surjective.

Reflections are of most interest in connection with walks on words containing odd palindromes. Let $f' \colon [1, m] \to [1, n]$ be a stroll, and u be a word of length n. We say that a reflection f on f' is *admissible for* u if it is either: (i) an initial reflection over $[1, k{+}1]$, and u has a palindrome of length $2k{+}1$ centred at $f'(k{+}1)$; (ii) a final reflection over $[m{-}k, m]$, and u has a palindrome of length $2k{+}1$ centred at $f'(m{-}k)$; or (iii) an internal reflection over $[j{-}k, j{+}k]$, and u has a palindrome of length $2k{+}1$ centred at $f'(j{-}k) = f'(j{+}k)$. We see by inspection of Fig. 6 that, if f is a reflection on f' admissible for u, then $u^f = u^{f'}$.

Suppose now f' and g' are walks with domain $[1, m]$ and co-domain $[1, n]$. If f and g are hesitations on f' and g', respectively, at some common point, we say that the pair of walks $\langle f, g \rangle$ is a *hesitation* on the pair $\langle f', g' \rangle$; similarly, if f and g are vacillations on f' and g' over some common interval, we say that the pair of walks $\langle f, g \rangle$ is a *vacillation* on the pair $\langle f', g' \rangle$. Evidently, if u is a word such that $u^{f'} = u^{g'}$ and $\langle f, g \rangle$ is a hesitation or vacillation on $\langle f', g' \rangle$, then $u^f = u^g$. If now f is a reflection on f' over some interval, we say that $\langle f, g' \rangle$ is a *reflection on* $\langle f', g' \rangle$, and also that $\langle g', f \rangle$ is a *reflection on* $\langle g', f' \rangle$. Evidently, if the reflection in question is (proper and) admissible for some word u such that $u^{f'} = u^{g'}$, then $u^f = u^{g'}$. Note that hesitations/vacillations on pairs of strolls are hesitations/vacillations on *both* of the strolls in question, while reflections on pairs of strolls are reflections on *either* of the strolls in question.

Now let f' and g' be walks, and suppose u is a word of length m such that $u^{f'} = u^{g'}$. We have seen that, if $\langle f, g \rangle$ is a hesitation or vacillation on $\langle f', g' \rangle$, or is a reflection on $\langle f', g' \rangle$ admissible for u, then $u^f = u^g$. The principal result of this section states that, for primitive words, this is essentially the only way in which we can arrive at distinct walks f and g such that $u^f = u^g$.

▶ **Lemma 7.** *Let u be a primitive word of length n, and let f and g be walks with domain $[1,m]$ and co-domain $[1,n]$ such that $u^f = u^g$. Then there exist sequences of walks $\{f_s\}_{s=0}^t$ and $\{g_s\}_{s=0}^t$, all having co-domain $[1,n]$, satisfying:* (i) *$f_0 = g_0$ is monotone;* (ii) *for all s $(0 \leq s < t)$, $\langle f_{s+1}, g_{s+1}\rangle$ is a hesitation on $\langle f_s, g_s\rangle$, a vacillation on $\langle f_s, g_s\rangle$, or a reflection on $\langle f_s, g_s\rangle$ admissible for u; and* (iii) *$f_t = f$ and $g_t = g$.*

Proof. Similar in character to the proof of Theorem 1. See full version [5] for details. ◀

Lemma 7 gives us everything we need for the proof of Theorem 4, which states that, for a primitive word u of length n, and walks $f, g\colon [1,m] \to [1,n]$, we have $u^f = u^g$ if and only if $\langle f(i), g(i)\rangle \in \Delta_u^*$ for all $i \in [1,m]$. Recall that Δ_u^* is the equivalence closure of Δ_u, the defect set of u.

The if-direction is almost trivial. Indeed, $\langle j,k\rangle \in \Delta_u$ certainly implies $u[j] = u[k]$, whence $\langle j,k\rangle \in \Delta_u^*$ also implies $u[j] = u[k]$. Thus, if $\langle f(i), g(i)\rangle \in \Delta_u^*$ for all $i \in [1,m]$, then $u[f(i)] = u[g(i)]$ for all $i \in [1,m]$, which is to say $u^f = u^g$.

For the only-if direction, we suppose that $u^f = u^g$. By Lemma 7, we may decompose the pair of walks $\langle f,g\rangle$ into a series $\{\langle f_s, g_s\rangle\}_{s=0}^t$ such that: (i) $f_0 = g_0$; (ii) for all s $(0 \leq s < t)$, the pair $\langle f_{s+1}, g_{s+1}\rangle$ is obtained by performing a hesitation, vacillation, or an admissible (for u) reflection on $\langle f_s, g_s\rangle$; and (iii) $\langle f_t, g_t\rangle = \langle f,g\rangle$. We establish that the following holds for all s $(0 \leq s \leq t)$:

$$\langle f_s(i), g_s(i)\rangle \in \Delta_u^* \text{ for all } i \text{ in the domain of } f_s \ (= \text{the domain of } g_s). \tag{1}$$

Putting $s = t$ then secures the required condition.

We proceed by induction on s. For the base case, where $s = 0$, we have $f_0 = g_0$, and there is nothing to do. For the inductive step, we suppose (1), and show that the same holds with s replaced by $s+1$. We have three cases.

Case 1. $\langle f_{s+1}, g_{s+1}\rangle$ is obtained by a hesitation on $\langle f_s, g_s\rangle$ at j. If $i \leq j$ then $f_{s+1}(i) = f_s(i)$ and $g_{s+1}(i) = g_s(i)$; and by (1), $\langle f_s(i), g_s(i)\rangle \in \Delta_u^*$. If $i > j$ then $f_{s+1}(i) = f_s(i-1)$ and $g_{s+1}(i) = g_s(i-1)$; and by (1), $\langle f_s(i-1), g_s(i-1)\rangle \in \Delta_u^*$. Either way, $\langle f_{s+1}(i), g_{s+1}(i)\rangle \in \Delta_u^*$.

Case 2. $\langle f_{s+1}, g_{s+1}\rangle$ is a vacillation on $\langle f_s, g_s\rangle$. We consider the case of an internal vacillation over some interval over $[j-k, j]$; initial and final vacillations are handled similarly. Again, if $i \leq j$ then $f_{s+1}(i) = f_s(i)$ and $g_{s+1}(i) = g_s(i)$; and by (1), $\langle f_s(i), g_s(i)\rangle \in \Delta_u^*$. If $j < i \leq j+k$, then $f_{s+1}(i) = f_s(j-(i-j))$ and $g_{s+1}(i) = g_s(j-(i-j))$; and by (1), $\langle f_s(j-(i-j)), g_s(j-(i-j))\rangle \in \Delta_u^*$. Finally, if $i > j+k$, then $f_{s+1}(i) = f_s(i-2k)$ and $g_{s+1}(i) = g_s(i-2k)$; and by (1), $\langle f_s(i-2k), g_s(i-2k)\rangle \in \Delta_u^*$.

Case 3. $\langle f_{s+1}, g_{s+1}\rangle$ is the result of a reflection on $\langle f_s, g_s\rangle$ over some interval $[j-k, j+k]$, with the reflection in question admissible for u. By exchanging f and g if necessary, we may assume that f_{s+1} is a reflection on f_s over $[j-k, j+k]$, and $g_{s+1} = g_s$; it does not matter for the ensuing argument whether the reflection in question is internal, initial or final. If $i \notin [j-k, j+k]$, then $f_{s+1}(i) = f_s(i)$ and $g_{s+1}(i) = g_s(i)$; and by (1), $\langle f_s(i), g_s(i)\rangle \in \Delta_u^*$. So suppose $i \in [j-k, k+j]$. Since the reflection over $[j-k, j+k]$ is admissible, the factor $u[f_s(j-k), f_s(j+k)]$ is a palindrome. Moreover, from the definition of reflection, either $u[f_{s+1}(i), f_s(i)]$ or $u[f_s(i), f_{s+1}(i)]$ is a palindromic factor of u (depending on whether $f_{s+1}(i) \leq f_s(i)$ or $f_{s+1}(i) \geq f_s(i)$). That is, either $\langle f_{s+1}(i), f_s(i)\rangle \in \Delta_u$ or $\langle f_s(i), f_{s+1}(i)\rangle \in \Delta_u$. But by (1), $\langle f_s(i), g_s(i)\rangle = \langle f_s(i), g_{s+1}(i)\rangle \in \Delta_u^*$. Hence $\langle f_{s+1}(i), g_{s+1}(i)\rangle \in \Delta_u^*$, again as required. This concludes the induction, and hence the proof of the only-if direction.

▶ **Corollary 8.** *Let v_1 and v_2 be primitive words of length n. Then v_1 and v_2 satisfy the same equations $u^f = u^g$, where f and g are walks with co-domain $[1, n]$, if and only if $\Delta_{v_1} = \Delta_{v_2}$.*

Proof. The if-direction is immediate from Theorem 4. For the only-if direction, suppose v_1 and v_2 satisfy the same equations $u^f = u^g$. If v_1 contains a non-trivial palindrome of (necessarily odd) length, say, $2k+1$ centred at i, let f and g be walks as depicted in Fig. 5, diverging at i and re-converging at $i+2k$. Thus $v_1^f = v_1^g$ and hence $v_2^f = v_2^g$. But considering f and g over the interval $[i, i+k]$, the equation $v_2^f = v_2^g$ clearly implies that v_2 has a a palindrome of length $2k+1$ centred at $f(i) = g(i) = i$, whence $\Delta_{v_2} \supseteq \Delta_{v_1}$. By symmetry, $\Delta_{v_1} \supseteq \Delta_{v_2}$ ◀

For a treatment of the problem of finding palindromes in words, see [2, Ch. 8].

6 Primitive generators of some morphic words

In this section, we prove Theorem 5, which states that, for each $k \geq 2$, all elements of the k-bonacci sequence $\{\alpha_n^{(k)}\}_{n\geq 1}$ from the kth onwards have the same primitive generator. In the sequel, we employ decorated versions of α, β, γ as constants denoting words.

We work with an alternative, recursive definition of the words $\alpha_n^{(k)}$. For all $k \geq 1$, let $\beta_k = \beta'_k \cdot k$, where β'_k is recursively defined by setting $\beta'_1 = \varepsilon$ and $\beta'_{k+1} = \beta'_k \cdot k \cdot \beta'_k$ for all $k \geq 1$. Now define, for any $k \geq 2$ the sequence $\{\alpha_n^{(k)}\}_{n\geq 1}$ by declaring, for all $n \geq 1$: $\alpha_n^{(k)} = \beta_n$ if $n \leq k$, and $\alpha_n^{(k)} = \alpha_{n-1}^{(k)}\alpha_{n-2}^{(k)} \cdots \alpha_{n-k}^{(k)}$ otherwise. A simple induction shows that this definition of the $\alpha_n^{(k)}$ coincides with that given in the introduction via morphisms. We remark that $|\beta_k| = 2^{k-1}$, for all $k \geq 1$.

We now define the primitive generators promised by Theorem 5. For all $k \geq 2$, let $\gamma'_k = (k-1) \cdot \beta'_{k-1}$ and $\gamma_k = \gamma'_k \cdot k$. We remark that $|\gamma_k| = 2^{k-2}+1$, for all $k \geq 2$. The following two claims are easily proved by induction.

▷ Claim 9. For all $k \geq 2$, β'_k is a palindrome over $\{1, \ldots, k-1\}$ containing exactly one occurrence of $k-1$ (in the middle); thus β_k contains exactly one occurrence of $k-1$ (at position $|\beta_k|/2$) and exactly one occurrence of k (at the end). For all $k \geq 3$, γ_k contains exactly one occurrence of each of k (at the end), $k-1$ (at the beginning) and $k-2$ (in the middle).

▷ Claim 10. For all $k \geq 2$, any position in the word γ_k is either occupied by the letter 1 or is next to a position occupied by the letter 1.

▷ Claim 11. For all $k \geq 2$, γ_k is primitive.

Proof. By induction on k. Certainly, $\gamma_2 = 12$ is primitive. For $k \geq 2$, by Claim 9, $\gamma_{k+1} = k \cdot \beta'_{k-1} \cdot (k-1) \cdot \beta'_{k-1} \cdot (k+1)$ contains exactly one occurrence of each of $k+1$, k and $k-1$. Considering the forms given by the four cases of Lemma 6, we see that γ_{k+1} does not have a prefix or suffix which is a non-trivial palindrome, and that any occurrence of either of the patterns aa or $axb\tilde{x}axb$ must be contained in one of the embedded occurrences of β'_{k-1} and hence in γ_k. By inductive hypothesis, γ_k is primitive, and therefore does not contain either of these patterns. But then γ_{k+1} is primitive by Lemma 6. ◁

▷ Claim 12. Let $k \geq 2$. For all h $(1 \leq h \leq |\gamma'_k|)$ such that $\gamma'_k[h] = 1$, there exists a walk f such that: (i) $\beta'_k = (\gamma'_k)^f$; (ii) $f(1) = h$; and (iii) $f(|\beta'_k|) = |\gamma'_k|$.

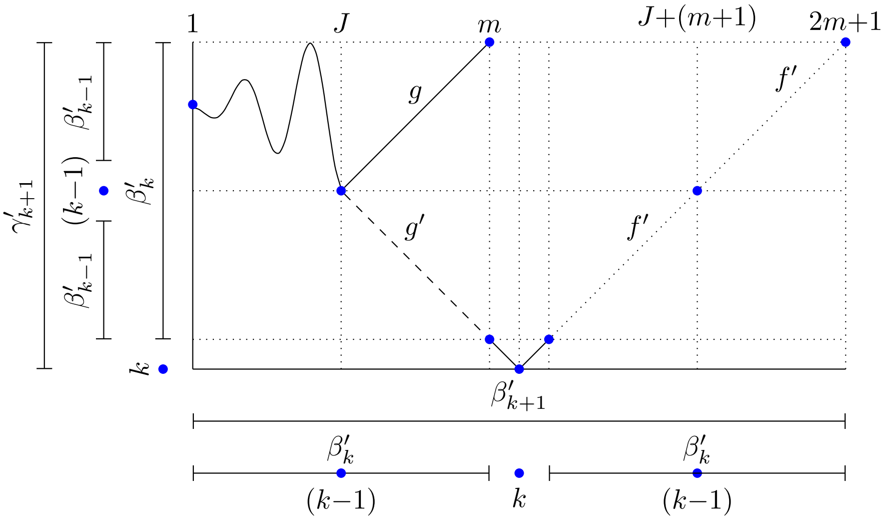


Figure 7 Proof of Claim 12: g (solid lines) is a shifted copy of a walk f on γ'_k yielding β'_k; g' (solid and dashed lines) is a final reflection on g over $[J, m]$; f' (solid, dashed and dotted lines) is a walk on $\gamma'_{k+1} = k \cdot \beta'_k \cdot (k+1)$ yielding $\beta'_{k+1} = \beta'_k \cdot k \cdot \beta'_k$.

Proof. We proceed by induction on k. For $k = 2$ and $k = 3$, the result is trivial, since $\beta'_2 = \gamma'_2 = 1$, $\beta'_3 = 121$ and $\gamma'_3 = 21$. Now suppose the claim holds for the value $k \geq 3$. For convenience, we write $m = |\beta'_k|$ and $n = |\gamma'_k| = |\beta'_{k-1}|+1$ (so $m = 2n-1$.). Remembering that $\beta'_{k+1} = \beta_k \cdot (k+1) \cdot \beta_k$, and $\gamma'_{k+1} = k \cdot \beta'_{k-1} \cdot (k-1) \cdot \beta'_{k-1} = k \cdot \beta'_{k-1} \cdot \gamma'_k$, we have $|\beta'_{k+1}| = 2m+1$ and $|\gamma'_{k+1}| = 2n$. To show that the claim also holds for the value $k+1$, pick any h' satisfying $(1 \leq h' \leq 2n)$ such that $\gamma'_{k+1}[h'] = 1$. We show that there exists a walk $f' \colon [1, 2m+1] \to [1, 2n]$ such that $\beta'_{k+1} = (\gamma'_{k+1})^{f'}$, $f'(1) = h'$, and $f'(2m+1) = 2n$.

Assume for the time being that $h' > n+1$, that is to say, h' is a position in $\gamma'_{k+1} = k \cdot \beta'_{k-1} \cdot (k-1) \cdot \beta'_{k-1}$ occupied by a 1 and lying in the *second* copy of β'_{k-1}. Then $h = h' - n$ is a position in $\gamma'_k = (k-1) \cdot \beta'_{k-1}$ occupied by a 1, so by inductive hypothesis, let $f \colon [1, m] \to [1, n]$ be a walk such that $\beta'_k = (\gamma'_k)^f$, $f(1) = h$, and $f(m) = n$. By Claim 9, β'_k contains exactly one occurrence of $k-1$ (this will be exactly in the middle), and γ'_k likewise contains exactly one occurrence of $k-1$ (this will be at the very beginning). Thus, f reaches the value 1 at just one point, namely $J = (m+1)/2$, and is otherwise strictly greater. (In fact, it is obvious that f must be a straight line from J onwards.) We first define a stroll $g \colon [1, m] \to [1, 2n]$ given by $g(i) = f(i)+n$ (Fig. 7, solid lines). Thus, $g(1) = h'$, and $g(m) = 2n$. Moreover, g reaches the value $n+1$ at just one point, namely $J = (m+1)/2$, and is otherwise strictly greater, as illustrated. Now let g' be the (final) reflection on g over the interval $[J, m]$ (Fig. 7, first solid, then dashed lines). Thus, g' is a stroll on γ'_{k+1} satisfying $g'(1) = h'$ and $g'(m) = 2$. Moreover, since $\beta'_k = \beta'_{k-1} \cdot (k-1) \cdot \beta'_{k-1}$ is a palindrome, we see by inspection that $(\gamma'_{k+1})^{g'} = (\gamma'_{k+1})^g = (\gamma'_k)^f = \beta'_k$. We now construct the desired walk $f' \colon [1, 2m+1] \to [1, 2n]$. For $i \in [1, m]$, we set $f'(i) = g'(i)$. Since $f'(m) = g'(m) = 2$, we set $f'(m+1) = 1$, and then proceed to define f' over the positions to the right, corresponding to the second copy of β'_k in the word $\beta'_{k+1} = \beta'_k \cdot k \cdot \beta'_k$. But this we can do by drawing a straight line, as shown in (Fig. 7). By inspection, f' has the required properties.

Finally, we consider the case where $h' \leq n+1$. Since $\gamma'_{k+1}[h'] = 1$ we in fact have $2 \leq h' \leq n$. And since β'_k is a palindrome, we may replace h' with the value $(n+1)+((n+1)-h)$ (i.e. reflect in the horizontal axis at height $n+1$) and construct f' as before. To re-adjust so that $f'(1)$ has the correct value, perform an initial reflection on f' over the interval $[1, J]$. ◁

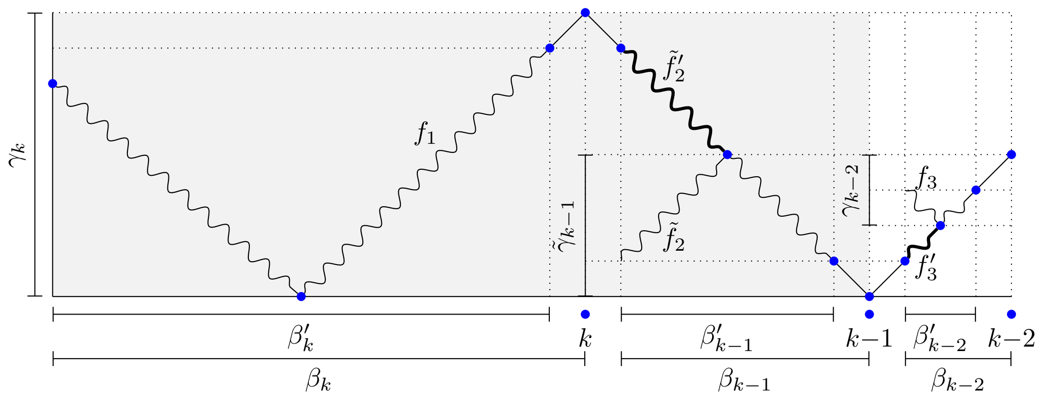


Figure 8 Proof of Claim 14 (schematic drawing): thin lines depict f_1, $\tilde{f}_2$ and f_3; thick lines denote the results $\tilde{f}'_2$ and f'_3 of performing initial reflections.

▷ Claim 13. Let $k \geq 2$. For all h ($1 \leq h \leq |\gamma_k|$) such that $\gamma_k[h] = 1$, there exists a walk f such that: (i) $\gamma_k^f = \beta_k$; (ii) $f(1) = h$; and (iii) $f(|\beta_k|) = |\gamma_k|$.

Proof. Take the walk guaranteed by Claim 12, and, noting that the final letters of β_k and γ_k are both k, extend f by setting $f(|\beta_k|) = |\gamma_k|$. ◁

▷ Claim 14. Let $k \geq 2$. For all h ($1 \leq h \leq |\gamma_n|$) such that $\gamma_k[h] = 1$, and for all p ($1 \leq p < k$) there exists a walk g such that: (i) $\gamma_k^f = \beta_k\beta_{k-1}\cdots\beta_{k-p}$ and (ii) $f(1) = h$.

Proof. Since $\beta_2\beta_1 = 121$ and $\gamma_2 = 12$, the claim is immediate for $k = 2$. Hence we may assume $k \geq 3$. By Claim 13, let f_1 be a walk on γ_k yielding β_k with $f_1(1) = h$ and $f_1(|\beta_k|) = |\gamma_k|$. Since f is a walk, and β_k contains only one occurrence of k, we have $f_1(|\beta_k|-1) = |\gamma_k|-1$. Set $g_1 = f_1$. By Claim 9, β'_{k-2} is a palindrome, whence $\gamma_k = (k-1) \cdot \beta'_{k-1} \cdot k = (k-1) \cdot \left(\beta'_{k-2} \cdot (k-2) \cdot \beta'_{k-2}\right) \cdot k = \tilde{\gamma}_{k-1} \cdot \beta'_{k-2} \cdot k$. Noting that the penultimate position of γ_{k-1} is always occupied by the letter 1, by Claim 13 let f_2 be a walk on γ_{k-1} yielding the word β_{k-1}, with $f_2(1) = |\gamma_{k-1}|-1$. By Claim 9, $f_2(i) = 1$ only when $i = |\beta_{k-1}|/2$, since that is the only position of β_{k-1} occupied by the letter $k-2$. It follows that the function $\tilde{f}_2$ defined by $\tilde{f}_2(i) = |\gamma_{k-1}|-(f_2(i)-1)$ is a walk on $\tilde{\gamma}_{k-1}$ yielding the word β_{k-1} with $\tilde{f}_2(1) = 2$, and achieving its maximum value $f(i) = |\gamma_{k-1}|$ only at $i = |\beta_{k-1}|/2$. Now regarding $\tilde{f}_2$ as a stroll on $\gamma_k = \tilde{\gamma}_{k-1} \cdot \beta'_{k-2} \cdot k$, let $\tilde{f}'_2$ be the initial reflection on $\tilde{f}_2$ over the interval $[1, |\beta_{k-1}|/2]$. Since $\gamma_k = (k-1) \cdot \beta'_{k-1} \cdot k$, with β'_{k-1} a palindrome, it follows that the stroll $\tilde{f}'_2$ on γ_k also yields the same result as $\tilde{f}_2$, namely β_{k-1}. Now let g_2 be the result of appending $\tilde{f}'_2$ to g_1, as shown in the shaded part of Fig. 8. (Most of the curves drawn schematically here will actually be straight lines, but no matter.) Formally, we define $g_2 : [1, |\beta_k| + |\beta_{k-1}|] \to [1, |\gamma_k|]$ to be the function:

$$g_2(i) = \begin{cases} g_1(i) & \text{if } 1 \leq i \leq |\beta_k| \\ \tilde{f}'_2(i - |\beta_k|) & \text{if } |\beta_k| < i \leq |\beta_k| + |\beta_{k-1}|. \end{cases}$$

Since $g_1(|\beta_k|) = |\gamma_k|$ and $\tilde{f}'_2(1) = |\gamma_k|-1$, we see that g_2 is indeed a walk on γ_k as shown (i.e. with no jumps), yielding $\beta_k\beta_{k-1}$. We remark that $g_2(|\beta_k| + |\beta_{k-1}|) = 1$. Notice that we needed to invert f_2 to yield $\tilde{f}_2$, so as to make the latter's reflection $\tilde{f}'_2$ join up to the end of g_1 properly.

We now repeat the above procedure, as shown in the unshaded part of Fig. 8. By Claim 13, and noting that the penultimate position of γ_{k-2} is occupied by the letter 1, let f_3 be a walk on γ_{k-2} yielding the word β_{k-2}, with $f_3(1) = |\gamma_{k-2}|-1$. By Claim 9, $f_3(i) = 1$ only when $i = |\beta_{k-2}|/2$, since that is the only position of β_{k-2} occupied by the letter $k-3$. Observing that $\gamma_{k-1} = \tilde{\gamma}_{k-2} \cdot \beta'_{k-3} \cdot k$, and hence $\tilde{\gamma}_{k-1} = k \cdot \tilde{\beta}'_{k-3} \cdot \gamma_{k-2}$, we see that, by shifting f_3 upwards by $|k \cdot \tilde{\beta}'_{k-2}|$, we can regard it as a stroll on γ_k. This (shifted) stroll reaches its minimum value $|k \cdot \tilde{\beta}'_{k-3}| + 1$ exactly once in the middle of its range. Let f'_3 be the initial reflection on of this stroll over the interval $[1, |\beta_{k-2}|/2]$. Since $\tilde{\gamma}_{k-1} = (k-1) \cdot \beta'_{k-2} \cdot (k-2)$ with β'_{k-2} a palindrome, we see by inspection that the stroll f'_3 on γ_k yields the same result as f_3, namely β_{k-2}. Now take g_3 to be the result of appending f'_3 to g_2, just as we earlier appended $\tilde{f}'_2$ to g_1. Thus, g_3 is a walk on γ_k yielding $\beta_k\beta_{k-1}\beta_{k-2}$. Notice that f_3, unlike f_2, did not need to be inverted to make its reflection f'_3 join up to the end of g_2. Evidently, this process may be continued until we obtain the desired walk g_{p+1} on γ_k yielding $\beta_k\beta_{k-1}\cdots\beta_{k-p}$, with the inversion step (producing $\tilde{f}_h$ from $\tilde{f}_h$) required only when h is even. ◁

We now prove Theorem 5, establishing by induction the following slightly stronger claim.

▷ Claim 15. Fix $k \geq 2$. For all $n \geq k$ and for all h $(1 \leq h \leq |\gamma_k|)$ such that $\gamma_n[h] = 1$, there exists a walk f such that $\alpha_n^{(k)} = \gamma_k^f$ and $f(1) = h$.

Proof. If $n = k$, then $\alpha_n^{(k)} = \beta_k$, and the result is immediate from Claim 13. If $n = k+1$, then $\alpha_n^{(k)} = \beta_k\beta_{k-1}\cdots\beta_1$, and the result is immediate from Claim 14, setting $p = k-1$.

For the inductive step we suppose $n \geq k+2$ and assume the result holds for values smaller than n. We consider first the slightly easier case where $n \geq 2k$. Set $h_1 = h$. Writing $\alpha_n^{(k)} = \alpha_{n-1}^{(k)} \cdots \alpha_{n-k}^{(k)}$, by inductive hypothesis, let g_1 be a walk such that $\alpha_{n-1}^{(k)} = \gamma_k^{g_1}$ and $g_1(1) = h_1$. Now let h'_1 be the final value of g_1, that is, $g_1(|\alpha_{n-1}^{(k)}|) = h'_1$. By Claim 10, there exists h_2 such that $|h_2 - h'_1| \leq 1$, and $\gamma_k[h_2] = 1$. Again, by inductive hypothesis, let g_2 be a walk such that $\alpha_{n-2}^{(k)} = \gamma_k^{g_2}$ and $g_2(1) = h_2$. Let h'_2 be the final value of g_2, and let h_3 be such that $|h_3 - h'_2| \leq 1$, and $\gamma_k[h_3] = 1$. Proceed in the same way, obtaining walks $g_3, \ldots, g_k$. Taking f to be the result of concatenating $g_1, g_2, g_3, \ldots, g_k$ in the obvious fashion yields the desired walk.

If $2k > n \geq k+2$, then we have $\alpha_n^{(k)} = \alpha_{n-1}^{(k)}\alpha_{n-2}^{(k)} \cdots \alpha_{k+1}^{(k)}\beta_k\beta_{k-1}\cdots\beta_{k-p}$, where $p = 2k-n$. We begin as in the previous paragraph: setting $h_1 = h$, by inductive hypothesis, let g_1 be a walk such that $\alpha_{n-1}^{(k)} = \gamma_k^{g_1}$ and $g_1(1) = h_1$. Now let h'_1 be the final value of g_1, that is, $g_1(|\alpha_{n-1}^{(k)}|) = h'_1$. By Claim 10, there exists h_2 such that $|h_2 - h'_1| \leq 1$, and $\gamma_k[h_2] = 1$. Now continue as before so as to obtain walks $g_2, g_3 \ldots$, with respective starting points $h_2, h_3, \ldots$, but stopping when we have obtained g_{k-p-1}, and the following starting point h_{k-p}. Observe that concatenating $g_1, g_2, g_3, \ldots, g_{k-p-1}$ gives a walk on γ_k which yields the word $\alpha_{n-1}^{(k)}\alpha_{n-2}^{(k)} \cdots \alpha_{k+1}^{(k)}$. By Claim 14, choose g_{k-p} to be a walk on γ_k yielding the word $\beta_k\beta_{k-1}\cdots\beta_{k-p}$ and with $g_{k-p}(1) = h_{k-p}$. Taking f to be the result of concatenating $g_1, g_2, g_3, \ldots, g_{k-p}$ establishes the claim. ◁

References

1 Bartosz Bednarczyk, Daumantas Kojelis, and Ian Pratt-Hartmann. On the Limits of Decision: the Adjacent Fragment of First-Order Logic. In Kousha Etessami, Uriel Feige, and Gabriele Puppis, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 111:1–111:21, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

2 M. Crochemore and W. Rytter. *Jewels of stringology.* World Scientific, Singapore and River Edge, NJ, 2002.

3 N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics.* Number 1794 in Lecture Notes in Mathematics, edited by V. Berthé, S. Ferenczi, C. Mauduit, and A. Siegel. Springer Verlag, Berlin, Heidelberg, New York, 2002.

4 M. Lothaire. *Applied Combinatorics on Words.* Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 2005.

5 Ian Pratt-Hartmann. Walking on words (v.2), 2024. `arXiv:2208.08913`.

6 Gérard Rauzy. Nombres algébriques et substitutions. *Bulletin de la Société Mathématique de France*, 110:147–178, 1982.