

# Walking on Words

Ian Pratt-Hartmann

Department of Computer Science  
University of Manchester, Manchester M13 9PL, United Kingdom

Instytut Informatyki  
Uniwersytet Opolski, 45-052 Opole, Poland

Take any word over some alphabet. If it is non-empty, go to any position and print out the letter being scanned. Now repeat the following any number of times (possibly zero): either stay at the current letter, or move one letter leftwards (if possible) or move one letter rightwards (if possible); then print out the letter being scanned. In effect, we are *going for a walk on* the input word. Let  $u$  be the infix of the input word comprising the visited positions, and  $w$  the word printed out (empty if the input word is). Since any unvisited prefix or suffix of the input word cannot influence  $w$ , we may as well discard them, and say that  $u$  *generates*  $w$ . It is obvious that every word generates itself (and its reversal), and that every non-empty word generates words of all lengths greater than itself. We ask about the converse. Given a word  $w$ , what words  $u$  generate it? The answer is surprising. Call  $u$  a *primitive generator* of  $w$  if  $u$  generates  $w$  and is not generated by any word shorter than  $u$ . We show that, excepting some degenerate cases, every word has precisely two primitive generators.

The present study belongs to the subject of *Combinatorics of Strings*. For a general introduction, see [1]. As far as the author is aware, the specific problem considered here is new. There is a rather tenuous connection to the study of random walks on graphs [3], with the graphs in question being linear orders; however, in the present paper, we shall not be concerned with transition *probabilities* at all. Indeed, much of the reasoning found in the sequel is reminiscent of standard text-matching algorithms, such [2], but that would appear to be the extent of the connection.

## 1 Primitive generators of words

For integers  $i, k$  denote by  $[i, k]$  the set  $\{j \in \mathbb{Z} \mid i \leq j \leq k\}$ . If  $m$  and  $n$  are positive integers ( $m \leq n$ ), a *walk* is a surjection  $f: [1, n] \rightarrow [1, m]$  such that  $|f(i+1) - f(i)| \leq 1$  for all  $i$  ( $1 \leq i < n$ ). (Regarding  $[1, n]$  and  $[1, m]$  as metric spaces,  $f$  is thus a metric map.) If  $u = a_1 \cdots a_m$  and  $w = c_1 \cdots c_n$  are words over some alphabet  $\Sigma$ , say that  $u$  *generates*  $w$  if there exists a walk  $f: [1, n] \rightarrow [1, m]$  such that  $c_i = a_{f(i)}$  for all  $i$  ( $1 \leq i \leq n$ ). Intuitively,  $f$  tells us where we should be in  $u$  at each time point in the range  $[1, n]$  on the walk generating  $w$ : the

condition  $|f(i+1)-f(i)| \leq 1$  for all  $i$  ( $1 \leq i < n$ ) ensures that we never change position by more than one letter at a time; the condition that  $f$  is surjective ensures that we visit every position of  $u$ . We may picture a walk as a piecewise linear function, with the generated word superimposed on the abscissa and the generating word on the ordinate. For example, Fig. 1 shows how  $u = \text{cbadefgh}$  generates  $w = \text{abcbaaadefedadefghgf}$ .



Figure 1: Generation of  $w = \text{abcbaaadefedadefghgf}$  from  $u = \text{cbadefgh}$ .

Generation is transitive: if  $u$  generates  $v$  and  $v$  generates  $w$ , then  $u$  generates  $w$ . If  $u = a_1 \cdots a_m$  is a word, denote the length of  $u$  by  $|u| = m$ , and the reversal of  $u$ —i.e. the word  $a_m \cdots a_1$ —by  $u^{-1}$ . Clearly, every word generates both itself and its reversal. Moreover, if  $u$  generates  $w$ , then  $|u| \leq |w|$ ; in fact,  $u$  and  $u^{-1}$  are the only words of length  $|u|$  generated by  $u$ . We call  $u$  *primitive* if it is not generated by any words shorter than itself, equivalently, if it is generated only by itself and its reversal. For example,  $\text{babcd}$  and  $\text{abcbcd}$  are not primitive, because they are generated by  $\text{abcd}$ ; but  $\text{abcbda}$  is primitive. Note that an infix (factor) of a primitive word need not be primitive. Define a *primitive generator* of  $w$  to be a generator of  $w$  which is itself primitive. It is obvious that every word  $w$  has some primitive generator  $u$ , and indeed,  $u^{-1}$  as well, since the reversal of a primitive generator is clearly a primitive generator. We show that there are no others:

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

Another way of stating the theorem is as follows. Suppose  $u$  and  $v$  are primitive words such that  $u \neq v$  and  $u \neq v^{-1}$ . Now suppose we go for a walk on  $u$  and, independently, go for a walk on  $v$ , under the stipulation that the two walks visit every position in the word in question. Then there is no possibility of coordinating these walks so that the same word is printed out.

A *palindrome* is a word  $u$  such that  $u = u^{-1}$ . No palindrome of length greater than 1 is primitive. Indeed, if  $|u|$  is even, then  $u$  has a double letter in the middle, and so is certainly not primitive; if  $|u|$  is odd, then it can be generated by its prefix (or suffix) of length  $(|u|+1)/2 < |u|$ . (The walk starts at one end, goes just over half way, and then returns to its starting point.) Call a word *unilateral* if it is of the form  $a^n$  for some letter  $a$  and some  $n \geq 0$ . Note that the empty word  $\epsilon$  counts as unilateral.

**Corollary 2.** *Every unilateral 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  $u^{-1}$  for some word  $u$ . The first statement of the corollary is obvious: if  $w = \epsilon$  then  $u = u^{-1} = \epsilon$ ; and if  $u = a^n$  for some  $n$  ( $n \geq 1$ ), then  $u = u^{-1} = a$ . If  $w$  is not uniliteral, then  $|u| > 1$ . But since  $u$  is primitive, we certainly cannot have  $u = u^{-1}$ .  $\square$

We have observed that no word containing an immediately repeated letter, i.e. no word of the form  $xaay$  for some words  $x, y$  and some letter  $a$ , can be primitive, since it is generated by  $xay$ . The key to understanding Theorem 1 is to generalize this observation.

**Lemma 3.** *No word of any of the forms (i)  $xaay$ , (ii)  $x^{-1}axy$ , (iii)  $yaxx^{-1}$  or (iv)  $yaxbx^{-1}axbz$ , where  $a, b$  are letters and  $x, y, z$  are words, is primitive.*

*Proof.* (i)  $xay$  generates  $xaay$ ; (ii)  $axy$  generates  $x^{-1}axy$  (start at the end of the prefix  $ax$ , go left to the beginning and then go to the end); (iii)  $yxa$  generates  $yaxx^{-1}$  (go to the end, turn back and redo the suffix  $xa$  in reverse); and (iv)  $yaxbz$  generates  $yaxbx^{-1}axbz$  (go to  $b$ , then turn back to  $a$ , then turn again and go to the end).  $\square$

It easily follows from parts (i) and (ii) of Lemma 3 that, over the alphabet  $\{0, 1\}$ , there are exactly five primitive words:  $\epsilon$ ,  $0$ ,  $1$ ,  $01$ , and  $10$ . However, over any larger alphabet, there are infinitely many. For example, over the alphabet  $\{0, 1, 2\}$ , all words of the form  $(012)^n$  (for  $n \geq 0$ ) are primitive, because no position in such a word has identical left and right neighbours. We mention in passing that a converse of Lemma 3 may easily be proved: if a word is not primitive; it has one of the forms (i)–(iv); however, we shall not require this fact. In the sequel, we denote the  $i$ th letter of a word  $w = a_1 \cdots a_n$  by  $w[i] = a_i$ , and denote the infix of  $w$  from the  $i$ th to  $j$ th letters by  $w[i, j] = a_i \cdots a_j$ . The concatenation of two words  $u$  and  $v$  is denoted  $uv$ , or sometimes, for clarity,  $u \cdot v$ .

## 2 Proof of main result

The following terminology will be useful. (Refer to Fig. 1 for motivation.) Let  $f: [1, n] \rightarrow [1, m]$  be a walk. By a *leg* of  $f$ , we mean a maximal interval  $[i, j] \subseteq [1, n]$  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 number  $h$  which forms the boundary between two consecutive legs will be called a *waypoint*. We count the numbers  $1$  and  $n$  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. A leg is *terminal* if either of the waypoints it connects is terminal, otherwise *internal*. 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.

We introduce some notation for excising segments from a walk. Let  $f: [1, n] \rightarrow [1, m]$  be a walk. As usual, for  $X \subseteq [1, n]$ , we write  $f(X)$  to denote  $\{f(x) \mid x \in$

$X\}$ . If  $1 < i \leq j < n$  such that (i)  $f(i) = f(j)$ , and (ii)  $f([1, i]) \cup f([j+1, n]) = [1, m]$ , define the function  $f': [1, n-j+i] \rightarrow [1, m]$  by

$$f'(h) = \begin{cases} f(h) & \text{if } 1 \leq h \leq i \\ f(h+j-i) & \text{ow.} \end{cases}$$

Intuitively,  $f'$  is just like  $f$ , but with the interval  $[i, j-1]$ —equivalently, the interval  $[i+1, j]$ —removed. From (i), we see that  $|f'(h+1) - f'(h)| \leq 1$  for all  $h$  ( $1 \leq h < n-j+i$ ), and from (ii), we see that  $f'$  is surjective. That is,  $f'$  is a walk, and we denote it by  $f/[i, j]$ . If  $i = 1$  or  $j = n$ , we change the definition slightly, because there is no analogue of condition (i). Specifically if  $1 \leq i, j \leq n$  such that  $f([j, n]) = [1, m]$  and  $f([1, i]) = [1, m]$ , define the functions  $f': [1, n-j+1] \rightarrow [1, m]$  and  $f'': [1, i] \rightarrow [1, m]$  by

$$f'(h) = f(j+h-1) \qquad f''(h) = f(h).$$

Intuitively,  $f'$  is just like  $f$ , but with the interval  $[1, j-1]$  removed, and  $f''$  is just like  $f$ , but with the interval  $[i+1, n]$  removed. That is,  $f'$  and  $f''$  are walks, and we denote them by  $f/[1, j]$  and  $f/[i, n]$ , respectively.

We prove Theorem 1 by contradiction, supposing that  $u$  and  $v$  are primitive words such that neither  $u = v$  nor  $u = v^{-1}$ , and  $w$  is a word generated from  $u$  by some walk  $f$  and from  $v$  by some walk  $g$ . Write  $|w| = n$ . 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. This assumption greatly simplifies the ensuing argument. In particular, 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  $w' = w[1, i] \cdot w[i+2, n]$ . Then  $w'$  is generated from  $u$  by  $f'$ , and also from  $v$  by  $g'$ , 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 hence 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  $w = u^{-1}$  and  $w = v$  or  $w = v^{-1}$ , whence  $u = v$  or  $u = v^{-1}$ , 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  $w = v^{-1}$ , 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  $k$  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  $[W, X]$ .

We suppose for the present that this leg is *internal*. Fig. 2 illustrates this situation where  $W$  is a peak and  $X$  a trough; but nothing essential would change if it were the other way around. Write  $U = W - k$  and  $Z = X + k$ . By the minimality of  $[W, X]$  (assumed internal), both  $U$  and  $Z$  must be in the interval  $[1, n]$ , and there can of course be no other waypoints in the interval  $[U+1, Z-1]$ . Now let  $w[U] = a$ ,  $w[W] = b$  and  $w[U+1, W-1] = x$ . Since  $W$  is

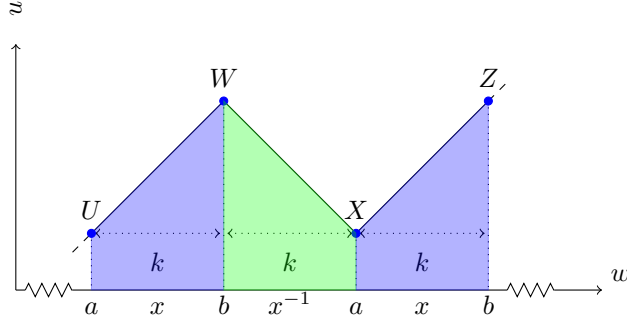


Figure 2: The minimal leg  $[W, X]$  of the walk  $f$  generating  $w$  from  $u$ .

a waypoint on  $f$ ,  $w[X] = a$  and  $w[W+1, X-1] = x^{-1}$ . Moreover, since  $X$  is a waypoint on  $f$ ,  $w[Z] = b$  and  $w[X+1, Z-1] = (x^{-1})^{-1} = x$ . We see immediately that  $g$  must have a waypoint in the interval  $[U+1, Z-1]$ , for otherwise,  $v$  (or  $v^{-1}$ ) contains an infix  $axbx^{-1}axb$ , contrary to the assumption that  $v$  is primitive (Lemma 3 (iv)). Let  $V$  be the waypoint on  $g$  which is closest to either of  $W$  or  $X$ . Replacing  $w$  by its reversal if necessary, assume that  $|V-W| \leq |V-X|$ , and write  $\ell = |V-W|$ . We consider possible values of  $\ell \in [0, k-1]$  in turn, deriving a contradiction in each case.

Case (i):  $\ell = 0$  (i.e.  $V = W$ ). For definiteness, let us suppose that  $V$  is as a peak, rather than a trough, but the reasoning is entirely unaffected by this determination. Certainly, we have  $g(U) = g(X)$ . By the minimality of the leg  $[W, X]$ ,  $g$  has no other waypoints in the interval  $[U+1, X-1]$ . By inspection of Fig. 2, it is also clear from the minimality of the leg  $[W, X]$  that  $f([1, U]) \cup f([X+1, n]) = [1, |u|]$ , whence the truncation  $f' = f/[U, X]$  is defined. We see immediately that  $g([1, U]) \cup g([X+1, n]) \neq [1, |v|]$ ; for otherwise,  $g' = g/[U, X]$  is defined, so that, writing  $w' = w[1, U] \cdot w[X+1, n]$ ,  $f'$  generates  $w'$  from  $u$  and  $g'$  generates  $w'$  from  $v$ , contrary to the assumption that  $w$  is a shortest counterexample. In other words, there are positions of  $v$  which  $g$  reaches over the range  $[U+1, X-1]$  that it does not reach outside this range. It follows that the position  $g(W) = g(V)$  in the string  $v$  is actually terminal. (Since we are assuming that  $V$  is a peak,  $g(V) = |v|$ ; but the following reasoning is unaffected if  $V$  is a trough and  $g(V) = 1$ .) It also follows that  $X$  itself cannot be a waypoint of  $g$ . For otherwise, the leg following  $X$ , which is of length at least  $k$ , covers all values in  $g([U, X])$ , thus ensuring that  $g([1, U]) \cup g([X+1, n]) = [1, |v|]$ , which we have just shown to be false. However,  $g$  must have some waypoint in  $[W+1, Z-1]$ . For if not, then  $g$  is decreasing between  $W$  and  $Z$  (remember that  $g(V) = g(W) = |v|$ ), and thus  $v$  has a suffix  $bx^{-1}axb$ , contrary to the assumption that  $v$  is primitive (Lemma 3 (iii)). By the minimality of the leg  $[W, X]$  we see that there is exactly one such waypoint, say  $Y$ . Since we have already shown that  $V$  is the only waypoint on  $g$  in  $[U+1, X-1]$ , and that  $X$  is not a waypoint on  $g$ , it follows that  $Y \in [X+1, Z-1]$ .

Now let  $j = Y - X$ . (Thus,  $1 \leq j < k$ .) If  $j > \frac{1}{2}k$ , we obtain the situation depicted in Fig. 3. Since  $g$  has a waypoint at  $Y$  and remembering that  $w[X+1, Z-1] = x$  and  $w[Z] = b$ , we see that  $x$  has the form  $ybzcz^{-1}$



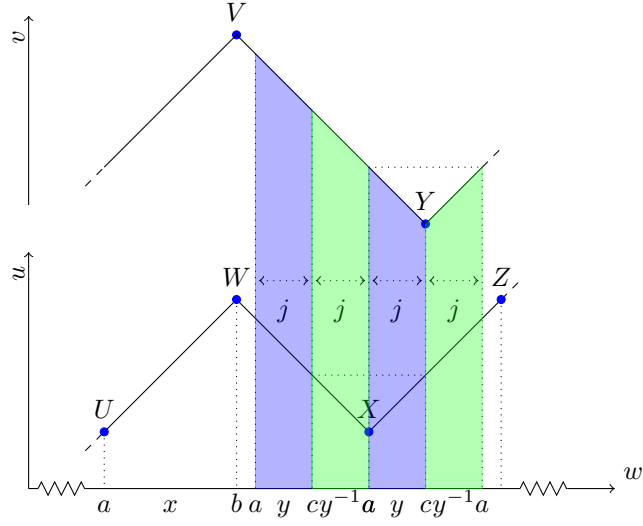


Figure 4: The walk  $g$  has waypoints at  $V = W$  and  $Y$  with  $j = Y - X < \frac{1}{2}k$ .

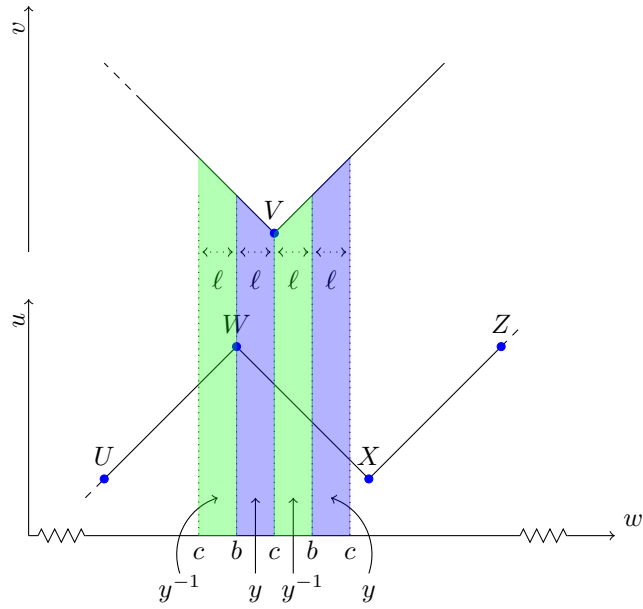


Figure 5: The walk  $g$  has a waypoint at  $V$  with  $\ell = |W - V| \leq \frac{1}{3}k$ ; for illustration, we assume  $V > W$ .

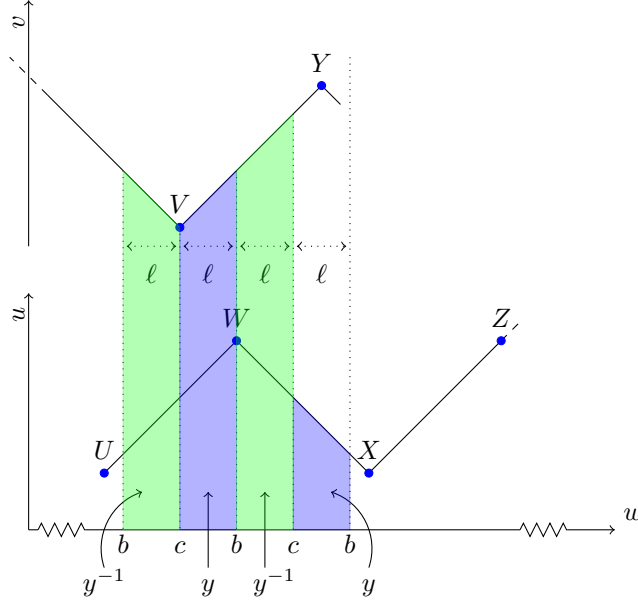


Figure 6: The walk  $g$  has a waypoint at  $V$  with  $\frac{1}{3}k < \ell = |W - V| < \frac{1}{2}k$ ; for illustration, we assume  $V < W$ .

$w[W-2\ell, W] = w[V-\ell, W] = by^{-1}cyb$ . Since  $W$  is a waypoint on  $f$ , we see that also  $w[W, W+2\ell] = by^{-1}cyb$ . Thus,  $u$  contains the infix  $by^{-1}cyb$  and  $v$  contains the infix  $cyby^{-1}c$ ; moreover  $w[V, V+3\ell] = cyby^{-1}cyb$ .

Now let  $Y$  be the next waypoint on  $g$  after  $V$ . It is immediate that  $Y - V < 3\ell$ , since otherwise,  $v$  contains the infix  $cyby^{-1}cyc$ , contrary to the assumption that  $v$  is primitive (Lemma 3 (iv)). We consider three possibilities for the point  $Y$ , depending on where, exactly,  $Y$  is positioned in  $[W+\ell, W+2\ell] = [V+2\ell, V+3\ell]$ . The three possibilities are indicated in Fig. 7, which shows the detail of Fig. 6 in that interval. Suppose (a) that  $W+\ell < Y < W+\frac{3}{2}\ell$ . Then, by inspection of Fig. 7(a),  $y$  must be of the form  $xdx^{-1}cz$  for some letter  $d$  and strings  $x$  and  $z$ . But we have already argued that  $u$  contains the infix

$$by^{-1}cyb = b(xdx^{-1}cz)^{-1}c(xdx^{-1}cz)b = b(z^{-1}cxdx^{-1})c(xdx^{-1}cz)b$$

and hence the infix  $cxdx^{-1}cxd$  contrary to the assumption that  $u$  is primitive (Lemma 3 (iv)). Suppose (b) that  $Y = W+\frac{3}{2}\ell$ . Then, by inspection of Fig. 7(b),  $y$  must be of the form  $xdx^{-1}$  for some letter  $d$  and string  $x$ , and furthermore,  $b = c$ . But in that case  $u$  contains the infix

$$by^{-1}cyb = c(xdx^{-1})^{-1}c(xdx^{-1})c = c(xdx^{-1})c(xdx^{-1})c$$

and hence the infix  $cxdx^{-1}cxd$  again. Suppose (c) that  $W+\frac{3}{2}\ell < Y < W+2\ell$ . Then by inspection of Fig. 7(c),  $y$  must be of the form  $zbx dx^{-1}$  for some letter  $d$  and strings  $x$  and  $z$ . But we have already argued that  $v$  contains the infix

$$cyby^{-1}c = c(zbx dx^{-1})b(zbx dx^{-1})^{-1}c = c(zbx dx^{-1})b(xdx^{-1}bz^{-1})c$$



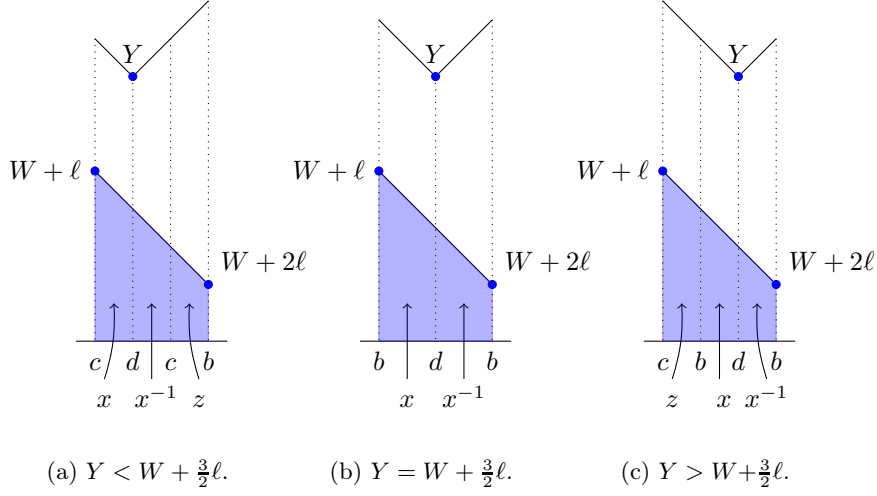


Figure 7: The location of  $Y$  with respect to  $W + \frac{3}{2}\ell$  in Case (iii).

and hence the infix  $bx dx^{-1} b x d$ , again contrary to the assumption that  $u$  is primitive. This eliminates all possibilities for the position of  $Y$ , and thus yields the desired contradiction.

Case (iv):  $\ell > \frac{1}{2}k$ . Since, by assumption,  $V$  is not closer to  $X$  than it is to  $W$ , we see that  $V$  cannot lie in the interval  $[W, X]$ , i.e.,  $V < W$ . Indeed,  $g$  cannot have any waypoint on the interval  $[W, X + \ell - 1]$ , so that we have the situation depicted in Fig. 8. Write  $w[U] = a$ ,  $w[V] = c$  and  $w[U+1, V-1] = y$ . In addition, let us write  $h = k - \ell < \frac{1}{2}k$ . Since  $V$  is a waypoint on  $g$ ,  $w[V, V+h] = cy^{-1}a$ , whence  $w[U, U+2h] = aycy^{-1}a$ . Using the fact that  $W$  is a waypoint on  $f$  with  $U$  and  $X$  symmetrically positioned with respect to  $W$ , we see that, also  $w[X-2h, X] = aycy^{-1}a$ . On the other hand,  $X$  is also a waypoint on  $f$ , whence  $w[X, V+h] = ayc$ . Thus  $w[X-2h, X+h] = aycy^{-1}ayc$ . But we have already observed that  $g$  has no waypoint on the interval  $[W, X + \ell - 1] \supset [W, X + h - 1]$ , whence  $v$  has an infix  $aycy^{-1}ayc$ , contrary to the assumption that  $v$  is primitive (Lemma 3 (iv)).

Case (v):  $\ell = \frac{1}{2}k$ . We first claim that  $g$  has a waypoint in the interval  $[W+1, X-1]$ . For otherwise, we have  $V < W$ , and the same situation as depicted in Fig. 8 arises, except that the unshaded region around  $W$  disappears, with  $a = b$ . But we have already argued that, in that case,  $v$  has an infix  $aycy^{-1}ayc$ , contrary to the assumption that  $v$  is primitive.

Thus, we may assume that  $V = W + \ell = X - \ell$ . By the minimality of the leg  $[W, X]$ , we know that  $g$  has no other waypoints in the interval  $[W - \ell + 1, X + \ell - 1]$ , and we have the situation depicted in Fig. 9. Observe that  $f(W - \ell) = f(X + \ell)$  and  $g(W - \ell) = g(X + \ell)$ . By inspection, we see that  $f([1, W - \ell]) \cup f([X + \ell, n]) = [1, |u|]$ , whence the truncation  $f' = f/[W - \ell, X + \ell]$  is defined. If, in addition  $g([1, W - \ell]) \cup g([X + \ell, n]) = [1, |v|]$ , we can define  $g' = g/[W - \ell, X + \ell]$ , and  $w' = w[1, W - \ell] \cdot w[X + \ell + 1, |u|]$ , so that  $f'$  generates  $w'$  from  $u$  and  $g'$  generates  $w'$  from  $v$ , contradicting the assumption that  $w$  is a shortest counterexample. Hence,  $g$  takes values on the interval  $[W - \ell + 1, X + \ell - 1] = [V - k + 1, V + k - 1]$  not taken outside this interval, whence  $g(V)$  is a terminal point of  $v$ . Since we have

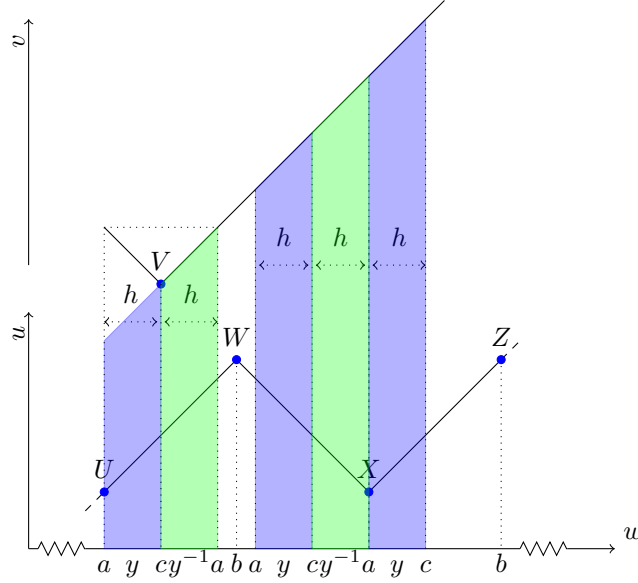


Figure 8: The walk  $g$  has a waypoint at  $V$  with  $\ell = W - V > \frac{1}{2}k$ ;  $h = k - \ell$ .

drawn  $V$  as a trough,  $g(V) = 1$ , but the same reasoning applies, with the obvious changes, if  $V$  is a peak. Write  $w[U] = a$ ,  $w[V] = c$  and  $w[W+1, V-1] = y$ . From the fact that  $W$  is a waypoint on  $f$ , we have  $w[X] = a$ , and from the fact that  $V$  is a waypoint on  $g$ , we have  $w[W+1, X-1] = y^{-1}$  and  $w[W] = a$ , whence  $w[W, X] = aycy^{-1}a$ . Using the fact that  $X$  is a waypoint on  $f$  again,  $w[X+1, X+\ell-1] = y$  and  $w[X+\ell] = c$ , whence  $w[V+1, V+2\ell] = cy^{-1}ayc$ . Recalling that  $g(V) = 1$ , we see that  $v$  has a prefix  $cy^{-1}ayc$  starting at  $V$ , contrary to the supposition that  $v$  is primitive (Lemma 3 (i)).

This deals with all cases in which the minimal leg  $[W, X]$  is *internal*. We turn finally to the few remaining cases where it is *terminal*. Without loss of generality, we may assume that we are dealing with an initial leg, i.e.  $W = 1$ ; hence  $X$  is an internal waypoint. Further, by replacing  $v$  with its reversal if necessary, we may assume that the initial leg of  $g$  is ascending. As before, let  $Z = X+k$ . By the minimality of  $[W, X]$ , we have  $Z \leq n$ . Again by the minimality of  $[W, X]$ , we have  $f([X, n]) = [1, |u|]$ . We claim that  $g([X, n]) \neq [1, |v|]$ . For otherwise, defining  $f' = f/[1, X]$ ,  $g' = g/[1, X]$  and  $w' = w[X, n]$ , we see that  $f'$  generates  $w'$  from  $u$  and  $g'$  generates  $w'$  from  $v$ , contrary to the assumption that  $w$  is the shortest counterexample. Thus,  $g$  achieves values on the interval  $[1, X-1]$  not achieved outside this interval, whence, since this is an ascending leg,  $g(1) = 1$ . It also follows that  $g$  does not have a waypoint at  $X$ , since, otherwise, by the minimality of the leg  $[W, X]$ , the interval  $[X, Z]$  is included in a leg of  $g$ , whence  $g([X, n]) = [1, |v|]$ , which we have shown is not true. Now write  $w[1] = a$ ,  $w[X] = b$  and  $w[2, X-1] = x$ . Since  $X$  is a waypoint on  $f$ , we have  $w[Z] = a$  and  $w[X+1, Z-1] = x^{-1}$ . Observe that  $g$  must have a waypoint, say  $V$ , in the interval  $[X+1, Z-1]$ , since otherwise  $v$  has a prefix  $axbx^{-1}a$ , contrary to the assumption that  $v$  is primitive (Lemma 3 (i)). By the minimality of the leg

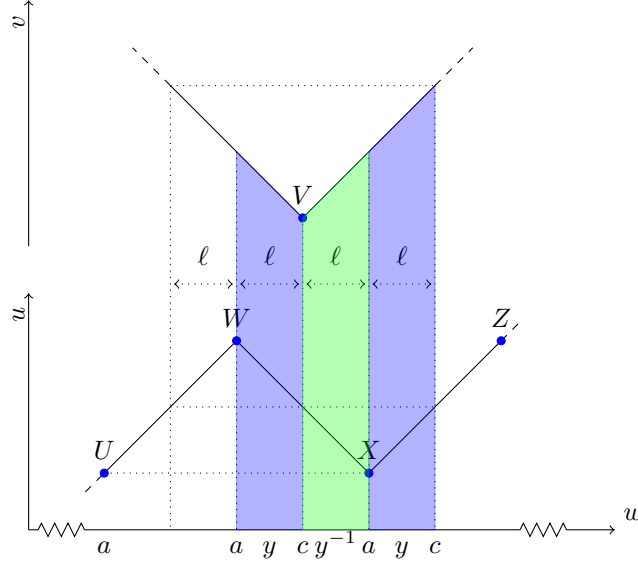


Figure 9: The walk  $g$  has a waypoint at  $V = W + \ell = X - \ell$ , where  $\ell = \frac{1}{2}k$ .

$[W, X]$ ,  $V$  is unique. Write  $c = w[V]$  and  $j = V - X$ . This time, we have just two cases.

Case (i):  $j \leq \frac{1}{2}k$ . The situation is shown in Fig. 10(a). Write  $y = w[X+1, V-1]$ , so that  $w[X, V] = w[V-j, V] = byc$ . Since  $V$  is a waypoint on  $g$ , we have  $w[V, V+j] = cy^{-1}b$ , and hence  $w[X, X+2j] = bycy^{-1}b$ . Therefore, since  $X$  is a waypoint on  $f$ , we have  $w[X-2j, X] = bycy^{-1}b$ , whence  $w[X-2j, V] = bycy^{-1}byc$ . But  $V$  is the first internal waypoint on  $g$ , whence  $v$  has an infix  $bycy^{-1}byc$ , contrary to the assumption that  $v$  is primitive (Lemma 3 (iv)).

Case (ii):  $j > \frac{1}{2}k$ . The situation is shown in Fig. 10(b). Setting  $h = k-j$ , we see that  $Z = V+h$  and  $h < j$ . Since  $V$  is a waypoint on  $g$ , we have  $w[V-h] = w[V+h] = w[Z] = a$ . Write  $y = w[V-h+1, V]$ , so that  $w[Z-2h, Z] = w[V-h+1, Z] = aycy^{-1}a$ . Therefore, since  $X$  is a waypoint on  $f$ , and the points  $W = 1$  and  $Z$  are symmetrically positioned about  $X$ , we see that also  $w[1, 2h] = ay^{-1}cya$ . But  $g$  certainly has no internal waypoint on this interval, whence  $v$  has a prefix  $ay^{-1}cya$ , again contrary to the assumption that  $v$  is primitive (Lemma 3 (ii)).

Thus, the assumption that  $w$  is a shortest word for which there exist primitive generators  $u$  and  $v$  such that  $u \neq v$  and  $u \neq v^{-1}$  leads in all cases to a contradiction. This completes the proof that no such  $w$  exists.

### Acknowledgements

This work is supported by the Polish NCN, grant number 2018/31/B/ST6/03662. The author would like to thank Vincent Michelini and Daumantas Kojelis for helpful comments.

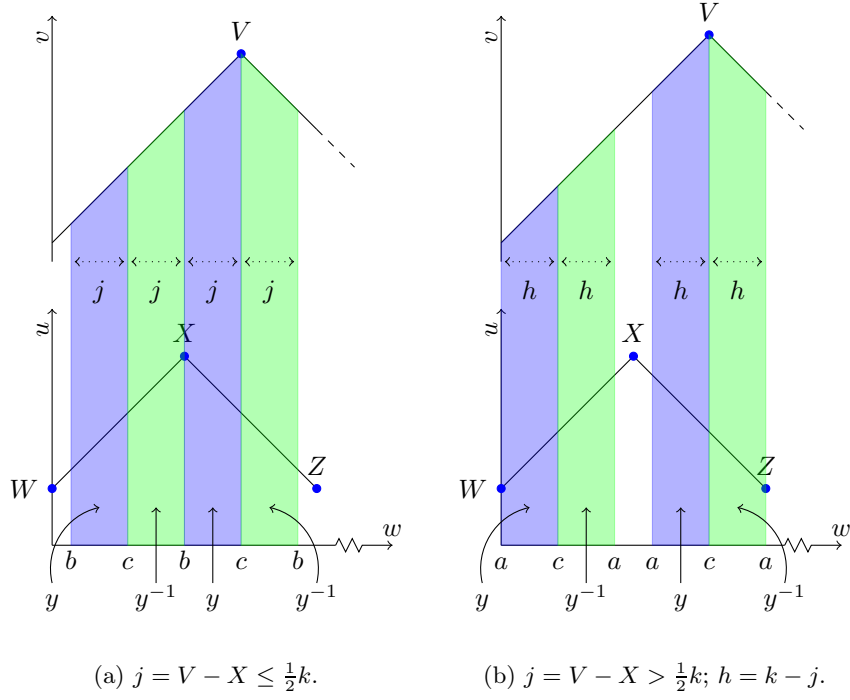


Figure 10: The walk  $g$  has its first internal waypoint at  $V \in [X+1, Z-1]$ .

## References

- [1] M. Crochemore and W. Rytter. *Jewels of stringology*. World Scientific, Singapore and River Edge, NJ, 2002.
- [2] D.E. Knuth, J.H. Morris Jr., and V.R. Pratt. Fast pattern matching in strings. 6:323–350, 1977.
- [3] Lovász László. Random walks on graphs: A survey, combinatorics, paul erdos is eighty. *Bolyai Soc. Math. Stud.*, 2, 01 1993.