NATURAL BIJECTIONS FOR CONTIGUOUS PATTERN AVOIDANCE IN WORDS

JULIA CARRIGAN, ISAIAH HOLLARS, AND ERIC ROWLAND

ABSTRACT. Two words p and q are avoided by the same number of length-n words, for all n, precisely when p and q have the same set of border lengths. However, known proofs of this result use generating functions and do not provide explicit bijections. We establish a natural bijection from the set of words avoiding p to the set of words avoiding q in the case that p and q have the same set of proper borders.

1. INTRODUCTION

Combinatorialists have studied pattern avoidance in multiple contexts. In this paper, we are interested in the avoidance of contiguous patterns in words. We say that a word w *avoids* a word p if w does not contain a contiguous occurrence of p. We refer to the word p as a *pattern*. For example, the word 010 avoids the pattern 00 but does not avoid 10. Let $\mathbb{N}$ denote the set of non-negative integers.

Definition 1.1. Let p and q be two words on a finite alphabet Σ. Define

$$A_n(p) = \{w \in \Sigma^n : \text{w avoids } p\}.$$

The words p and q are *avoidant-equivalent* if $|A_n(p)| = |A_n(q)|$ for all $n \in \mathbb{N}$.

This notion of equivalence is analogous to Wilf equivalence for non-contiguous permutation patterns, which has been studied extensively. When two permutation patterns are avoided by the same number of permutations, researchers seek a bijective explanation. See for example the survey by Claesson and Kitaev [1] of bijections between permutations that avoid 321 and permutations that avoid 132.

Analogously, when two words p, q are avoidant-equivalent, we would like a natural bijection from $A_n(p)$ to $A_n(q)$ for each n, since this provides a combinatorial explanation for the equivalence and therefore a deeper understanding of the relationship between these two structures. Since the set of all finite words avoiding p is a regular language, one can use the framework of abstract numeration systems developed by Lecomte and Rigo [5] to construct an explicit bijection $A_n(p) \to A_n(q)$ by composing a bijection $A_n(p) \to \{1, 2, \dots, |A_n(p)|\}$ with a bijection $\{1, 2, \dots, |A_n(q)|\} \to A_n(q)$. However, the use of an intermediate

This work was done in the 2022 New York City Discrete Mathematics REU and was supported by NSF award DMS-2051026.

indexing set makes this bijection somewhat arbitrary, and it is not likely to preserve combinatorial statistics of interest. In this paper, we establish a more direct bijection for certain pairs of patterns.

A sufficient condition for two patterns to be avoidant-equivalent has essentially been known since the work of Solov'ev [9, 10], who determined the expected time required for a pattern p to appear in a word built randomly letter by letter. Solov'ev showed that the expected time depends only on the lengths of the borders of p.

Definition 1.2. Let p be a word. A non-empty word x is a *border* of p if x is both a prefix and a suffix of p. Let

$$b(p) = \{|x| : x \text{ is a border of } p\}.$$

We call $b(p)$ the *border length set* of p. A non-empty word x is a *proper border* of p if x is a border of p and $x \neq p$.

Example. Let $\Sigma = \{0, 1\}$. The borders of $p = 0110$ are 0 and 0110. These borders can be thought of as the ways p can overlap itself:

$$\begin{array}{l} \boxed{0}110 \\ 011\boxed{0} \end{array} \qquad \begin{array}{l} \boxed{01}10 \\ 01\boxed{10} \end{array} \qquad \begin{array}{l} \boxed{011}0 \\ 0\boxed{110} \end{array} \qquad \begin{array}{l} \boxed{0110} \\ \boxed{0110} \end{array}$$

The border length set is $b(0110) = \{1, 4\}$. The only proper border of p is 0.

It follows from the paper of Solov'ev, and more explicitly from the work of Guibas and Odlyzko [3], that if $b(p) = b(q)$ then p and q are avoidant-equivalent. Moreover, Guibas and Odlyzko give a method for computing the generating function of the number of words avoiding a pattern (or set of patterns). Let $k = |\Sigma|$ be the size of the alphabet, let $l = |p|$, and define the polynomial $B(x) = \sum_{i \in b(p)} x^{l-i}$. Then

$$\sum_{n \geq 0} |A_n(p)| x^n = \frac{B(x)}{(1 - kx)B(x) + x^l}. \tag{1}$$

This generating function can also be obtained by the Goulden–Jackson cluster method [2]; see the treatment by Noonan and Zeilberger [6] for friendly introduction.

Example. For the word $p = 0110$, the border length set is $b(p) = \{1, 4\}$. For $q = 1011$, we have $b(q) = \{1, 4\}$ as well. Therefore $b(p) = b(q)$, and the series expansion of $\frac{x^3+1}{1-2x+x^3-x^4}$ gives the sizes of both $A_n(p)$ and $A_n(q)$ for all $n \in \mathbb{N}$. In particular, p and q are avoidant-equivalent.

The main result of this paper (Theorem 3.1) is the following. Suppose p and q are words on a finite alphabet Σ. If the set of proper borders of p is equal to the set of proper borders of q, then the map ϕ_L, which is defined in Section 2 and iteratively replaces occurrences of q with p, is a bijection from $A_n(p)$ to $A_n(q)$ for all n. Note that here the condition is that the sets of proper borders themselves are equal, as opposed to the sets of border lengths.

For words on the binary alphabet $\Sigma = \{0, 1\}$, there are 103764 pairs of length-10 avoidant-equivalent patterns, and our theorem provides a bijection for 71058 of these pairs. Additionally, there are two types of trivial bijections — left–right reversal and permutations of Σ. Compositions of all these bijections provide bijections for 103460 pairs, which is 99.7% of avoidant-equivalent pairs of length-10 patterns. See Section 3.1 and Table 1 for more data. The smallest pair of avoidant-equivalent patterns on $\{0, 1\}$ for which we do not have a natural bijection is 0010010 and 0110110, which have a border length set of $\{1, 4, 7\}$.

Example. Let $p = 1001$ and $q = 1101$. Since p and q have the same set of proper borders, namely $\{1\}$, the replacement function ϕ_L forms a bijection from $A_n(p)$ to $A_n(q)$. We would also like a bijection from $A_n(0110)$ to $A_n(q)$, since $b(0110) = \{1\} = b(q)$. The patterns 0110 and q do not have the same set of proper borders, since 0 is a border of 0110 but is not a border of q. However, if we let σ be the letter permutation function, which replaces 0's with 1's and 1's with 0's, then σ forms a bijection from $A_n(0110)$ to $A_n(p)$. Therefore the composition $\phi_L \circ \sigma$ is a bijection from $A_n(0110)$ to $A_n(q)$.

We mention that the sufficient condition $b(p) = b(q)$ for the patterns p, q to be avoidant-equivalent is also necessary. This follows from the rational generating function in Equation (1), which provides a linear recurrence satisfied by $|A_n(p)|$. Namely, let $k = |\Sigma|$ and $l = |p|$ again, and let $s(n) = |A_n(p)|$. Then

$$s(n) = k\, s(n-1) - s(n-l) + \sum_{\substack{i \in b(p) \\ i \neq l}} \Big(k\, s(n+i-l-1) - s(n+i-l) \Big).$$

Using this recurrence, one can show that if $b(p) \neq b(q)$ then the sequence $(|A_n(p)|)_{n \geq 0}$ first differs from $(|A_n(q)|)_{n \geq 0}$ at

$$n = \begin{cases} \min(|p|, |q|) & \text{if } |p| \neq |q| \\ 2|p| - \max\big(b(p) \triangle b(q)\big) & \text{if } |p| = |q| \end{cases}$$

where $\triangle$ denotes symmetric difference. Therefore, the patterns p and q are avoidant-equivalent if and only if $b(p) = b(q)$.

In Section 2, we define replacement functions ϕ_L and ϕ_R. Section 3 is dedicated to proving the main theorem, namely that ϕ_L establishes a bijection from $A_n(p)$ to $A_n(q)$ under the condition that the proper borders of p and q are identical. Finally, in Section 4, we discuss limitations of other natural options for constructing bijections, including automata and tree representations for words.

2. Replacement functions

In this section, we define the function ϕ_L that, under certain conditions, gives a bijection $\phi_L \colon A_n(p) \to A_n(q)$ in Section 3. The general idea is to systematically replace each occurrence of q in a word with p. We accomplish this with an iterative replacement process. We will define ϕ_L to take a p-avoiding word and scan from left to right looking for occurrences of q. If it finds q, it replaces the first occurrence of q with p and then starts the left-to-right scan over. The replacement process ends when no more q's remain. We will prove in Lemma 2.3 below that this process terminates.

In the following definitions, we assume that we have two patterns p and q such that $b(p) = b(q)$. In particular, $|p| = |q|$. Let $f^k(w)$ be the word obtained by iteratively applying k iterations of the function f to w.

Definition 2.1. For a given p-avoiding word w, the *single scan function* L replaces the leftmost q in w with p. If no q exists, L acts as the identity function. Define $\phi_L(w) = L^i(w)$, where i is the least non-negative integer such that $L^i(w)$ contains no q's.

Example. Let $p = 011$ and $q = 001$. The iterative replacement process of ϕ_L on the word $0001001 \in A_7(p)$ is as follows:

$$\begin{aligned} 0001001 &\overset{L}{\mapsto} 0011001 \\ = 0011001 &\overset{L}{\mapsto} 0111001 \\ = 0111001 &\overset{L}{\mapsto} 0111011 \\ &= 0111011. \end{aligned}$$

Thus, $\phi_L(0001001) = 0111011$. We have $0111011 \in A_7(q)$ as desired.

To prove that ϕ_L forms a bijection from $A_n(p)$ to $A_n(q)$, we will prove that there exists a natural inverse function ϕ_R. To this end, we define the functions R and ϕ_R, which are built to undo their counterparts L and ϕ_L.

Definition 2.2. For a given q-avoiding word w, the *single scan function* R replaces the rightmost p in w with q. If no p exists, R acts as the identity function. Define $\phi_R(w) = R^j(w)$, where j is the least non-negative integer such that $R^j(w)$ contains no p's.

Example. Using $p = 011$ and $q = 001$ as in the previous example, one checks that $\phi_R(0111011) = 0001001$, so $\phi_R(\phi_L(0001001)) = 0001001$.

Lemma 2.3. *Let p and q be equal-length patterns such that $p \neq q$, and let $n \in \mathbb{N}$. For every $w \in A_n(p)$, we have $\phi_L(w) \in A_n(q)$.*

Proof. Since $p \neq q$, either $p < q$ or $p > q$ lexicographically. Assume $p < q$, since the other case is analogous. If w contains q, then $L(w) < w$. Therefore, iteratively applying L produces lexicographically smaller words until the image no longer contains q. Since there are only finitely many length-n words on Σ, this happens after finitely many steps, at which point we have a word in $A_n(q)$. $\square$

For a word w, we define $\overline{w}$ to be the reverse of w. Let $\overline{L}$ be the function that replaces the leftmost occurrence of $\overline{p}$ in a word with $\bar{q}$. Similarly, let $\overline{R}$ be the function that replaces the rightmost $\overline{q}$ in a word with $\overline{p}$.

Definition 2.4. We now define the functions $\overline{\phi_L}\colon A_n(\bar{q}) \to A_n(\bar{p})$ and $\overline{\phi_R}\colon A_n(\bar{p}) \to A_n(\bar{q})$ in a similar fashion to ϕ_L and ϕ_R. Define $\overline{\phi_L} = \overline{L}^i(w)$, where i is the least non-negative integer such that $\overline{L}^i(w)$ contains no $\overline{p}$'s. Define $\overline{\phi_R} = \overline{R}^j(w)$, where j is the least non-negative integer such that $\overline{R}^j(w)$ contains no $\overline{q}$'s.

Lemma 2.5. *Let $w \in A_n(p)$ and $v \in A_n(q)$. We have*

$$\phi_R(v) = \overline{\overline{\phi_L}(\overline{v})} \tag{2}$$

$$\phi_L(w) = \overline{\overline{\phi_R}(\overline{w})}. \tag{3}$$

Intuitively speaking, Equation (3) says the functions ϕ_L and $\overline{\phi_R}$ are conjugate under word reversal.

Example. Let $p = 011$ and $q = 001$, and let $w = 0001001$. We will show Equation (3) holds. An example above shows the computation of $\phi_L(w) = 0111011$. Next we evaluate $\overline{\overline{\phi_R}(\overline{w})}$. Firstly, we have $\overline{w} = 1001000$. Secondly, we evaluate $\overline{\phi_R}(\overline{w})$. Recall that $\overline{\phi_R}$ will scan right to left replacing $\overline{q}$'s with $\overline{p}$'s. The iterative replacement gives

$$\begin{aligned} 1001000 &\overset{\overline{R}}{\mapsto} 1001100 \\ &= 1001100 \overset{\overline{R}}{\mapsto} 1001110 \\ &= 1001110 \overset{\overline{R}}{\mapsto} 1101110 \\ &= 1101110. \end{aligned}$$

This shows $\overline{\phi_R}(\overline{w}) = 1101110$. Since $\overline{1101110} = 0111011$, we have that $\overline{\overline{\phi_R}(\overline{w})} = 0111011$ as expected.

Proof of Lemma 2.5. We prove Equation (2) by induction on the number of replacement steps, denoted k. Then Equation (3) will follow by symmetry.

Let j be the number of steps required by ϕ_R applied to v. We set out to show

$$R^k(v) = \overline{\left(\overline{L}^k(\overline{v})\right)}. \tag{4}$$

for $0 \le k \le j$. It helps to first establish that, for any v that still has some p to replace, we have

$$R(v) = \overline{\left(\overline{L}(\overline{v})\right)}. \tag{5}$$

To see why this is true, observe that replacing the rightmost p is equivalent to

- reversing the word,

- replacing the leftmost $\bar{p}$, and then
- reversing again.

For the base case, the left-hand side of Equation (4) equals v because, when $k = 0$, there are no p's to replace in v. Similarly, $\overline{\left(\overline{L}^k(\overline{v})\right)} = \overline{(\overline{v})} = v$, because there are no $\bar{p}$'s to replace in $\overline{v}$.

Inductively, assume Equation (4) holds for some value of k where $0 \le k < j$. We have

$$\begin{aligned}
\overline{\left(\overline{L}^{k+1}(\overline{v})\right)} &= \overline{\overline{L}\left(\overline{L}^k(\overline{v})\right)} \\
&= \overline{\overline{L}\left(\overline{R^k(v)}\right)} && \text{by the inductive hypothesis} \\
&= R\left(R^k(v)\right) && \text{using Equation (5)} \\
&= R^{k+1}(v).
\end{aligned}$$

This establishes Equation (4), which gives us Equation (2). □

3. The main theorem

With all this background, we are ready for the main result of the paper.

Theorem 3.1. *Let Σ be a finite alphabet, and let p and q be distinct, equal-length words on Σ. If the set of proper borders of p is equal to the set of proper borders of q, then $\phi_L \colon A_n(p) \to A_n(q)$ forms a bijection for all $n \in \mathbb{N}$.*

For example, the set of proper borders for each of the words 0100 and 0110 is $\{0\}$. On the other hand, 0110 and 1011 do not have the same set of proper borders, despite $b(0110) = \{1, 4\} = b(1011)$.

The outline of the proof is as follows. Firstly, the case where p and q have no proper borders is slightly different from the case where p and q have one or more proper borders. Case A will be the case where they have no proper borders ("A" for "anti-border"), and Case B will be the case where they have one or more proper borders ("B" for "borders").

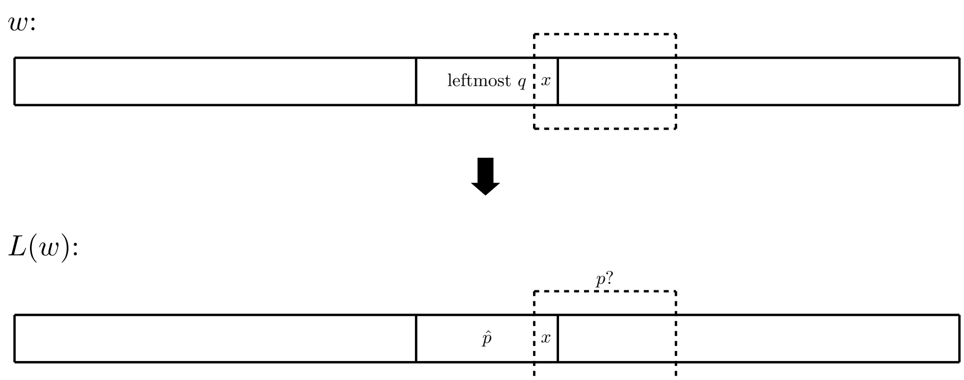
By Lemma 2.3, we have that ϕ_L is a map from $A_n(p)$ to $A_n(q)$, so it remains to show that ϕ_L is a bijection. To show bijectivity, we will show that ϕ_R is its inverse function. Thus, we will need to show that $\phi_R\left(\phi_L(w)\right) = w$ for $w \in A_n(p)$ and also that $\phi_L\left(\phi_R(w)\right) = w$ for $w \in A_n(q)$. In other words, we show that scanning from left to right and replacing q's with p's is undone by scanning from right to left and replacing p's with q's. More specifically, we show that each one-step replacement L that takes place in $\phi_L(w)$ is undone by a one-step R replacement.

Proof of Theorem 3.1. Let $n \in \mathbb{N}$ and $w \in A_n(p)$. Let i be the number of steps required by ϕ_L applied to w. We will show by induction that $L^{k-1}(w) = R\left(L^k(w)\right)$ for all k satisfying $1 \le k \le i$, so that R is the left inverse of L. It will then follow that $\phi_R\big(\phi_L(w)\big) = w$. Let $v \in A_n(q)$; then

$$\begin{aligned}
\phi_L\big(\phi_R(v)\big) &= \phi_L\big(\overline{\overline{\phi_L}(\overline{v})}\big) \quad \text{by Equation (2)} \\
&= \overline{\overline{\phi_R}\big(\overline{\phi_L}(\overline{v})\big)} \quad \text{by Equation (3), letting } w = \big(\overline{\phi_L}(\overline{v})\big) \\
&= \overline{(\overline{v})} \quad \text{because } \overline{\phi_R} \text{ is the left inverse of } \overline{\phi_L} \\
&= v.
\end{aligned}$$

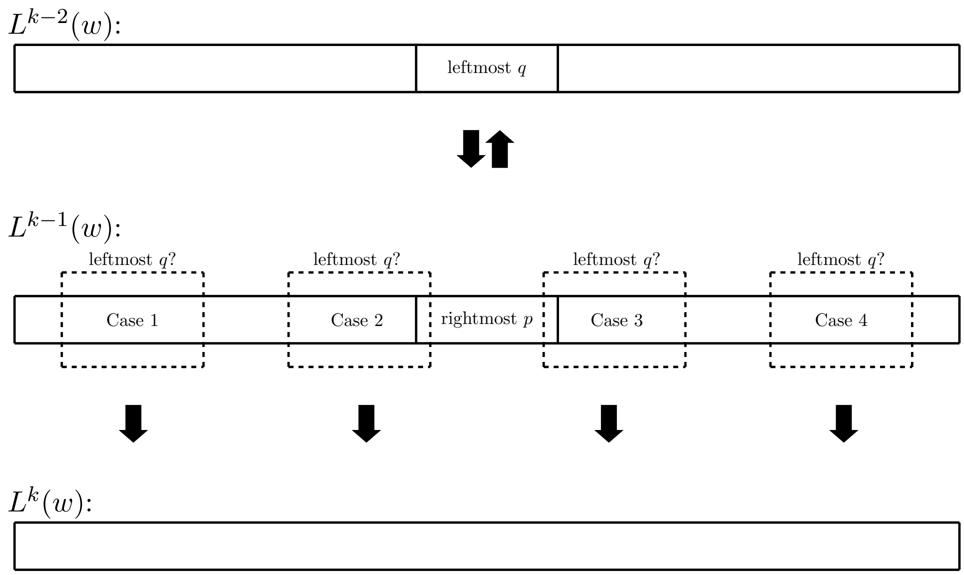
Thus, we will have also shown that $\phi_L\big(\phi_R(v)\big) = v$, so that ϕ_R is both the left inverse and right inverse of ϕ_L. It will follow that $\phi_L\colon A_n(p) \to A_n(q)$ is a bijection.

It remains to prove that $L^{k-1}(w) = R\big(L^k(w)\big)$. For the base case $k = 1$, the left-hand side of $L^{k-1}(w) = R\big(L^k(w)\big)$ is equivalent to applying zero L operations on w, so it trivially equals w. The right-hand side of this equation is $R\big(L(w)\big)$. Note that if the newly-inserted p — we commonly denote it $\hat{p}$ — in $L(w)$ is the rightmost p, then the R step function will find it first and will replace $\hat{p}$ back with a q. So, we must check that $\hat{p}$ is the rightmost p in $L(w)$. Since w is p-avoiding, the only way for $\hat{p}$ to not be rightmost is if $\hat{p}$ overlaps another p to its right in $L(w)$:

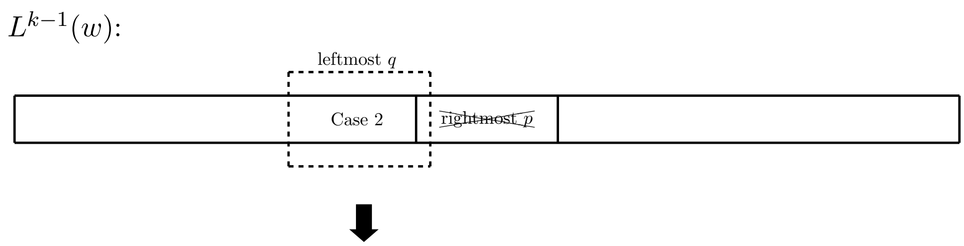


In Case A (reminder: p has no proper borders), there cannot be a p overlapping $\hat{p}$. So, we consider Case B (reminder: p and q have identical proper borders). In this case, $\hat{p}$ overlaps this other p in a border x of p. Since p and q have the same borders, this x is also in w (it wasn't changed by swapping in the $\hat{p}$). This implies that this overlapping p is also in w. This is a contradiction because w is p-avoiding. Therefore, $L^{k-1}(w) = R\big(L^k(w)\big)$ holds for $k = 1$.

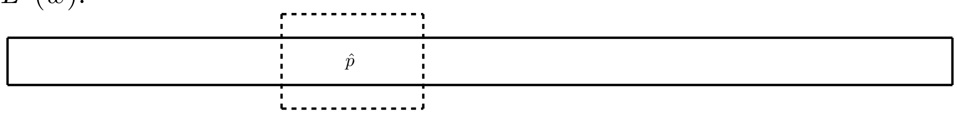
Inductively, assume that $L^{k-2}(w) = R\big(L^{k-1}(w)\big)$ for some k between 1 and the number of steps required by ϕ_L. This assumption means that once we replace the leftmost q in $L^{k-2}(w)$ with p, this new p must be the rightmost p in $L^{k-1}(w)$ because we assumed that the R function maps $L^{k-1}(w)$ back to $L^{k-2}(w)$ (the R function scans from right to left); this is indicated by an arrow in each direction in the diagram below. To show the inductive hypothesis holds for $k+1$, we need to show this same relationship holds between words $L^{k-1}(w)$ and $L^k(w)$. Thus, we wish to show that, once we replace the leftmost q in $L^{k-1}(w)$ with p, this new p in $L^k(w)$ is the rightmost. The proof is split into cases based on the position of the leftmost q in the word $L^{k-1}(w)$. For each case we demonstrate that $\hat{p}$ in $L^k(w)$ is the rightmost p, or that the case is impossible. In the rest of the proof, we refer to cases in terms of the number (1–4) and a letter (A or B). For example, in Case 2B, the 2 refers to the case where the leftmost q in $L^{k-1}(w)$ is in position 2 (see below diagram) and the B indicates that p and q have identical proper borders.



Case 1A. Since this q doesn't overlap with the rightmost p in $L^{k-1}(w)$, the q must have also existed in the previous word, $L^{k-2}(w)$. This would put it left of the leftmost q in $L^{k-2}(w)$, which is a contradiction. Therefore this case is impossible.
Case 1B. This case is also impossible due to the same reason as Case 1A.
Case 2A. Since p's do not overlap in this case, when we replace the leftmost q in $L^{k-1}(w)$ with $\hat{p}$, we will also be "destroying" the rightmost p. For the rightmost p to remain, it would have to overlap $\hat{p}$, and it doesn't. Therefore, $\hat{p}$ will be the rightmost p in $L^k(w)$. Clearly, no p can be to the right of $\hat{p}$ in $L^k(w)$ without having overlapped another p or being right of the rightmost p in $L^{k-1}(w)$.



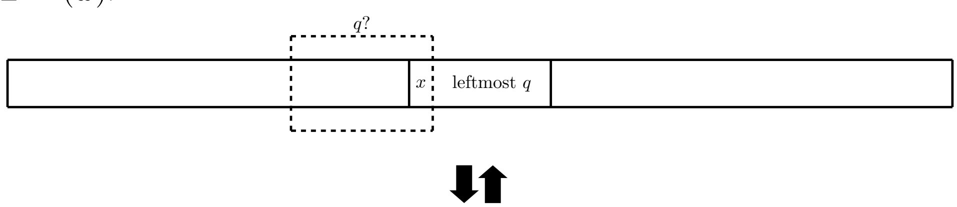
$L^k(w)$:



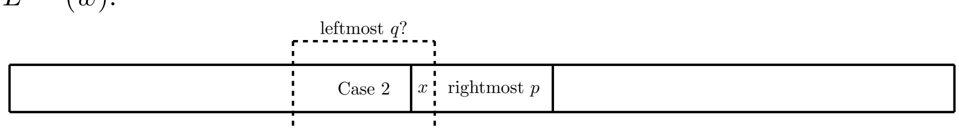
Case 2B. There are technically 2 sub-cases to cover in Case 2B. The first case assumes the leftmost q is overlapping the rightmost p in $L^{k-1}(w)$ along a border segment. This case is shown in the below diagram. The second case assumes the overlap segment is not a border. Let us first consider the case where the overlap segment is a border. In this case (similar to the base case), we must have had a q existing in the same spot in $L^{k-2}(w)$. This q would be left of the leftmost q in $L^{k-2}(w)$, so we have a contradiction.

On the other hand, if it were the case that the leftmost q in $L^{k-1}(w)$ did not overlap the rightmost p on a border segment, then upon substituting the q in $L^{k-1}(w)$ with $\hat{p}$, we will see that $\hat{p}$ becomes the rightmost p in $L^{k-2}(w)$ due to the same "destroying" phenomenon seen in Case 2A.

$L^{k-2}(w)$:

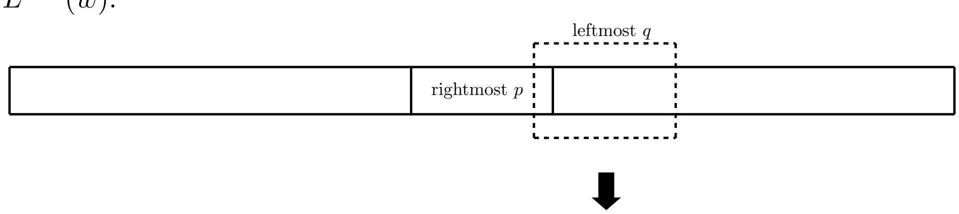


$L^{k-1}(w)$:

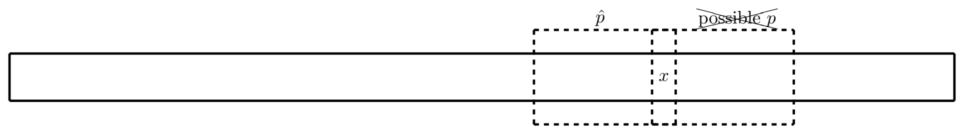


Case 3A. In this case, we have that the newly-inserted $\hat{p}$ in $L^k(w)$ is farther right than the rightmost p in the previous word $L^{k-1}(w)$. Furthermore, substituting $\hat{p}$ cannot create another p to the right because this would require the p's to overlap at some border. This contradicts the assumption of the A cases that p doesn't have any border. Therefore, $\hat{p}$ is the rightmost p in $L^k(w)$.

$L^{k-1}(w)$:

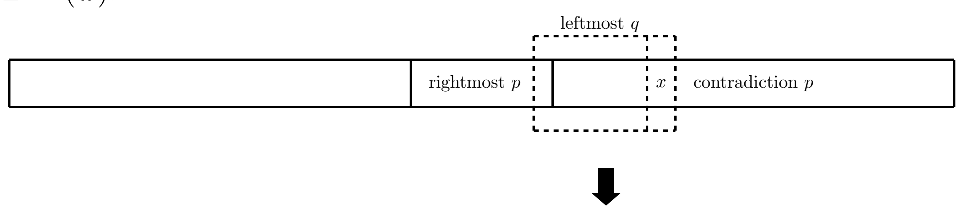


$L^k(w)$:

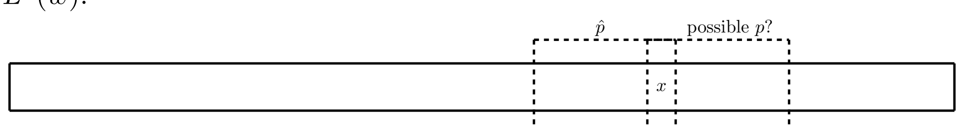


Case 3B. Similarly to Case 3A, the newly-inserted $\hat{p}$ will be the rightmost p in $L^k(w)$ as long as another p wasn't created to the right of $\hat{p}$. We will assume toward a contradiction that there does exist this other p to the right of $\hat{p}$. The words $\hat{p}$ and p must overlap in a border x. But since q and p share the same border, x wasn't changed when $\hat{p}$ was substituted in. This implies that this p to the right of $\hat{p}$ also must have existed in the previous word $L^{k-1}(w)$. This is a contradiction because this p would be existing to the right of the rightmost p in $L^{k-1}(w)$. Therefore $\hat{p}$ must be rightmost in $L^k(w)$.

$L^{k-1}(w)$:



$L^k(w)$:



Case 4A. Upon substituting in $\hat{p}$, we see that it must be the rightmost p in $L^k(w)$ by using the same argument as Case 3A.
Case 4B. The same argument as in Case 3B shows that the newly-inserted $\hat{p}$ in $L^k(w)$ is the rightmost p. $\square$

We contextualize the proof with an example and a counterexample.

Example. Let $p = 0110$ and $q = 0010$. Note that the set of proper borders for both p and q is $\{0\}$, so ϕ_L is a bijection from $A_n(p)$ to $A_n(q)$. Let $w = 1001001011 \in A_{10}(p)$. This example demonstrates how each single scan function L is undone by the function R. Observe that the first replacement aligns with the base case of the proof for Theorem 3.1, while the second replacement aligns with Case 3B. Running ϕ_L on w gives

$$\begin{aligned}1001001011 &\overset{L}{\mapsto} 1011001011\\ &= 1011001011 \overset{L}{\mapsto} 1011011011\\ &= 1011011011.\end{aligned}$$

Now we will run $\phi_L(w) = 1011011011$ through ϕ_R to see that we get w back. We also see that single scan R successfully undoes every replacement made by an L. This gives us

$$\begin{aligned} 1011011011 &\overset{R}{\mapsto} 1001011011 \\ &= 1001011011 \overset{R}{\mapsto} 1001001011 \\ &= 1001001011 = w. \end{aligned}$$

Example. We now present a short counterexample. Let $p = 1011$ and $q = 0100$. Note that $b(p) = \{1, 4\} = b(q)$, but 1 is a proper border of p and not a proper border of q. So, Theorem 3.1 does not guarantee ϕ_L will form a bijection from $A_n(p)$ to $A_n(q)$. For the word $w_1 = 0101011 \in A_7(p)$, we have

$$0101011 \overset{L}{\mapsto} 0100100 = 0100100.$$

For another word $w_2 = 1011100 \in A_7(p)$, we have

$$1011100 \overset{L}{\mapsto} 0100100 = 0100100.$$

Observe that $\phi_L(w_1) = 0100100 = \phi_L(w_2)$ so that ϕ_L does not provide a bijection.

3.1. **How many bijections do we obtain?** One might wonder how many of the possible bijections ϕ_L provides. We know that ϕ_L forms a bijection from $A_n(p)$ to $A_n(q)$ if $b(p) = \{|p|\} = b(q)$. Words with no proper borders, such as these, are known as *borderless* words. The density of borderless words on a finite alphabet has been analyzed in detail. Silberger [8] first discovered a recursive formula to count borderless words, and Holub & Shallit [4] investigated the probability that a random word is borderless. Notably, a long binary word p chosen randomly has $\approx 27\%$ chance of being borderless and $\approx 30\%$ chance of having the border length set $\{1, |p|\}$. The function ϕ_L provides a bijection for all borderless pairs and almost half of the pairs whose border length set is $\{1, |p|\}$. These cases alone account for a sizable chunk of possible avoidant-equivalent word pairs, which is why the percentage of pairs for which we have natural bijections is so high.

Table 1 contains data on the number of pairs of patterns on $\Sigma = \{0, 1\}$ for which we have a natural bijection. The "Equivalent pairs" column gives the total number of unordered pairs of patterns p and q for which $b(p) = b(q)$. The second column counts pairs for which ϕ_L establishes a bijection. Additionally, if we allow compositions with the reversal function and letter permutation function, we are able to obtain even more bijections; these pairs are counted in the third column.

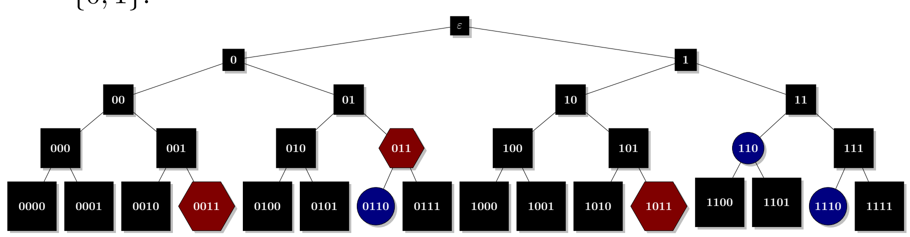
4. Other possible bijections

In this section we briefly discuss two promising approaches to finding bijections that nonetheless do not work, the purpose being to spare the reader from rediscovering these dead ends. In the effort of coming up

Pattern length	Pairs ϕ_L bijects	Pairs bijected by compositions	Equivalent pairs
1	1	1	1
2	1	2	2
3	6	8	8
4	21	32	32
5	88	120	120
6	312	460	460
7	1212	1708	1716
8	4649	6764	6780
9	18264	26072	26168
10	71058	103460	103764
11	279946	403836	405404
12	1107836	1613132	1618556

TABLE 1. Summary of bijections between patterns on $\{0, 1\}$.

FIGURE 1. Tree representation of words ending with **011** (hexagons) and words ending with **110** (circles) on $\Sigma = \{0, 1\}$.



with a universal bijection, it is natural to search for alternate ways to represent words or to encode them. Two structures we investigated were trees (Section 4.1) and finite automata (Section 4.2). We outline how these encodings work and offer up some natural algorithms and their pitfalls.

4.1. **Tree restructuring.** The set of finite words on Σ naturally has the structure of an infinite, ordered $|\Sigma|$-ary tree, where the parent of each non-empty length-n word is its length-$(n-1)$ prefix. The idea is that natural bijections from words avoiding p to words avoiding q might correspond to natural operations on this tree (for example, an automorphism).

For example, Figure 1 shows the first several levels in the tree of binary words. Let $p = 011$ and $q = 110$. We have marked nodes ending with 011 as hexagons and nodes ending with 110 as circles. Since $b(p) = \{3\} = b(q)$, the patterns p and q are avoidant-equivalent, so on each level the number of hexagon nodes is equal to the number of circle nodes. We would like an operation that swaps their positions. A natural attempt is to try permuting certain sibling nodes. Unfortunately, this does not always work.

For example, two nodes that start as aunt and niece will remain so after any sibling permutation. This is a nice property, but it is too restrictive to accomplish our goals. Observe that circle nodes 110 and 1110 have an aunt–niece relationship. On the other hand, hexagon 011 doesn't have a hexagon niece. Therefore, a sibling permutation will be unable to align the hexagons into the circle positions of the tree.

More complicated structural replacements were considered by Rowland [7], but these do not provide the bijections we seek either.

4.2. **Finite automata.** Since the set of p-avoiding words on Σ forms a regular language, it is tempting to consider the minimal finite automaton accepting this language. For example, let $\Sigma = \{0, 1\}$. The automata that recognize words avoiding 0010 and 0110 are shown in Figure 2 and Figure 3 respectively. In both diagrams, q_4 is the only fail state. Any words ending in the fail state will be rejected, since they contain the pattern we are trying to avoid.

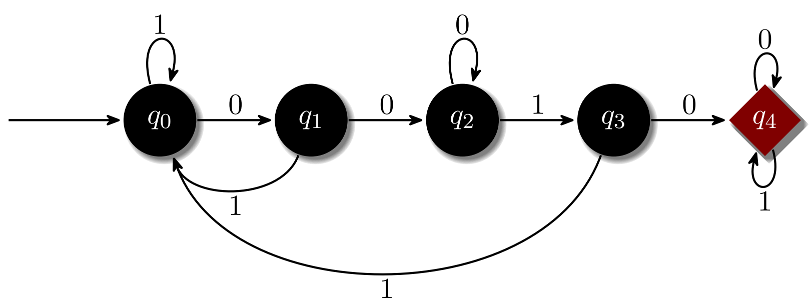


FIGURE 2. Automaton accepting words avoiding 0010.

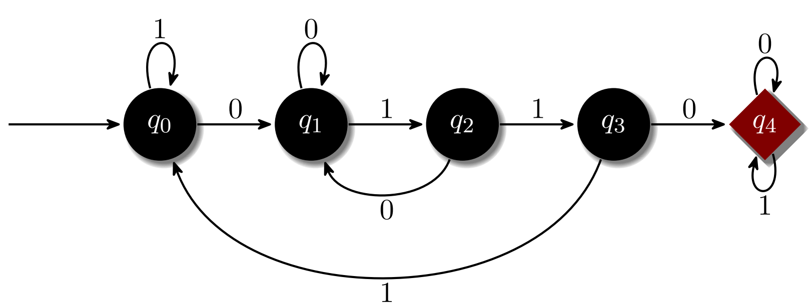


FIGURE 3. Automaton accepting words avoiding 0110.

Note that these two patterns are avoidant-equivalent (they both have the border length set $\{1, 4\}$). Every word can be associated with a unique sequence of states, which is generated by tracing the word through the automaton. For example, the word 010 generates the state sequence q_0, q_1, q_0, q_3

when traced through the automaton in Figure 2. We have now translated the problem of mapping words to a problem of mapping sequences of states. Unfortunately, it seems this translation is no magic bullet. One glaring issue is that certain state sequences are possible for the automaton in Figure 2 that aren't possible for the automaton in Figure 3. For example, we saw that 010 generates the state sequence q_0, q_1, q_0, q_3 for the first automaton; however, no word generates this state sequence in the second automaton since there is no edge from q_1 to q_0. This fact rules out the possibility of a bijection algorithm that preserves state sequences.

References

[1] Anders Claesson and Sergey Kitaev, Classification of bijections between 321- and 132-avoiding permutations, *Séminaire Lotharingien de Combinatoire* **60** (2008) Article B60d.

[2] Ian Goulden and David Jackson, An inversion theorem for cluster decompositions of sequences with distinguished subsequences, *Journal of the London Mathematical Society* **(2) 20** (1979) 567–576.

[3] Leo J. Guibas and Andrew M. Odlyzko, String overlaps, pattern matching, and nontransitive games, *Journal of Combinatorial Theory, Series A* **30** (1981) 183–208.

[4] Štěpán Holub and Jeffrey Shallit, Periods and borders of random words, *Symposium on Theoretical Aspects of Computer Science* (2016).

[5] Pierre B. A. Lecomte and Michel Rigo, Numeration systems on a regular language, *Theory of Computing Systems* **34** (2001) 27–44.

[6] John Noonan and Doron Zeilberger, The Goulden–Jackson cluster method: extensions, applications and implementations, *Journal of Difference Equations and Applications* **5** (1999) 355–377.

[7] Eric Rowland, Pattern avoidance in binary trees, *Journal of Combinatorial Theory, Series A* **117** (2010) 741–75.

[8] D. M. Silberger, How many unbordered words?, *Commentationes Mathematicae (Prace Matematyczne)* **22** (1980) 143–145.

[9] A. D. Solov'ev, A combinatorial identity and its application to the problem about the first occurrence of a rare event, *Teoriya Veroyatnostei i ee Primeneniya* **11** (1966) 313–320.

[10] A. D. Solov'ev, A combinatorial identity and its application to the problem concerning the first occurrence of a rare event (translated from Russian by N. A. Brunswick), *Theory of Probability and Its Applications* **11** (1966) 276–282.

Mathematics Department, Occidental College, Los Angeles, CA, USA

Mathematics Department, Belmont University, Nashville, TN, USA

Department of Mathematics, Hofstra University, Hempstead, NY, USA