The sigma-sequence and occurrences of some patterns, subsequences and subwords

SERGEY KITAEV

Matematik
Chalmers tekniska högskola och Göteborgs universitet
S-412 96 Göteborg
Sweden
kitaev@math.chalmers.se

Abstract

We consider sigma-words, which are words used by Evdokimov in the construction of the sigma-sequence [A. A. Evdokimov, Mat. Zametki 6 (1969), 309–319]. We then find the number of occurrences of certain patterns, subsequences and subwords in these words.

1 Introduction and Background

We write permutations as words \( \pi = a_1a_2 \cdots a_n \), whose letters are distinct and usually consist of the integers 1, 2, \ldots, \( n \).

An occurrence of a pattern \( p \) in a permutation \( \pi \) is “classically” defined as a subsequence in \( \pi \) (of the same length as the length of \( p \) ) whose letters are in the same relative order as those in \( p \). Formally speaking, for \( r \leq n \), we say that a permutation \( \sigma \) in the symmetric group \( S_n \) has an occurrence of the pattern \( p \in S_r \) if there exist \( 1 \leq i_1 < i_2 < \cdots < i_r \leq n \) such that \( p = \sigma(i_1)\sigma(i_2)\cdots\sigma(i_r) \) in reduced form. The reduced form of a permutation \( \sigma \) on a set \( \{j_1, j_2, \ldots, j_r\} \), where \( j_1 < j_2 < \cdots < j_r \), is a permutation \( \sigma_1 \) obtained by renaming the letters of the permutation \( \sigma \) so that \( j_i \) is renamed \( i \) for all \( i \in \{1, \ldots, r\} \). For example, the reduced form of the permutation 3651 is 2431. The first case of classical patterns studied was that of permutations avoiding a pattern of length 3 in \( S_3 \). Knuth [13] found that, for any \( \tau \in S_3 \), the number \( |S_n(\tau)| \) of \( n \)-permutations avoiding \( \tau \) is \( C_n \), the \( n \)th Catalan number. Later, Simion and Schmidt [16] determined the number \( |S_n(P)| \) of permutations in \( S_n \) simultaneously avoiding any given set of patterns \( P \subseteq S_3 \).

In [1] Babson and Steingrímsson introduced generalized permutation patterns that allow the requirement that two adjacent letters in a pattern must be adjacent in the permutation. In order to avoid confusion we write a “classical” pattern, say 231, as 2-3-1, and if we write, say 2-31, then we mean that if this pattern occurs in the permutation, then the letters in the permutation that correspond to 3 and 1 are...
adjacent. For example, the permutation $\pi = 516423$ has only one occurrence of the pattern 2-31, namely the subword 564, whereas the pattern 2-3-1 occurs, in addition, in the subwords 562 and 563. A motivation for introducing these patterns in [1] was the study of Mahonian statistics. A number of interesting results on generalized patterns were obtained in [6]. Relations to several well studied combinatorial structures, such as set partitions, Dyck paths, Motzkin paths and involutions, were shown there.

Burstein [2] considered words instead of permutations. In particular, he found the number $|k^n(P)|$ of words of length $n$ in a $k$-letter alphabet that avoid all patterns from a set $P \subseteq S_3$ simultaneously. Burstein and Mansour [3] (resp. [4, 5]) considered forbidden patterns (resp. generalized patterns) with repeated letters.

The most attention, in the papers on classical or generalized patterns, is paid to finding exact formulas and/or generating functions for the number of words or permutations avoiding, or having $k$ occurrences of, certain patterns. In [11] the authors suggested another problem, namely counting the number of occurrences of certain patterns in certain words. These words were chosen to be the set of all finite approximations of a sequence generated by a morphism with certain restrictions. A motivation for this choice was the interest in studying classes of sequences and words that are defined by iterative schemes [14, 15]. In [12] the authors also studied the number of occurrences of certain patterns in certain words. But there they choose these words to be the subdivision stages from which the Peano curve is obtained. The authors called these words the Peano words. The Peano curve was studied by the Italian mathematician Giuseppe Peano in 1890 as an example of a continuous space filling curve.

In the present paper we consider the sigma-words, which are words used by Evdokimov in construction of the sigma-sequence [7]. Evdokimov used this sequence to construct chains of maximal length in the $n$-dimensional unit cube. Independent interest to the sigma-sequence appears in connection with the well-known Dragon curve, discovered by physicist John E. Heighway and defined as follows: we fold a sheet of paper in half, then fold in half again, and again, etc. and then unfold in such a way that each crease created by the folding process is opened out into a 90-degree angle. The “curve” refers to the shape of the partially unfolded paper as seen edge on. If one travels along the curve, some of the creases will represent turns to the left and others turns to the right. Now if 1 indicates a turn to the right, and 2 to the left, and we start travelling along the curve indicating the turns, we get the sigma-sequence [8]. In [10] the sigma-sequence was studied from another point of view. It was proved there that this sequence cannot be defined by iterated morphism.

Since the sigma-sequence $w_\sigma$ is a sequence in a 2-letter alphabet, we consider patterns in 2-letter alphabets. Moreover, the patterns in a 1-letter alphabet (for example 1-1-1) correspond to two subsequences (for this example, these subsequences are 1-1-1 and 2-2-2), whereas the patterns in a 2-letter alphabet (with at least one letter 2) uniquely determine the subsequences in $w_\sigma$ that correspond to them, and conversely. For example, an occurrence of the pattern 1-2-1 is an occurrence of the subsequence 1-2-1, whereas an occurrence of the subsequence (subword) 211 is an occurrence of the pattern 211. Thus, any of our results for a pattern, can be
interpreted in term of subsequences or subwords, depending on the context, and conversely.

In our paper we give either an explicit formula or recurrence relation for the number of occurrences for some classes of patterns, subwords and subsequences in the sigma-words. In particular, Theorem 4, allows to find the number of occurrences of an arbitrary generalized pattern without internal dashes of length $\ell$, provided we know four certain numbers that can be easily calculated for the sigma-words $C_k$, $D_k$, $C_{k+1}$ and $D_{k+1}$ (to be defined below), where $k = \lceil \log_2 \ell \rceil$. Theorem 9 gives a recurrence relation for counting occurrences of patterns of the form $\tau_1-\tau_2$. In Section 6 we discuss occurrences of patterns of the form $\tau_1-\tau_2-\cdots-\tau_k$, where the pattern $\tau_i$ does not overlap with the patterns $\tau_{i-1}$ and $\tau_{i+1}$ for $i = 1, 2, \ldots, k-1$. Finally, Section 7 deals with patterns of the form $[\tau_1-\tau_2-\cdots-\tau_k]$, $[\tau_1-\tau_2-\cdots-\tau_k]$ and $(\tau_1-\tau_2-\cdots-\tau_k)$ in Babson and Steingrímsson notation, which means that we use $"[x"]$ in a pattern $p$ to indicate that in an occurrence of $p$, the letter corresponding to the $x$ must be the first letter of a word under consideration, whereas if we use "$y$"", we mean that the letter corresponding to $y$ must be the last (rightmost) letter in the word.

2 Preliminaries

In [7], Evdokimov constructed chains of maximal length in the $n$-dimensional unit cube using the sigma-sequence. The sigma-sequence $w_\sigma$ was defined there by the following inductive scheme:

\[
\begin{align*}
C_1 &= 1, & D_1 &= 2, \\
C_{k+1} &= C_k D_k, & D_{k+1} &= C_k 2D_k
\end{align*}
\]

and $w_\sigma = \lim_{k \to \infty} C_k$. Thus, the initial letters of $w_\sigma$ are 1121121122121212. We call the words $C_k$ the sigma words. The first four values of the sequence \{C_k\}_{k \geq 1} are 1, 112, 1121122, 112112211122122. In [10] an equivalent definition of $w_\sigma$ was given: any natural number $n \neq 0$ can be presented unambiguously as $n = 2^t(4s + \sigma)$, where $\sigma < 4$, and $t$ is the greatest natural number such that $2^t$ divides $n$. If $n$ runs through the natural numbers then $\sigma$ runs through some sequence consisting of 1 and 3. If we substitute 3 by 2 in this sequence, we get $w_\sigma$.

In this paper we count occurrences of patterns in the sigma-words, which are particular initial subwords of $w_\sigma$. However, the challenging question is to find the number of occurrences of patterns or subwords in an arbitrary initial subword of $w_\sigma$, or more generally, in a subword of $w_\sigma$ starting in the position $i$ and ending in the position $j$.

It turns out that for counting occurrences of certain patterns or subwords in $C_n$, one needs to know the number of occurrences of certain patterns in $D_n$. So, in the most cases, we give results for both $C_n$ and $D_n$. However, our main purpose is the words $C_n$ for $n \geq 1$, and in some propositions and examples we do not consider $D_n$. 

SIGMA-SEQUENCE AND OCCURRENCES OF SOME PATTERNS 189
In what follows, we give initial values for the words $C_i$ and $D_i$:

\begin{align*}
C_1 &= 1 & D_1 &= 2 \\
C_2 &= 112 & D_2 &= 122 \\
C_3 &= 1121122 & D_3 &= 1122122 \\
C_4 &= 112112211122122 & D_4 &= 11211221122122 \\
C_5 &= 1121122111221221112112221122122 & D_5 &= 1121122111221222112112221122122
\end{align*}

We now give some other definitions.

A descent (resp. rise) in a word $w = a_1a_2\ldots a_n$ is an $i$ such that $a_i > a_{i+1}$ (resp. $a_i < a_{i+1}$). It follows from the definitions that an occurrence of a descent (resp. rise) is an occurrence of the pattern 21 (resp. 12).

Let $c_n^\tau$ (resp. $d_n^\tau$) denote the number of occurrences of the pattern $\tau$ in $C_n$ (resp. $D_n$).

Suppose a word $W = AaB$, where $A$ and $B$ are some words of the same length, and $a$ is a letter. We define the kernel of order $k$ for the word $W$ to be the subword consisting of the $k - 1$ rightmost letters of $A$, the letter $a$, and the $k - 1$ leftmost letters of $B$. We denote it by $K_k(W)$. For example, $K_3(11211221) = 12112$. If $|A| < k - 1$ then we assume $K_k(W) = \epsilon$, that is, the kernel in this case is the empty word. Also, $m_k(\tau, W)$ denotes the number of occurrences of the pattern (or the word, or the subsequence depending on the context) $\tau$ in $K_k(W)$.

We denote $x-x-\ldots-x$ ($\ell$ times) by $x^\ell$. Also, $\lceil a \rceil$ denotes the least natural number $b$ such that $a \leq b$.

## 3 Patterns 1-1-\ldots-1, 1-2 and 2-1

It is easy to see that $|C_n| = |D_n| = 2^n - 1$. The following lemma gives the number of the letters 1 and 2 in $C_n$ and $D_n$.

**Lemma 1.** The number of 1s (resp. 2s) in $C_n$ is $2^{n-1}$ (resp. $2^{n-1} - 1$). The number of 1s (resp. 2s) in $D_n$ is $2^{n-1} - 1$ (resp. $2^{n-1}$).

**Proof.** It is enough to find the number of 1s $c_n$ and $d_n$ in $C_n$ and $D_n$ respectively, since the number of 2s in $C_n$ and $D_n$ are obviously equal to $|C_n| - c_n$ and $|D_n| - d_n$ respectively.

It is easy to see from the structure of $C_n$ and $D_n$ that

\[
\begin{aligned}
c_n &= c_{n-1} + d_{n-1} + 1, \\
d_n &= c_{n-1} + d_{n-1},
\end{aligned}
\]

together with $c_1 = 1$ and $d_1 = 0$. The solution to this recurrence is $c_n = 2^{n-1}$ and $d_n = 2^{n-1} - 1$. 

\[\square\]
Proposition 2. The number occurrences of the subsequence $1^k$ (resp. $2^k$) in $C_n$ is \( \binom{2n-1}{k} \) (resp. \( \binom{2n-2}{k} \)). Thus, the number of occurrences of the pattern $1^k$ in $C_n$ is equal to

\[
c_n^k = \binom{2n-1}{k} + \binom{2n-1}{k} = 2^n - k \binom{2n-1}{k}.
\]

Proof. From Lemma 1, there are $2^{n-1}$ (resp. $2^{n-1} - 1$) occurrences of the letter 1 (resp. 2) in $C_n$, and thus there are \( \binom{2n-1}{k} \) (resp. \( \binom{2n-1}{k} \)) occurrences of the subsequence $1^k$ (resp. $2^k$) there.

Proposition 3. We have that for all $n \geq 2$, $c_n^{1-2} = d_n^{1-2} = 2 \cdot 4^{n-2} + (n - 2) \cdot 2^{n-2}$, and $c_n^{2-1} = d_n^{2-1} = 2 \cdot 4^{n-2} - n \cdot 2^{n-2}$.

Proof. Let us first consider the pattern 1-2. An occurrence of this pattern in $C_n = C_{n-1}D_{n-1}$ is either inside $C_{n-1}$, or inside $D_{n-1}$, or the letter 1 is from the word $C_{n-1}$, whereas the letter 2 is from the word $D_{n-1}$. Thus

\[
c_n^{1-2} = c_n^{1-2} + d_n^{1-2} + \{ \text{the number of 1s in } C_{n-1} \} \cdot \{ \text{the number of 2s in } D_{n-1} \}.
\]

Using the same considerations for $D_n = C_{n-1}2D_{n-1}$, one can get

\[
d_n^{1-2} = c_n^{1-2} + d_n^{1-2} + \{ \text{the number of 1s in } C_{n-1} \} \cdot \{ \text{the number of 2s in } D_{n-1} \} + 1.
\]

The number of 1s and 2s in $C_{n-1}$ and $D_{n-1}$ is given in Lemma 1. So,

\[
\begin{align*}
\begin{pmatrix}
c_n^{1-2} \\
d_n^{1-2}
\end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix}
c_{n-1}^{1-2} \\
d_{n-1}^{1-2}
\end{pmatrix} + \begin{pmatrix} 2^{n-2} \cdot (2^{n-2} + 1) \\ 2^{n-2} \cdot (2^{n-2} + 1) \end{pmatrix} \\
\end{align*}
\]

\((1)\)

together with $c_2^{1-2} = 2$ and $d_2^{1-2} = 2$. Here, and several times in what follows, we need to solve recurrence relations of the form

\[
x_n = Ax_{n-1} + b,
\]

where $A$ is a matrix, and $x_n, x_{n-1}$ and $b$ are some vectors, where $b$ sometimes depends on $n$. We recall from linear algebra that such relations can be solved by diagonalization of the matrix $A$, that is, by writing $A = VDV^{-1}$, where $D$ is a diagonal matrix consisting of eigenvalues of $A$, and the columns of $V$ are eigenvectors of $A$. For example, if $A$ is a $2 \times 2$ matrix that consists of 1s, then we use

\[
\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix}
\]

for computing powers of $A$, and thus for solving the recurrence relations. For the recurrence 1, we get that for all $n \geq 2$, $c_n^{1-2} = d_n^{1-2} = 2 \cdot 4^{n-2} + (n - 2) \cdot 2^{n-2}$.
In the same manner, we can get that for the pattern 2-1,
\[
\begin{align*}
c_n^{2-1} &= c_{n-1}^{2-1} + d_{n-1}^{2-1} + 2^{n-2} \cdot (2^{n-2} - 1), \\
d_n^{2-1} &= c_{n-1}^{2-1} + d_{n-1}^{2-1} + 2^{n-2} \cdot (2^{n-2} - 1),
\end{align*}
\]
together with \(c_3^{2-1} = 2\) and \(d_3^{2-1} = 2\). This gives, that for all \(n \geq 2\), \(c_n^{2-1} = d_n^{2-1} = 2 \cdot 4^{n-2} - n \cdot 2^{n-2}\).

Proposition 3 shows that asymptotically, the numbers of occurrences of the patterns, or the subsequences, 1-2 and 2-1 in \(C_n\) or \(D_n\) are equal.

4 Patterns without internal dashes

Recall the definitions in Section 2.

**Theorem 4.** Let \(\tau = \tau_1 \tau_2 \ldots \tau_\ell\) be an arbitrary generalized pattern without internal dashes that consists of 1s and 2s. Suppose \(k = \lfloor \log_2 \ell \rfloor\), \(a = m_\ell(\tau, D_k 1C_k)\), and \(b = m_\ell(\tau, D_k 2C_k)\). Then for \(n > k + 1\), we have
\[
\begin{align*}
c_n = (a + b + c_{k+1} + d_{k+1}) \cdot 2^{n-k-2} - b, \\
d_n = (a + b + c_{k+1} + d_{k+1}) \cdot 2^{n-k-2} - a.
\end{align*}
\]

**Proof.** Suppose \(n > k + 1\). In this case, \(C_n = C_{n-1} D_{n-1} = W_1 K_\ell(1C_k) W_2\), for some words \(W_1\) and \(W_2\) such that \(|W_1| = |W_2|\). Because of the definition of the kernel \(K_\ell(D_k 1C_k)\), an occurrence of the pattern \(\tau\) in \(C_n\) is in either \(C_{n-1}\), or \(D_{n-1}\), or \(K_k(D_k 1C_k)\) (from the definitions \(|C_{n-1} \cap K_k(D_k 1C_k)| = |D_{n-1} \cap K_k(D_k 1C_k)| = \ell - 1\) and thus these intersections cannot be an occurrence of \(\tau\). So,
\[
c_n = c_{n-1} + d_{n-1} + a,
\]
whereas in the same way, we can obtain that
\[
d_n = c_{n-1} + d_{n-1} + b.
\]
By solving these recurrence relations, we get the desirable result.

In particular, Theorem 4 is valid for \(\ell = 1\), in which case the number of occurrences of \(\tau\) in \(C_n\) (or \(D_n\)) is the number of letters in \(C_n\) (or \(D_n\)). Indeed, in this case, \(k = 0\), \(a = b = c_1 = d_1 = 1\), hence \(c_n = d_n = 2^n - 1 = |C_n| = |D_n|\). Also, as a corollary to Theorem 4 we have, that if \(a = b = c_{k+1} = d_{k+1} = 0\) for some pattern \(\tau\), then this pattern never appears in sigma-sequence.

All of the following examples are corollaries to Theorem 4.
Example 5. Suppose $\tau = 12$. We have that $k = 1$, $a = m_2(12, D_1 C_1) = 0$ and $b = m_2(12, D_1 C_2) = 0$. Besides, $c_{12}^1 = 1$ and $d_{12}^1 = 1$. Thus using Theorem 4, for all $n > 2$, $c_{12}^n = 2^{n-2}$. So, the number of rises in $C_n$ is equal to $2^{n-2}$, for $n \geq 2$.

If $\tau = 21$, again $k = 1$, but now $a = m_2(21, D_1 C_1) = 1$ and $b = m_2(21, D_1 C_2) = 1$. Besides, $c_{31}^1 = 1$ and $d_{31}^1 = 1$. From Theorem 4, for all $n > 3$, $c_{31}^n = 2^{n-2} - 1$, which shows that the number of descents in $C_n$ is one less than the number of rises.

Since in both cases $a = b$, using the recurrences in Theorem 4, we have that $c_{12}^n = d_{12}^n = 2^{n-2}$, whereas $c_{31}^n = d_{31}^n = 2^{n-2} - 1$.

Example 6. Suppose $\tau = 122$. We have that $k = 2$, $a = m_3(112, D_2 C_2) = 0$, and $b = m_3(112, D_2 C_2) = 0$. Besides, $c_{12}^1 = 2$ and $d_{12}^1 = 1$. Now, from Theorem 4, we have that for all $n > 3$, $c_{12}^n = d_{12}^n = 3 \cdot 2^{n-4}$.

Example 7. Suppose $\tau = 221$. We have that $k = 2$, $a = m_3(221, D_2 C_2) = 1$, and $b = m_3(221, D_2 C_2) = 1$. Besides, $c_{21}^2 = 0$ and $d_{21}^2 = 1$. Now, from Theorem 4, we have that for all $n > 3$, $c_{21}^n = d_{21}^n = 3 \cdot 2^{n-4} - 1$.

Example 8. If $\tau = 2212221$ then $k = 3$, $a = m_7(221, D_3 C_3) = 0$, $b = m_7(221, D_3 C_3) = 1$, $c_{4212221} = 0$, and $d_{4212221} = 0$. Thus for $n \geq 4$, $c_{4212221} = 2^{n-4} - 1$.

5 Patterns of the form $\tau_1 - \tau_2$

Theorem 9. Let $p = \tau_1, \tau_2$ be a generalized pattern such that $|\tau_1| = k_1$ and $|\tau_2| = k_2$. Suppose $k = [\log_2(k_1 + k_2 - 1)]$. The following denote the number of occurrences of the subwords $\tau_1$ and $\tau_2$ in the certain kernels: $a_{\tau_1} = m_{k_1}(\tau_1, D_k C_k)$, $a_{\tau_2} = m_{k_2}(\tau_2, D_k C_k)$, $b_{\tau_1} = m_{k_1}(\tau_1, D_k C_k)$, and $b_{\tau_2} = m_{k_2}(\tau_2, D_k C_k)$. Also, let $r_{\tau_1}^a$ (resp. $r_{\tau_2}^a$, $r_{\tau_1}^b$, $r_{\tau_2}^b$) denote the number of occurrences of overlapping subwords $\tau_1$ and $\tau_2$ in the word $D_k C_k$ (resp. $D_k C_k$, $D_k C_k$, $D_k C_k$), where $\tau_1 \in K_{k_1}(D_k C_k)$ and $\tau_2 \in C_k$ (resp. $\tau_1 \in D_k C_k$ and $\tau_2 \in K_{k_1}(D_k C_k)$, $\tau_1 \in K_{k_1}(D_k C_k)$ and $\tau_2 \in C_k$, $\tau_1 \in D_k C_k$ and $\tau_2 \in K_{k_2}(D_k C_k)$). Besides, we assume that we know $c_n^{\tau_1}$ and $d_n^{\tau_2}$ for $n > n_i$, $i = 1, 2$. Then for $n > \max(k + 1, n_1 + 1, n_2 + 1)$, $c_n^{\tau_1}$ and $d_n^{\tau_2}$ are given by the following recurrence:

\[
\begin{pmatrix}
  c_n^{\tau_1} \\
  d_n^{\tau_2}
\end{pmatrix} = \begin{pmatrix}
  1 & 1 \\
  1 & 1
\end{pmatrix} \begin{pmatrix}
  c_{n-1}^{\tau_1} \\
  d_{n-1}^{\tau_2}
\end{pmatrix} + \begin{pmatrix}
  \alpha_n \\
  \beta_n
\end{pmatrix},
\]

where

\[
\alpha_n = (c_{n-1}^{\tau_1} + a_{\tau_1} - r_{\tau_1}^a)d_{n-1}^{\tau_2} + (a_{\tau_2} - r_{\tau_2}^a)c_{n-1}^{\tau_1},
\]

and

\[
\beta_n = (c_{n-1}^{\tau_1} + b_{\tau_1} - r_{\tau_1}^b)d_{n-1}^{\tau_2} + (b_{\tau_2} - r_{\tau_2}^b)c_{n-1}^{\tau_1}.
\]

Proof. Suppose $n > \max(k + 1, n_1 + 1, n_2 + 1)$. Let us find a recurrence for the number $c_n^{\tau_1}$ (one can use the same considerations for $d_n^{\tau_2}$).

An occurrence of the pattern $\tau$ in $C_n = C_{n-1} D_{n-1}$ is either inside $C_{n-1}$, or inside $D_{n-1}$, or begins in $C_{n-1}$ or the letter 1 between $C_{n-1}$ and $D_{n-1}$ and ends in $D_{n-1}$ or
the letter 1. The first two cases obviously give \( c_{n-1} \) and \( d_{n-1} \) occurrences of \( \tau \). To count the contribution of the last to cases, we work with words instead of patterns. We do it to take in account the situations when \( \tau_1 \) or \( \tau_2 \) consists of copies of only one letter. In this case, we cannot count occurrence of these patterns separately, and then use this information, since, for instance, occurrences of the pattern \( \tau_1 = 111 \) are subwords 111 and 222 (the last one of these subwords we do not need), whereas occurrences of the pattern \( \tau_1 = 222 \) are not defined at all (222 is not a pattern).

If an occurrence of \( \tau_1 - \tau_2 \) does not entirely belong to \( C_{n-1} \) or \( D_{n-1} \) then we only have one of the following possibilities:

(a) the subword \( \tau_1 \) entirely belongs to \( C_{n-1} \) and the subword \( \tau_2 \) entirely belongs to \( D_{n-1} \);

(b) the subword \( \tau_1 \) belongs entirely to \( C_{n-1} \) and the subword \( \tau_2 \) belongs to the kernel \( K_{k_2}(D_k 1 C_k) \), where \( k = \lceil \log_2 (k_1 + k_2 - 1) \rceil \) is the least number that allow to control, in \( C_n \) \((n > k)\), overlapping occurrences of subwords \( \tau_1 \) and \( \tau_2 \) where \( \tau_1 \) is entirely from \( C_{n-1} \) and \( \tau_2 \in K_{k_2}(D_k 1 C_k) \);

(c) the subword \( \tau_2 \) belongs entirely to \( D_{n-1} \) and the subword \( \tau_1 \) belongs to the kernel \( K_{k_3}(D_k 1 C_k) \).

In (a) we obviously have \( c_{n-1}^{\tau_1} \cdot d_{n-1}^{\tau_2} \) possibilities.

In (b) we have \( c_{n-1}^{\tau_1} \cdot a_{\tau_2} - c_{n-1}^{\tau_1} \cdot r_{\tau_2}^n \) possibilities, since we need to subtract those occurrences of \( \tau_1 \) and \( \tau_2 \) that overlap.

Similarly to (b), in (c) we have \( d_{n-1}^{\tau_2} \cdot a_{\tau_1} - d_{n-1}^{\tau_2} \cdot r_{\tau_1}^n \) possibilities, which completes the proof.

\[ \square \]

**Remark 10.** For using Theorem 9, one needs to know \( c_{\tau}^n \) and \( d_{\tau}^n \) for patterns \( \tau \) without internal dashes. These numbers could be obtained by using Theorem 4.

The following corollary to Theorem 9 is straightforward to prove, using the fact that for non-overlapping patterns \( \tau_1 \) and \( \tau_2 \), \( r_{\tau_1}^n = r_{\tau_2}^n = r_{\tau_1}^b = r_{\tau_2}^b = 0 \).

**Corollary 11.** We make the same assumptions as those in Theorem 9. Suppose additionally that the words \( \tau_1 \) and \( \tau_2 \) are not overlapping in the following sense: no one suffix of \( \tau_1 \) is a prefix of \( \tau_2 \). Then for \( n > \max(k + 1, n_1 + 1, n_2 + 1) \), \( c_{\tau}^n \) and \( d_{\tau}^n \) are given by the same recurrence as that in Theorem 9 with

\[
\alpha_n = (c_{n-1}^{\tau_1} + a_{\tau_1}) d_{n-1}^{\tau_2} + a_{\tau_2} c_{n-1}^{\tau_1}
\]

and

\[
\beta_n = (c_{n-1}^{\tau_1} + b_{\tau_1}) d_{n-1}^{\tau_2} + b_{\tau_2} c_{n-1}^{\tau_1}.
\]

**Remark 12.** Corollary 11 is valid under weaker assumptions, namely we only need the property of non-overlapping of the patterns \( \tau_1 \) and \( \tau_2 \) when one of them is in its kernel and the other one is not in its kernel. Example 15 deals with the pattern \( \tau \) that has overlapping blocks \( \tau_1 \) and \( \tau_2 \), but Corollary 11 can be applied. However,
from practical point of view, checking the fact if two subwords are non-overlapping
is more easy than considering the kernels and checking the non-overlapping of the
subwords there.

**Example 13.** Suppose $\tau = 12-21$. We have that $|\tau_1| = |\tau_2| = 2$. Now, in the
statement of Theorem 9 we have that $k = 2, a_{\tau_1} = 0, a_{\tau_2} = 1, b_{\tau_1} = 0$ and $b_{\tau_2} = 1$.
Also, since there are no overlapping occurrences of the subwords 12 and 21 in $K_3(1221112)$ and $K_3(1222112)$, we have $r^a_1 = 0, r^a_2 = 0, r^b_1 = 0$ and $r^b_2 = 0$. Besides,
from example 5, $c_{3}^{12} = d_{3}^{12} = 2^{n-2}$ and $c_{3}^{21} = d_{3}^{21} = 2^{n-2} - 1$. Thus, $\alpha_n = \beta_n = 4^n - 3$.
Using the fact that $c_3^{12-21} = 0$ and $d_3^{12-21} = 1$, this allows us to get an explicit formula
for $c_n^{12-21}$ and $d_n^{12-21}$ for $n > 3$:

$$c_n^{12-21} = d_n^{12-21} = \frac{1}{2} 4^{n-2} - 3 \cdot 2^{n-4}.$$ 

In particular $c_5^{12-21} = 5$.

**Example 14.** Suppose $\tau = 1-221$. We have that $|\tau_1| = 1$ and $|\tau_2| = 3$. Moreover,
the words $\tau_1$ and $\tau_2$ are not overlapping, hence we can use Corollary 11. We have
that $k = 2, a_{\tau_1} = 1, a_{\tau_2} = 1, b_{\tau_1} = 0$ and $b_{\tau_2} = 1$. From example 7, $d_n^{221} = 3 \cdot 2^{n-4} - 1$.
Also, the number of occurrences of the letter 1 (the subword $\tau_1 = 1$) is given by
Lemma 1: $c_n^1 = 2^{n-1}$. So, $\alpha_n = 6 \cdot 4^n - 3 \cdot 2^{n-5} - 1$ and $\beta_n = 6 \cdot 4^n - 4$.
One can get
now an explicit formula for $c_n^{1-221}$ and $d_n^{1-221}$ for $n > 4$:

$$c_n^{1-221} = \frac{1}{2} 4^{n-2} + 27 \cdot 2^{n-5} - n - 7,$$
$$d_n^{1-221} = \frac{1}{2} 4^{n-2} + 21 \cdot 2^{n-5} - 8.$$ 

In particular, $c_5^{1-221} = 47$.

**Example 15.** Suppose $\tau = 112-21$. We have that $|\tau_1| = k_1 = 3$ and $|\tau_2| = k_2 = 2$.
The other parameters in Theorem 9 are $k = 3, a_{\tau_1} = 0, a_{\tau_2} = 1, b_{\tau_1} = 0, b_{\tau_2} = 1,$
$r^a_1 = r^a_2 = r^b_1 = r^b_2 = 0$. From Example 6, for $n \geq 4$, $c_n^{112} = 3 \cdot 2^{n-4}$, and from
Example 5, $d_n^{21} = 2^{n-2} - 1$. Thus, in Theorem 9, $\alpha_n = \beta_n = c_n^{112}(d_n^{21} + 1) = 3 \cdot 4^{n-4}$.
Now, we solve the recurrence relation from the theorem to get, that for $n > 3$

$$c_n^{112-21} = d_n^{112-21} = \frac{3}{2} 4^{n-3} - 2^{n-4}.$$ 

### 6 Counting occurrences of $\tau_1\tau_2\cdots\tau_k$

In this section we study the number of occurrences of a pattern $\tau = \tau_1\tau_2\cdots\tau_k$,
where $\tau_i$ are patterns without internal dashes. We say that $\tau$ consists of $k$ blocks.
We assume that for $i = 1, 2, \ldots, k - 1$, the pattern $\tau_i$ does not overlap with the
patterns $\tau_{i-1}$ and $\tau_{i+1}$. In this case we give a recurrence relation for the number of
occurrences of $\tau$, provided we know the number of occurrences of certain patterns
consisting of less than, or equal to, $k - 1$ blocks, as well as $2k$ certain numbers
which can be calculated by considering the words $D_1C_\ell$ and $C_\ell 2D_\ell$, where $\ell$ is the
maximum number such that \( \ell \leq \max_i [\log_2 |\tau_i|] \). The cases of \( k = 1 \) and \( k = 2 \) are studied in the previous sections; they give the bases for our calculations. However, the case of overlapping patterns \( \tau_i \) is not solved, and it remains as a challenging problem, since an answer to this problem gives the way to count occurrences of an arbitrary generalized pattern, or an arbitrary subsequence, in \( \sigma \)-words.

**Theorem 16.** Let \( \tau = \tau_1 \tau_2 \cdots \tau_k \) be a generalized pattern such that \( |\tau_i| = k_i \) for \( i = 1, 2, \ldots, k \). We assume that for \( i = 1, 2, \ldots, k-1 \), the subword \( \tau_i \) does not overlap with the subwords \( \tau_{i-1} \) and \( \tau_{i+1} \) in the following sense: no one suffix of \( \tau_i \) is a prefix of \( \tau_{i-1} \) and no one suffix of \( \tau_i \) is a prefix of \( \tau_{i+1} \). Suppose \( \ell_i = [\log_2 k_i] \), \( \ell = \max_i \ell_i \), and for the subwords \( \tau_i \) we have \( a_i = m_{k_i}(\tau_i, D\ell_i C \ell_i) \) and \( b_i = m_{k_i}(\tau_i, D\ell_i 2C \ell_i) \), for \( i = 1, 2, \ldots, k \). We assume that we know \( c^*_n = c^*_{n-1} \) and \( d^*_n = d^*_{n-1} \) for each \( 1 \leq i \leq k-1 \) and for all \( n > n^* \). Then for all \( n > \max(\ell + 1, n^* + 1) \), \( c^*_n \) and \( d^*_n \) are given by the following recurrence:

\[
\begin{pmatrix}
  c^*_n \\
  d^*_n
\end{pmatrix} =
\begin{pmatrix}
  1 & 1 \\
  1 & 1
\end{pmatrix}
\begin{pmatrix}
  c^*_n \\
  d^*_n
\end{pmatrix} + \sum_{i=1}^{k-1} \begin{pmatrix}
  c^*_n \\
  d^*_n
\end{pmatrix}
\begin{pmatrix}
  c^*_{n-1} \cdots c^*_{i+1} d^*_{n-1} \cdots d^*_{i+1} \\
  c^*_{n-1} \cdots c^*_{i+1} d^*_{i+1} \cdots d^*_{k}
\end{pmatrix}
+ \sum_{i=1}^{k} \begin{pmatrix}
  a_i \cdot c^*_{n-1} \cdots c^*_{i-1} d^*_{n-1} \cdots d^*_{i+1} \\
  b_i \cdot c^*_{n-1} \cdots c^*_{i-1} d^*_{i+1} \cdots d^*_{k}
\end{pmatrix}.
\]

**Proof.** We consider only \( c^*_n \), since the same arguments can be applied to \( d^*_n \). We use the considerations similar to those in Theorem 9.

An occurrence of the pattern \( \tau \) in \( C_n = C_{n-1} D_{n-1} \) can be entirely in \( C_n \) or \( D_n \). The first term counts such occurrences. Otherwise, we have two possibilities: either the letter 1 between the words \( C_{n-1} \) and \( D_{n-1} \) does not belong to an occurrence of \( \tau \), or it does do it, in which case there exist \( i \) (exactly one) such that the subword \( \tau_i \) occurs in its kernel. The first sum in the statement is obviously responsible for the first of this cases, whereas the second sum is responsible for the second case (in the last case we use the fact that subwords \( \tau_i \) are not overlapping). \( \square \)

As a corollary to Theorem 16, we have Corollary 11.

The following example is another corollary to Theorem 16.

**Example 17.** Suppose \( \tau = 2 - 1 - 221 \), that is, \( \tau_1 = 2 \), \( \tau_2 = 1 \) and \( \tau_3 = 221 \). So, parameters in Theorem 16 are the following: \( k_1 = k_2 = 1 \), \( k_3 = 3 \), \( \ell_1 = \ell_2 = 1 \), \( \ell_3 = 2 \), \( \ell = 2 \). From \( D_1 C_1 = 211 \) we obtain \( a_1 = 0 \), \( a_2 = 1 \). From \( D_2 C_2 = 1221112 \) we obtain \( a_3 = 1 \). From \( D_1 2C_1 = 221 \) we get \( b_1 = 1 \), \( b_2 = 0 \). From \( D_2 2C_2 = 1222112 \) we get \( b_3 = 1 \). Besides, from Proposition 3, Examples 7 and 14, we have

\[
\begin{align*}
  c^*_n & = c^*_2 = 2^n - 2 \cdot 4^{n-2} - n \cdot 2^{n-2}, \text{ for } n > 1; \\
  d^*_n & = d^*_n = 3 \cdot 2^{n-4} - 1, \text{ for } n > 3; \\
  d^*_n & = d^*_n = \frac{1}{2} \cdot 4^{n-2} + 21 \cdot 2^{n-5} - 8, \text{ for } n > 4.
\end{align*}
\]

Also, the number of occurrences of the subword \( \tau_1 = 2 \) in \( C_n \) is given by Proposition 2: \( c^*_n = c^*_n = 2^n - 1 \). So, the number of occurrences of the pattern \( \tau \) in \( C_n \) and \( D_n \),
for \( n > 5 \), satisfies the following recurrence relation:
\[
\begin{pmatrix}
    c_n^\tau \\
    d_n^\tau
\end{pmatrix}
= \begin{pmatrix}
    1 & 1 \\
    1 & 1
\end{pmatrix}
\begin{pmatrix}
    c_{n-1}^\tau \\
    d_{n-1}^\tau
\end{pmatrix}
+ \left( \frac{5}{1024} s^n + \frac{25-3n}{256} 4^n - \frac{171}{64} 2^n + 9 \right) \left( \frac{5}{1024} s^n + \frac{21-3n}{256} 4^n - 2^{n+1} \right),
\]
with initial conditions \( c_0^\tau = 70 \) and \( d_0^\tau = 74 \).

7 Patterns of the form \([\tau_1\tau_2\cdots\tau_k], [\tau_1\tau_2\cdots\tau_k] \) and \((\tau_1\tau_2\cdots\tau_k)\)

We recall that according to Babson and Steingrímsson notation for generalized patterns, if we use \( \tau \) in a pattern, for example if we write \( p = [1-2] \), we indicate that in an occurrence of \( p \), the letter corresponding to the 1 must be the first letter of a word under consideration, whereas if we write, say, \( p = (1-2] \), then the letter corresponding to 2 must be the last (rightmost) letter of the word.

In the theorems of this section, we assume that we can find the numbers \( c_n^{\tau_1}\tau_2\cdots\tau_k \) and \( d_n^{\tau_1}\tau_2\cdots\tau_k \) for any patterns \( \tau_i, i = 1, 2, \ldots, k \), without internal dashes. For certain special cases, these numbers can be obtained using the theorems of Sections 5 and 6.

**Theorem 18.** Suppose \( \tau_1 \) and \( \tau_2 \) are two patterns without internal dashes such that \( |\tau_1| = k_1 \) and \( |\tau_2| = k_2 \). Also, suppose \( \ell_1 = \log_2(k_1 + 1), \ell_2 = \log_2(k_2 + 1) \) and \( \ell = \log_2(k_1 + k_2 + 1) \). Let \( a(\tau_1, \tau_2) \) be the number of overlapping subwords \( \tau_1 \) and \( \tau_2 \) in \( C_\ell \) such that \( \tau_1 \) occurs as \( k_1 \) leftmost letters of \( C_\ell \); \( b(\tau_1, \tau_2) \) is the number of overlapping subwords \( \tau_1 \) and \( \tau_2 \) in \( C_\ell \) such that \( \tau_2 \) occurs as \( k_2 \) rightmost letters of \( C_\ell \). We assume that we know \( c_n^{\tau_1} \) and \( d_n^{\tau_2} \) for \( i = 1, 2 \) and for all \( n > n^* \).

i. For \( n \geq \max(\ell_1, n^*) \),
\[
c_n^{[\tau_1\tau_2]} = \begin{cases}
0, & \text{if } C_{\ell_1} \text{ does not begin with } \tau_1, \\
c_n^{\tau_2} - a(\tau_1, \tau_2), & \text{otherwise}.
\end{cases}
\]

ii. For \( n \geq \max(\ell_2, n^*) \),
\[
c_n^{(\tau_1\tau_2)} = \begin{cases}
0, & \text{if } C_{\ell_2} \text{ does not end with } \tau_2, \\
c_n^{\tau_1} - b(\tau_1, \tau_2), & \text{otherwise}.
\end{cases}
\]

iii. For \( n \geq \ell \),
\[
c_n^{[\tau_1\tau_2]} = \begin{cases}
0, & \text{if } C_\ell \text{ does not begin with } \tau_1 \text{ or end with } \tau_2, \\
1, & \text{otherwise}.
\end{cases}
\]

iv. For \( n \geq \max(\ell_1, n^*) \),
\[
d_n^{[\tau_1\tau_2]} = \begin{cases}
0, & \text{if } D_{\ell_1} \text{ does not begin with } \tau_1, \\
d_n^{\tau_2} - a(\tau_1, \tau_2), & \text{otherwise}.
\end{cases}
\]
v. For $n \geq \max(\ell_2, n^*)$,
\[
d_{n}^{(\tau_1 - \tau_2)} = \begin{cases} 0, & \text{if } D_{\ell_2} \text{ does not end with } \tau_2, \\ d_{n}^{\tau_1} - b(\tau_1, \tau_2), & \text{otherwise.} \end{cases}
\]

vi. For $n \geq \ell$,
\[
d_{n}^{[\tau_1 - \tau_2]} = \begin{cases} 0, & \text{if } D_{\ell} \text{ does not begin with } \tau_1 \text{ or end with } \tau_2, \\ 1, & \text{otherwise.} \end{cases}
\]

Proof. We prove case i, all the other cases are then easy to see.

Clearly, if $C_{\ell_1}$ does not begin with $\tau_1$ then $C_n$ does not begin with $\tau_1$ for all $n \geq \ell_1$, which means that $c_n^{\tau_1 - \tau_2} = 0$ in this case. Otherwise, to count occurrences of the pattern $[\tau_1 - \tau_2]$ is the same as to find the number of occurrences of the pattern $\tau_2$ in $C_n$ and then subtract the number of such occurrences of $\tau_2$ that begin from the $i$-th letter of $C_n$, where $1 \leq i \leq k_1$.

The following two examples are corollaries to Theorem 18.

Example 19. Suppose we have the patterns $\sigma_1 = [1122 - 2111]$ and $\sigma_2 = (21221 - 12]$. From Theorem 18, $c_n^{\sigma_1} = d_n^{\sigma_1} = 0$, since $C_3$ does not begin with 1122 ($\ell_1 = 3$). Also, $c_n^{\sigma_2} = d_n^{\sigma_2} = 0$, since $C_3$ does not end with 12 ($\ell_2 = 3$).

Example 20. Suppose $\tau = [112-21)$. We have that $k_1 = 3$, $\ell_1 = 2$ and $C_2$ begins with the subword 112. Besides, $a(112, 21) = 1$ and, from Example 5, $c_n^{21} = d_n^{21} = 2^{n-2} - 1$. Theorem 18 now yields, that for $n > 3$, we have $c_n^{[112-21]} = c_n^{\tau_2} - a(\tau_1, \tau_2) = 2^{n-2} - 2$.

The following theorem is straightforward to prove under the assumption that certain subwords do not overlap.

Theorem 21. Let $\{\tau_1, \tau_2, \ldots, \tau_k\}$ be a set of generalized patterns without internal dashes. Suppose $|\tau_1| = s_1$, $|\tau_k| = s_k$, $\ell_1 = \log_2(s_1 + 1)$ and $\ell_k = \log_2(s_k + 1)$. Also, $\ell = \max(\ell_1, \ell_k)$.

i. With the assumption that the subword $\tau_1$ does not overlap with the subword $\tau_2$, that is, no one suffix of $\tau_1$ is a prefix of $\tau_2$, we have

(a)\[
c_n^{\tau_1 - \tau_2 \cdots - \tau_k} = \begin{cases} 0, & \text{if } C_{\ell_1} \text{ does not begin with } \tau_1, \\ c_n^{\tau_2 - \tau_3 \cdots - \tau_k}, & \text{otherwise.} \end{cases}
\]

(b)\[
d_n^{\tau_1 - \tau_2 \cdots - \tau_k} = \begin{cases} 0, & \text{if } D_{\ell_1} \text{ does not begin with } \tau_1, \\ d_n^{\tau_2 - \tau_3 \cdots - \tau_k}, & \text{otherwise.} \end{cases}
\]
ii. With assumption that the subword \( \tau_{k-1} \) does not overlap with the subword \( \tau_k \), that is, no one suffix of \( \tau_{k-1} \) is a prefix of \( \tau_k \), we have

\[
(c) \quad c_n^{[\tau_1-\tau_2-\cdots-\tau_k]} = \begin{cases} 
0, & \text{if } C_{\ell_k} \text{ does not end with } \tau_k, \\
\frac{\tau_1-\tau_2-\cdots-\tau_{k-1}}{c_n}, & \text{otherwise.}
\end{cases}
\]

\[
(b) \quad d_n^{[\tau_1-\tau_2-\cdots-\tau_k]} = \begin{cases} 
0, & \text{if } D_{\ell_k} \text{ does not end with } \tau_k, \\
\frac{\tau_1-\tau_2-\cdots-\tau_{k-1}}{d_n}, & \text{otherwise.}
\end{cases}
\]

iii. With the assumption that the subword \( \tau_1 \) does not overlap with the subword \( \tau_2 \), and the subword \( \tau_{k-1} \) does not overlap with the subword \( \tau_k \), we have

\[
(a) \quad c_n^{[\tau_1-\tau_2-\cdots-\tau_k]} = \begin{cases} 
0, & \text{if } C_{\ell_1} \text{ does not begin with } \tau_1 \text{ or does not end with } \tau_k, \\
\frac{\tau_2-\tau_3-\cdots-\tau_{k-1}}{c_n}, & \text{otherwise.}
\end{cases}
\]

\[
(b) \quad d_n^{[\tau_1-\tau_2-\cdots-\tau_k]} = \begin{cases} 
0, & \text{if } D_{\ell_1} \text{ does not begin with } \tau_1 \text{ or does not end with } \tau_k, \\
\frac{\tau_2-\tau_3-\cdots-\tau_{k-1}}{d_n}, & \text{otherwise.}
\end{cases}
\]

The following example is a corollary to Theorem 21.

**Example 22.** Suppose \( \tau = [112-1-221-22] \). The parameters of Theorem 21 are \( k_1 = 3, k_2 = 2, \ell_1 = 2, \ell_2 = 2, \ell = 2 \). \( C_3 \) begins with the subword 112 and ends with the subword 22. Thus by Theorem 21 and Example 14, \( c_n^{[112-1-221-22]} = c_n^{1-221} = \frac{1}{2}n^4 - 27 \cdot 2n - 5 - n - 7 \).

**References**

[1] E. Babson and E. Steingrímsson, Generalized permutation patterns and a classification of the Mahonian statistics, Sém. Lothar. Combin. 44 (2000), Art. B44b, 18 pp.

[2] A. Burstein, Enumeration of words with forbidden patterns, Ph.D. Thesis, University of Pennsylvania, (1998).

[3] A. Burstein and T. Mansour, Words restricted by patterns with at most 2 distinct letters, Electron. J. Combin. 9(2) (2002), #R3.
[4] A. Burstein and T. Mansour, Words restricted by 3-letter generalized multipermutation patterns, preprint CO/0112281.

[5] A. Burstein and T. Mansour, Counting occurrences of some subword patterns, preprint CO/0204320.

[6] A. Claesson: Generalised Pattern Avoidance, European J. Combin. 22 (2001), 961–971.

[7] A. A. Evdokimov, The maximal length of a chain in the unit n-dimensional cube, Mat. Zametki 6 (1969), 309–319. (In Russian.)

[8] Private communication (2001).

[9] B. Gelbaum and J. Olmsted, Counterexamples in Analysis, Holden-day, San Francisco, London, Amsterdam, (1964).

[10] S. Kitaev, There are no iterated morphisms that define the Arshon sequence and the sigma-sequence, J. Automata, Languages and Combinatorics (2002), to appear.

[11] S. Kitaev and T. Mansour, Counting the occurrences of generalized patterns in words generated by a morphism, preprint CO/0210170.

[12] S. Kitaev and T. Mansour, The Peano curve and counting occurrences of some patterns, preprint CO/0210268.

[13] D. E. Knuth, The Art of Computer Programming, 2nd ed. Addison Wesley, Reading, MA, (1973).

[14] M. Lothaire, Combinatorics on Words, Encyclopedia of Mathematics, Vol. 17, Addison-Wesley (1986). Reprinted in the Cambridge Mathematical Library, Cambridge University Press, Cambridge UK (1997).

[15] A. Salomaa, Jewels of Formal Language Theory, Computer Science Press (1981).

[16] R. Simion and F. Schmidt, Restricted permutations, European J. Combin. 6 (1985), 383–406.

[17] S. V. Yablonsky, Discrete mathematics and mathematical problems of cybernetics, Nauka, Vol. 1, Moscow (1974), 112–116. (In Russian.)

(Received 19 Nov 2002)