COUNTING THE OCCURRENCES OF GENERALIZED PATTERNS IN WORDS GENERATED BY A MORPHISM

Sergey Kitaev and Toufik Mansour [1]

Matematik, Chalmers tekniska högskola och Göteborgs universitet,
S-412 96 Göteborg, Sweden

kitaev@math.chalmers.se, toufik@math.chalmers.se

ABSTRACT

We count the number of occurrences of certain patterns in given words. We choose these words to be the set of all finite approximations of a sequence generated by a morphism with certain restrictions. The patterns in our considerations are either classical patterns 1-2, 2-1, 1-1-····-1, or arbitrary generalized patterns without internal dashes, in which repetitions of letters are allowed. In particular, we find the number of occurrences of the patterns 1-2, 2-1, 12, 21, 123 and 1-1-····-1 in the words obtained by iterations of the morphism $1 \to 123$, $2 \to 13$, $3 \to 2$, which is a classical example of a morphism generating a nonrepetitive sequence.

1. INTRODUCTION AND BACKGROUND

We write permutations as words $\pi = a_1 a_2 \cdots a_n$, whose letters are distinct and usually consist of the integers $1, 2, \ldots, n$.

An occurrence of a pattern p in a permutation π is "classically" defined as a subsequence in π (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 σ in the symmetric group $\mathcal{S}_n$ has an occurrence of the pattern $p \in \mathcal{S}_r$ if there exist $1 \leq i_1 < i_2 < \cdots < i_r \leq n$ such that $p = \sigma(i_1)\sigma(i_2)\ldots\sigma(i_r)$ in reduced form. The *reduced form* of a permutation σ on a set $\{j_1, j_2, \ldots, j_r\}$, where $j_1 < j_2 < \cdots < j_r$, is a permutation σ_1 obtained by renaming the letters of the permutation σ 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 $\mathcal{S}_3$. Knuth [Knuth] found that, for any $\tau \in \mathcal{S}_3$, the number $|\mathcal{S}_n(\tau)|$ of n-permutations avoiding τ is C_n, the nth Catalan number. Later, Simion and Schmidt [SimSch] determined the number $|\mathcal{S}_n(P)|$ of permutations in $\mathcal{S}_n$ simultaneously avoiding any given set of patterns $P \subseteq \mathcal{S}_3$.

In [BabStein] Babson and Steingrímsson introduced *generalised 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 [BabStein] was the study of Mahonian statistics. A number of interesting results on generalised patterns were obtained in [Claes]. Relations to several well studied combinatorial structures, such as set partitions, Dyck paths, Motzkin paths and involutions, were shown there.

Burstein [Burstein] considered words instead of permutations. In particular, he found the number $|[k]^n(P)|$ of words of length n in k-letter alphabet that avoid each pattern from a set $P \subseteq \mathcal{S}_3$ simultaneously. Burstein and Mansour [BurMans1] (resp. [BurMans2, BurMans3]) considered forbidden patterns (resp. generalized patterns) with repeated letters.

The most attention, in the papers on classical or generalized patterns, is paid to counting exact formulas and/or generating functions for the number of words or permutations avoiding, or having k occurrences of, certain pattern. In this paper we suggest another problem, namely counting the number of occurrences of a particular pattern τ in given words. We choose these words to be a set of all finite approximations (to be defined below) of a sequence generated by a morphism with certain restrictions. A motivation for such a choice is big interest in studying classes of sequences and words that are defined by iterative schemes [Lothaire, Salomaa]. The pattern τ in our considerations is either a classical pattern from the set $\{1\text{-}2, 2\text{-}1, 1\text{-}1\text{-}\cdots\text{-}1\}$, or an arbitrary generalized pattern without internal dashes, in which repetitions of letters are allowed. In particular, we find that there are $(3 \cdot 4^{n-1} + 2^n)$ occurrences of the pattern 1-2 in the n-th finite approximation of the sequence w defined below, which is a classical example of a nonrepetitive sequence.

Let Σ be an alphabet and $\Sigma^\star$ be the set of all words of Σ. A map $\varphi : \Sigma^\star \to \Sigma^\star$ is called a *morphism*, if we have $\varphi(uv) = \varphi(u)\varphi(v)$ for any $u, v \in \Sigma^\star$. It is easy to see that a morphism φ can be defined by defining $\varphi(i)$ for each $i \in \Sigma$. The set of all rules $i \to \varphi(i)$ is called a *substitution system*. We create words by starting with a letter from the alphabet Σ and iterating the substitution system. Such a substitution system is called a *D0L (Deterministic, with no context Lindenmayer) system* [LindRoz]. D0L systems are classical objects of formal language theory. They are interesting from mathematical point of view [Frid], but also have applications in theoretical biology [Lind]. Let $|X|$ denote the length of a word X, that is the number of letters in X.

Suppose a word $\varphi(a)$ begins with a for some $a \in \Sigma$, and that the length of $\varphi^k(a)$ increases without bound. The symbolic sequence $\lim_{k \to \infty} \varphi^k(a)$ is said to be *generated* by the morphism φ. In particular, $\lim_{k \to \infty} \varphi^k(a)$ is a *fixed point* of φ. However, in this paper we are only interesting in the *finite approximations* of $\lim_{k \to \infty} \varphi^k(a)$, that is in the words $\varphi^k(a)$ for $k = 1, 2, \ldots$.

An example of a sequence generated by a morphism can be the following sequence w. We create words by starting with the letter 1 and iterating the substitution system ϕ_w: $1 \to 123$, $2 \to 13$, $3 \to 2$. Thus, the initial letters of w are 123132123213.... This sequence was constructed in connection with the problem of constructing a nonrepetitive sequence on a 3-letter alphabet, that is, a sequence that does not contain any subwords of the type $XX = X^2$, where X is any non-empty word over a 3-letter alphabet. The sequence w has that property. The question of the existence of such a sequence, as well as the questions of the existence of sequences avoiding other kinds of repetitions, were studied in algebra [Adian, Justin, Kol],

discrete analysis [Carpi, Dekk, Evdok, Ker, Pleas] and in dynamical systems [MorseHedl]. In Examples 2.2, 2.6 and 3.3 we give the number of occurrences of the patterns 1-2, 2-1, 1-1-$\cdots$-1, 12, 123 and 21 in the finite approximations of w.

To proceed further, we need the following definitions. Let $N_\phi^\tau(n)$ denote the number of occurrences of the pattern τ in a word generated by some morphism ϕ after n iterations. We say that an occurrence of τ is *external* for a pair of words (X, Y), if this occurrence starts in X and ends in Y. Also, an occurrence of τ for a word X is *internal*, if this occurrence starts and ends in this X.

2. PATTERNS 1-2, 2-1 AND 1-1-...-1

Theorem 2.1. *Let $\mathcal{A} = \{1, 2, \ldots, k\}$ be an alphabet, where $k \geq 2$ and a pattern $\tau \in \{1\text{-}2, 2\text{-}1\}$. Let X_1 begins with the letter 1 and consists of ℓ copies of each letter $i \in \mathcal{A}$ ($\ell \geq 1$). Let a morphism ϕ be such that*

$$1 \to X_1, \ 2 \to X_2, \ 3 \to X_3, \ldots, k \to X_k,$$

where we allow X_i to be the empty word ϵ for $i = 2, 3, \ldots, k$ (that is, ϕ may be an erasing morphism), $\sum_{i=2}^{k} |X_i| = k \cdot d$, and each letter from $\mathcal{A}$ appears in the word $X_2 X_3 \ldots X_k$ exactly d times. Besides, let $e_{i,j}$ (resp. e_i) be the number of external occurrences of τ for (X_i, X_j) (resp. (X_i, X_i)), where $i \neq j$. Let s_i be the number of internal occurrences of τ in X_i. In particular, $s_i = e_i = e_{i,j} = e_{j,i} = 0$, whenever $X_i = \epsilon$; also, $e_i = |X_i| \cdot (|X_i| - 1)/2$, whenever there are no repetitive letters in X_i. Then $N_\phi^\tau(1) = s_1$ and for $n \geq 2$, $N_\phi^\tau(n)$ is given by

$$(d + \ell)^{n-2} \sum_{i=1}^{k} s_i + \binom{(d+\ell)^{n-2}}{2} \sum_{i=1}^{k} e_i + (d + \ell)^{2n-4} \sum_{1 \leq i < j \leq k} e_{i,j}.$$

Proof. We assume that $\tau = 1\text{-}2$. All the considerations for this τ remain the same for the case $\tau = 2\text{-}1$.

If $n = 1$ then the statement is trivial.

Suppose $n \geq 2$. Using the fact that $X_1 X_2 X_3 \ldots X_k$, has exactly $d + \ell$ occurrences of each letter i, $i = 1, 2, \ldots, k$, one can prove by induction on n, that the word $\phi^n(1)$ is a permutation of $(d + \ell)^{n-2}$ copies of each word X_i, where $i = 1, 2, \ldots, k$. This implies, in particular, that $|\phi^n(1)| = k \cdot (d + \ell)^{n-1}$.

An occurrence of τ in $\phi^n(1)$ can be either internal, that is when τ occurs inside a word X_i, or external, which means that τ begins in a word X_i and ends in another word X_j. In the first of these cases, since there are $(d + \ell)^{n-2}$ copies of each X_i, we have $(d + \ell)^{n-2} \sum_{i=1}^{k} s_i$ possibilities. In the second case, either $i = j$, which gives $\binom{(d+\ell)^{n-2}}{2} \sum_{i=1}^{k} e_i$ possibilities, or $i \neq j$, in which case there are $(d+\ell)^{n-2}$ possibilities to choose X_i (resp. X_j) among $(d+\ell)^{n-2}$ copies of X_i (resp. X_j), and using the fact that $e_{i,j} = e_{j,i}$ (the order in which the words X_i and X_j occur in $\phi^n(1)$ is unimportant), we have $(d + \ell)^{2n-4} \sum_{1 \leq i < j \leq k} e_{i,j}$ possibilities. Summing all the possibilities, we finish the proof. $\square$

Let s (resp. e) denote the vector $(s_1, s_2, \ldots, s_k)$ (resp. $(e_1, e_2, \ldots, e_k)$), where s_i and e_j are defined in Theorem 2.1. All of the following examples are corollaries to Theorem 2.1.

Example 2.2. *If we consider the morphism ϕ_w defined in Section 1 and the pattern $\tau = 1\text{-}2$ then $d = \ell = 1$, $s = (3, 1, 0)$, $e = (3, 1, 0)$ and $e_{1,2} = e_{2,1} = 2$, $e_{1,3} = e_{3,1} = 1$, $e_{2,3} = e_{3,2} = 1$. Hence, the number of occurrences of τ is given by $N_{\phi_w}^{1\text{-}2}(1) = 3$ and, for $n \geq 2$, $N_{\phi_w}^{1\text{-}2}(n) = (3 \cdot 4^{n-1} + 2^n)/2$. If $\tau = 2\text{-}1$ then $s = (0, 0, 0)$, $e = (3, 1, 0)$ and $e_{1,2} = e_{2,1} = 2$, $e_{1,3} = e_{3,1} = 1$, $e_{2,3} = e_{3,2} = 1$. Hence, $N_{\phi_w}^{2\text{-}1}(1) = 0$ and, for $n \geq 2$, $N_{\phi_w}^{2\text{-}1}(n) = (3 \cdot 4^{n-1} - 2^n)/2$.*

Example 2.3. *If we consider the morphism ϕ: $1 \to 1324$, $2 \to \epsilon$, $3 \to 14$, and $4 \to 23$ then for the pattern $\tau = 1\text{-}2$, we have $d = \ell = 1$, $s = (5, 0, 1, 1)$, $e = (6, 0, 1, 1)$, and $e_{i,j}$, for $i \neq j$, are elements of the matrix*

$$
\begin{pmatrix}
- & 0 & 3 & 3 \\
0 & - & 0 & 0 \\
3 & 0 & - & 2 \\
3 & 0 & 2 & -
\end{pmatrix}.
$$

Hence, $N_{\phi}^{1\text{-}2}(1) = 5$ and, for $n \geq 2$, $N_{\phi}^{1\text{-}2}(n) = 3 \cdot 4^{n-1} + 11 \cdot 2^{n-2}$.

Example 2.4. *If we consider the morphism ϕ: $1 \to 13542$, $2 \to 423$, $3 \to \epsilon$, $4 \to 5115$, and $5 \to 234$ then for the pattern $\tau = 1\text{-}2$, we have $\ell = 1$, $d = 2$, $s = (6, 1, 0, 2, 3)$, $e = (10, 3, 0, 4, 3)$, and $e_{i,j}$, for $i \neq j$, are elements of the matrix*

$$
\begin{pmatrix}
- & 6 & 0 & 8 & 6 \\
6 & - & 0 & 6 & 3 \\
0 & 0 & - & 0 & 0 \\
8 & 6 & 0 & - & 6 \\
6 & 3 & 0 & 6 & -
\end{pmatrix}.
$$

Hence, $N_{\phi}^{1\text{-}2}(1) = 6$ and, for $n \geq 2$, $N_{\phi}^{1\text{-}2}(n) = 5 \cdot 9^{n-1} + 2 \cdot 3^{n-2}$.

Using the proof of Theorem 2.1, we have the following.

Theorem 2.5. *Let a morphism ϕ satisfy all the conditions in the statement of Theorem 2.1 and the pattern $\tau = \underbrace{1\text{-}1\text{-}\cdots\text{-}1}_{r \ times}$. Then, for $n \geq 2$, the number of occurrences of τ in $\phi^n(1)$ is given by $k \cdot \binom{(d+\ell)^{n-1}}{r}$, whereas for $n = 1$, by $k \cdot \binom{\ell}{r}$.*

Proof. From the proof of Theorem 2.1, we have that if $n \geq 2$ (resp. $n = 1$) then $\phi^n(1)$ has exactly $(d+\ell)^{n-1}$ (resp. ℓ) copies of each letter from $\mathcal{A}$. We can choose r of them in $\binom{(d+\ell)^{n-1}}{r}$ (resp. $\binom{\ell}{r}$) ways to form the pattern τ. The rest is clear. $\square$

The following example is a corollary to Theorem 2.5.

Example 2.6. *If we consider the morphism ϕ_w defined in Section 1 and the pattern $\tau = 1\text{-}1\text{-}1\text{-}1$ then $d = \ell = 1$, $r = 4$, hence the number of occurrences of τ in $\phi^n(1)$ is 0, whenever $n = 1$ or $n = 2$, and $3 \cdot \binom{2^{n-1}}{4}$ otherwise.*

3. Patterns without internal dashes

In what follows we need to extend the notion of an external occurrence of a pattern. Suppose $W = AXBYC$, where A, X, B, Y and C are some subwords. We say that an occurrence of τ in W is external for a pair of words (X, Y), if this occurrence starts in X, ends in Y and is allowed to have some of its letters in B. For instance, if $W = 12324245$, where $A = 1$,

$X = 23$, $B = 2$ and $Y = 424$ then an occurrence of the generalized pattern 213, namely the subword 324 is an external occurrence for (X, Y).

Theorem 3.1. *Let $\mathcal{A} = \{1, 2, \ldots, k\}$ be an alphabet and a generalized pattern τ has no internal dashes. Let X_1 begins with the letter 1 and consists of ℓ copies of each letter $i \in \mathcal{A}$ ($\ell \geq 1$). Let a morphism ϕ be such that*

$$1 \to X_1, \ 2 \to X_2, \ 3 \to X_3, \ldots, k \to X_k,$$

where we allow X_i to be the empty word ϵ for $i = 2, 3, \ldots, k$ (that is, ϕ may be an erasing morphism), $\sum_{i=2}^{k} |X_i| = k \cdot d$, and each letter from $\mathcal{A}$ appears in the word $X_2 X_3 \ldots X_k$ exactly d times. Besides, we assume that there are no external occurrences of τ in $\phi^n(1)$ for the pair (X_i, X_j) for each i and j. Let s_i be the number of internal occurrences of τ in X_i. In particular, $s_i = 0$, whenever $X_i = \epsilon$. Then $N_\phi^\tau(1) = s_1$ and for $n \geq 2$, $N_\phi^\tau(n) = (d + \ell)^{n-2} \sum_{i=1}^{k} s_i$.

Proof. The theorem is straightforward to prove by observing that for $n \geq 2$, $\phi^n(1)$ has $(d + \ell)^{n-2}$ occurrences of each word X_i (see the proof of Theorem 2.1). $\square$

Remark 3.2. *In order to use Theorem 3.1, we need to control the absence of external occurrences of a pattern τ for given τ (without internal dashes) and a morphism ϕ. To do this, we need, for any pair (X_i, X_j), to consider all the words $X_i W X_j$, where $|W| < |\tau| - 1$, and W is a permutation of a number of words from the set $\{X_1, X_2, \ldots, X_k\}$.*

The following examples are corollaries to Theorem 3.1.

Example 3.3. *If we consider the morphism ϕ_w defined in Section 1 and the pattern $\tau = 12$ then all the conditions of Theorems 3.1 hold. In this case $d = \ell = 1$ and $s = (2, 1, 0)$. Hence, the number of occurrences of the patterns 12, that is the number of rises, is given by $N_{\phi_w}^{12}(1) = 2$ and, for $n \geq 2$, $N_{\phi_w}^{12}(n) = 3 \cdot 2^{n-2}$. If $\tau = 123$ then we can apply the theorem to get that for $n \geq 2$, $N_{\phi_w}^{123}(n) = 2^{n-2}$.*

If we want to count the number of occurrences of the pattern $\tau = 21$, that is the number of descents, then we cannot apply Theorem 2.1, since for instance, the pair $(X_1, X_2) = (123, 13)$ has an external occurrence of τ. However, it is obvious that the number of descents in $\phi^n(1)$ is equal to $|\phi^n(1)| - N_{\phi_w}^{12}(1) - 1 = 3 \cdot 2^{n-2} - 1$.

Example 3.4. *If we consider the morphism ϕ: $1 \to 1243$, $2 \to 3$, $3 \to \epsilon$, and $4 \to 124$ then for the pattern $\tau = 123$, all the conditions of Theorems 3.1 hold. In this case $d = \ell = 1$, $s = (1, 0, 0, 1)$. Hence, for $n \geq 1$, $N_\phi^{123}(n) = 2^{n-1}$. For $\tau = 321$ we cannot apply Theorem 3.1, since the pair (X_4, X_1) has an external occurrence of τ (look at $X_4 X_2 X_1 = 12\mathbf{431}243$). Consideration of the words $X_4 X_2$ and $X_4 X_1$ implies that the theorem cannot be apply for the patterns 132 and 231 respectively. However, we can apply the theorem to the pattern 213 to prove that it does not occur in $\phi^n(1)$ for any n.*

References

[Adian] Adian S. I.: *The Burnside problem and identities in groups.* Translated from the Russian by John Lennox and James Wiegold. Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas], 95. Springer-Verlag, Berlin-New York, (1979). xi+311 pp.

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

[Burstein] Burstein A., Enumeration of words with forbidden patterns, Ph.D. thesis, University of Pennsylvania, 1998.

[BurMans1] Burstein A. and Mansour T., Words restricted by patterns with at most 2 distinct letters, *Electronic J. of Combinatorics*, to appear (2002).

[BurMans2] Burstein A. and Mansour T., Words restricted by 3-letter generalized multipermutation patterns, preprint CO/0112281.

[BurMans3] Burstein A. and Mansour T., Counting occurrences of some subword patterns, preprint CO/0204320.

[Carpi] Carpi A.: On the number of abelian square-free words on four letters, Discrete Appl. Mathematics, Elsevier, 81 (1998), 155–167.

[Claes] A. Claesson: Generalised Pattern Avoidance, European J. Combin. 22 (2001), no. 7, 961–971.

[Dekk] Dekking F. M.: Strongly non-repetitive sequences and progression-free sets, Journal Com. Theory, Vol. 27-A, No. 2 (1979), 181–185.

[Evdok] Evdokimov A. A.: Strongly asymmetric sequences generated by a finite number of symbols, Dokl. Akad. Nauk SSSR, 179 (1968), 1268–1271. (Russian) English translation in: Soviet Math. Dokl., 9 (1968), 536–539.

[Frid] Frid A. E.: On the frequency of factors in a DOL word, J. Automata, Languages and Combinatorics, Otto-von-Guericke-Univ., Magdeburg 3(1) (1998), 29–41.

[Justin] Justin J.: Characterization of the repetitive commutative semigroups, Journal of Algebra (1972), no. 21, 87–90.

[Ker] Keränen V.: Abelian squares are avoidable on 4 letters, In W. Kuich, editor, Proc. ICALP'92, Lecture Notes in Comp. Sci., 623, Springer-Verlag, Berlin (1992), 41–52.

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

[Kol] Kolotov A. T.: Aperiodic sequences and functions of the growth of algebras, Algebra i Logika 20 (1981), no. 2, 138–154. (Russian)

[Lind] Lindenmayer A.: Mathematical models for cellular interaction in development, Parts I and II, Journal of Theoretical Biology, 18 (1968), 280–315.

[LindRoz] Lindenmayer A., Rozenberg G.: *Automata, languages, development*, North-Holland Publishing Co., Amsterdam-New York-Oxford (1976), viii+529 pp.

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

[MorseHedl] Morse M., Hedlung G.: Unending chess, symbolic dynamics and a problem in semigroups, Duke Math. Journal, Vol. 11, No. 1 (1944), 1–7.

[Pleas] Pleasants P.: Non-repetitive sequences, Proc. Camb. Phil. Soc., Vol. 68 (1970), 267–274.

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

[SimSch] R. Simion, F. Schmidt: Restricted permutations, European J. Combin. 6, no. 4 (1985), 383–406.