ON THE NUMBER OF ABELIAN BORDERED WORDS (WITH AN EXAMPLE OF AUTOMATIC THEOREM-PROVING)

Daniel GOČ

School of Computing, Queen's University,
Kingston, Ontario K7L 3N6, Canada

Narad RAMPERSAD

Dept. of Math. and Statistics, University of Winnipeg,
515 Portage Ave. Winnipeg, MB, R3B 2E9, Canada,
Narad.Rampersad@gmail.com

Michel RIGO, Pavel SALIMOV*

University of Liège, Department of Mathematics, Grande traverse 12 (B37)
B-4000 Liège, Belgium
M.Rigo@ulg.ac.be

In the literature, many bijections between (labeled) Motzkin paths and various other combinatorial objects are studied. We consider abelian (un)bordered words and show the connection with irreducible symmetric Motzkin paths and paths in $\mathbb{Z}$ not returning to the origin. This study can be extended to abelian unbordered words over an arbitrary alphabet and we derive expressions to compute the number of these words. In particular, over a 3-letter alphabet, the connection with paths in the triangular lattice is made. Finally, we characterize the lengths of the abelian unbordered factors occurring in the Thue–Morse word using some kind of automatic theorem-proving provided by a logical characterization of the k-automatic sequences.

Keywords: bordered word; Motzkin paths; automatic sequences; first-order logic.

1991 Mathematics Subject Classification: 68Q45, 68R15

1. Introduction

A finite word is *bordered* if it has a proper prefix that is also a suffix of the whole word. Otherwise, the word is said to be *unbordered*. Such properties have been investigated for a long time in combinatorics on words. For instance, the famous

conjecture of Duval about the relationship between the length of a word and the maximum length of its unbordered factors has been solved in [9]. A classic result by Ehrenfeucht and Silberger [6] states that if an infinite word has only finitely many unbordered factors, then it is ultimately periodic.

Let us denote the *Parikh vector* of the word u over A by $\Psi(u)$: i.e., $\Psi(u)$ is the element of $\mathbb{N}^A$ representing the number of occurrences of each letter in u. Two words u and v are *abelian equivalent*, if $\Psi(u) = \Psi(v)$. The notions of (un)bordered words are naturally extended to their abelian analogues by replacing equality with abelian equivalence. Such an extension is considered, for example, in [10].

This paper is organized as follows. Below, we make precise the basic definitions. In Section 2, we show that abelian unbordered words over a two letter alphabet are in one-to-one correspondence with particular Motzkin paths. In Section 3, abelian unbordered words over a two letter alphabet are shown to be in one-to-one correspondence with n-step walks in $\mathbb{Z}$ starting from the origin but not returning to it. In particular, the number of these n-step walks is well-known and is given by the sequence A063886 in Sloane's Encyclopedia [14]. In Section 4, we extend the size of the alphabet and are still able to relate abelian unbordered words with specific paths and then derive a recursion formula to get the number of such words of length n. Finally, in Section 5, we consider the abelian unbordered factors occurring in abelian periodic automatic sequences (such as the Thue–Morse word).

This paper is an updated version of the paper [11] presented during the DLT 2013 conference. Theorem 21 was only conjectured in [11]. We are now able to describe a technique leading to the proof. We emphasize this part of the paper. It provides a non-trivial example about how Büchi's theorem and the logical characterization of k-automatic sequences can be useful in combinatorics on words [3, 8].

Definition 1. *A word $u \in A^*$ is* abelian bordered *if there exist $v, v', x, y \in A^+$ such that $u = vx = yv'$ and $\Psi(v) = \Psi(v')$. In that case v is an* abelian border *of u. Otherwise, u is said to be* abelian unbordered.

It is easy to see that if u is abelian bordered, it has an abelian border of length at most $|u|/2$. Note that a word u over $\{a, b\}$ is abelian unbordered if and only if its complement $\overline{u}$, where all a's are replaced with b's and all b's with a's, is also abelian unbordered. If a word is bordered, then it is trivially abelian bordered. But, in general, the converse does not hold. For instance, *aabbabab* is abelian bordered but not bordered.

Example 2. *We consider the first few abelian unbordered words over $\{a, b\}$ that start with a: a, ab, aab, abb, aaab, aabb, abbb, aaaab, aaabb, aabbb, abbbb, aabab, ababb. The first few values for the number of abelian unbordered words of length $n \geq 0$ over $\{a, b\}$ are: 2, 2, 4, 6, 12, 20, 40, 70, 140, 252, 504, 924, 1848, 3432, 6864,... These values match Sloane's sequence* A063886.

Remark 3. *The language of abelian bordered words is not context-free* [11].

2. Connection with Motzkin Words

The following is an immediate consequence of the definition "abelian unbordered".

Lemma 4. *Let $n \geq 1$. A word $u_1 \cdots u_n c v_n \cdots v_1$, where for all $i \in \{1, \ldots, n\}$, $u_i \in A$, $v_i \in A$, and $c \in \{\varepsilon\} \cup A$, is abelian unbordered if and only if, for all $i \in \{1, \ldots, n\}$, $\Psi(u_1 \cdots u_i) \neq \Psi(v_i \cdots v_1)$.*

Let us fix the alphabet $A = \{a, b\}$. If $x = x_1 \cdots x_n$ and $y = y_1 \cdots y_n$ are words of length n over A, we define

$$\binom{x}{y} := \binom{x_1}{y_1} \cdots \binom{x_n}{y_n} \in (A \times A)^* \text{ and } \pi_1 : (A \times A)^* \to A, \binom{x}{y} \mapsto x.$$

We write x^R to denote the *reversal* of x; that is, $x^R = x_n \cdots x_1$. We now define the map m and a context-free language $P \subseteq (A \times A)^*$

$$m : A^* \to (A \times A)^*, \ u \mapsto \binom{u}{u^R}, \quad P = m(A^*) = \left\{ \binom{u}{u^R} \mid u \in A^* \right\}.$$

Lemma 4 can be restated as follows.

Lemma 5. *A word $u \in A^+$ is abelian bordered if and only if there exists a non-empty proper prefix p of $m(u)$ such that the numbers of occurrences of $\binom{a}{b}$ and $\binom{b}{a}$ in p are the same.*

Definition 6. *A* Grand Motzkin path *of length n is a lattice path of $\mathbb{N}^2$ running from $(0,0)$ to $(n,0)$, whose permitted steps are the up diagonal step $(1,1)$, the down diagonal step $(1,-1)$ and the horizontal step $(1,0)$, called* rise, fall *and* level *step, respectively.*

A Motzkin path *is a Grand Motzkin path that never passes below the x-axis.*

An irreducible *(or* elevated*) Motzkin path is a Motzkin path that does not touch the x-axis except for the origin and the final destination* [1].

If the level steps are labeled by k colors (here colors will be letters from the alphabet A) we obtain a k-colored Motzkin path [12]. *A k-colored Motzkin path is described by a word over the alphabet $\{R, F, L_1, \ldots, L_k\}$ and the context-free language of the k-colored Motzkin paths is denoted by $\mathcal{M}_k$. In particular, a Motzkin path described by a word over $\{R, F\}$ is a* Dyck path.

Let $h : (A \times A)^* \to \{R, F, L_a, L_b\}^*$ be the coding

$$h : \binom{a}{b} \mapsto R, \ \binom{b}{a} \mapsto F, \ \binom{a}{a} \mapsto L_a, \ \binom{b}{b} \mapsto L_b.$$

Note that if p belongs to P, then $h(p)$ is a *symmetric Grand Motzkin path* having a symmetry axis $x = n/2$. Let $\iota : \{R, F, L_a, L_b\}^* \to \{R, F, L_a, L_b\}^*$ defined by $\iota(R) = F$, $\iota(F) = R$, $\iota(L_a) = L_a$ and $\iota(L_b) = L_b$. If w is a word over $\{R, F, L_a, L_b\}$,

then $\widetilde{w} = \iota(w^R)$. A symmetric Grand Motzkin path is described by a word of the kind $w\, c\, \widetilde{w}$ where $c \in \{\varepsilon, L_a, L_b\}$.

Example 7. *Two Motzkin paths colored with the letters a and b are represented in Figure 1. The left one is described by the word RL_aRFRL_bFFRF and the right one is symmetric and irreducible. It corresponds to the word $h(m(aaaababababb)) = RRL_aRFRFRFL_aFF$.*
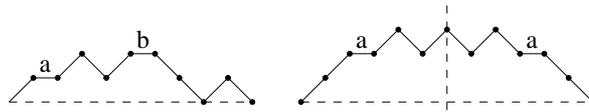
Fig. 1. Two Motzkin paths.

Now we can restate Lemma 5.

Lemma 8. *A word u starting with a is abelian unbordered if and only if $h(m(u))$ is a symmetric and irreducible Motzkin path.*

Proposition 9. *The set of abelian unbordered words over $\{a, b\}$ starting with a and of length at least 2 is $a\pi_1(P \cap h^{-1}(\mathcal{M}_2))b$.*

Proof. Note that $h(P \cap h^{-1}(\mathcal{M}_2))$ is the set of all symmetric 2-colored Motzkin paths. Now observe that if u belongs to $a\pi_1(P \cap h^{-1}(\mathcal{M}_2))b$, then $h(m(u))$ starts with R and ends with F. So the corresponding 2-colored Motzkin path is irreducible. Conversely, if u is abelian unbordered and starts with a, then by Lemma 8, $h(m(u)) = RMF$, where M is a symmetric 2-colored Motzkin path. Thus, $u \in a\pi_1(P \cap h^{-1}(\mathcal{M}_2))b$. □

Remark 10. *Any symmetric 2-colored Motzkin path can be built by reflecting a prefix of a 2-colored Motzkin path. Let $w \in \{R, F, L_a, L_b\}^*$ be a prefix of length $k - 1$ of a word in $\mathcal{M}_2$. By the previous proposition, we get that $a\pi_1[h^{-1}(w\,\widetilde{w})]b$, $a\pi_1[h^{-1}(w\,L_a\,\widetilde{w})]b$ and $a\pi_1[h^{-1}(w\,L_b\,\widetilde{w})]b$ are respectively an abelian unbordered word of length $2k$, of length $2k + 1$ having a as central letter, of length $2k + 1$ having b as central letter.*

- *The set of abelian unbordered words of length $2k$ starting with a is in one-to-one correspondence with the set of prefixes of length $k - 1$ of words in $\mathcal{M}_2$. Equivalently, the set of abelian unbordered words of length $2k$ starting with a is in one-to-one correspondence with the set of prefixes of length k of irreducible 2-colored Motzkin paths.*
- *The set of abelian unbordered words of length $2k + 1$ starting with a and having a central letter equal to a (resp. b) is in one-to-one correspondence*

with the set of prefixes of length $k-1$ of words in $\mathcal{M}_2$. Equivalently, the set of abelian unbordered words of length $2k+1$ starting with a is in one-to-one correspondence with the set of prefixes of length k of irreducible 2-colored Motzkin paths.

3. Connection with the Sequence `A063886`

The sequence `A063886` gives the number $s(n)$ of n-step walks in $\mathbb{Z}$ starting from the origin but not returning to it. Such walks can be described by words over $\{r, \ell\}$ for right and left steps. The aim of this section is to show that the set of abelian unbordered words over a binary alphabet is in one-to-one correspondence with the n-step walks in $\mathbb{Z}$ starting from the origin but not returning to it. Let us first collect some well-known facts. The generating function for $s(n)$ is $\sqrt{\frac{1+2x}{1-2x}}$.

Consider a word $w = u_1 \cdots u_n v_n \cdots v_1 \in \{a, b\}^*$ of length $2n$. We consider the map c by

$$c : \binom{a}{b} \mapsto rr, \quad \binom{b}{a} \mapsto \ell\ell, \quad \binom{a}{a} \mapsto \ell r, \quad \binom{b}{b} \mapsto r\ell.$$

Applying c to the prefix of length n of $m(w)$ provides a unique path of length $2n$ in $\mathbb{Z}$. This path is denoted by $\mathfrak{p}(w)$. It is clear that $\mathfrak{p}$ is a one-to-one correspondence between the words of length $2n$ over $\{a, b\}$ and the paths of length $2n$ in $\mathbb{Z}$ starting from the origin. The following proposition follows immediately from Lemma 5.

Proposition 11. *A word w over $\{a, b\}$ of even length is abelian unbordered if and only if the path $\mathfrak{p}(w)$ does not return to the origin.*

We extend the definition of $\mathfrak{p}$ to words of odd length by

$$\mathfrak{p}(u_1 \cdots u_n \alpha v_n \cdots v_1) = \begin{cases} \mathfrak{p}(u_1 \cdots u_n v_n \cdots v_1)\, \ell \text{ , if } \alpha = a; \\ \mathfrak{p}(u_1 \cdots u_n v_n \cdots v_1)\, r \text{ , if } \alpha = b. \end{cases}$$

With this definition, $\mathfrak{p}$ is a one-to-one correspondence between the abelian unbordered words of length $2n+1$ over $\{a, b\}$ and the paths of length $2n+1$ in $\mathbb{Z}$ starting from the origin and not returning to it. It is therefore easy to get a result similar to the above proposition for words of odd length.

Proposition 12. *A word w over $\{a, b\}$ of odd length is abelian unbordered if and only if the path $\mathfrak{p}(w)$ does not return to the origin.*

4. Larger Alphabets

Let $k \geq 2$. Consider the alphabet $A = \{a_1, \ldots, a_k\}$, or simply $\{1, \ldots, k\}$, and $\mathbb{Z}^k$ equipped with the usual unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$, whose coordinates are all equal to zero except one which is equal to 1. To be able to define k-colored paths, we assume that at each point in $\mathbb{Z}^k$, there are exactly k loops colored with the k different letters.

6   *D. Goč, N. Rampersad, M. Rigo, P. Salimov*

We first consider a word $u_1 \cdots u_n v_n \cdots v_1$ of even length $2n$. Take the prefix of length n of $m(u_1 \cdots u_n v_n \cdots v_1)$ and apply to it the morphism $h_k : (A \times A)^* \to \{\mathbf{e}_i - \mathbf{e}_j \mid 1 \le i, j \le k\}^* \subset (\mathbb{Z}^k)^*$ defined by

$$h_k \begin{pmatrix} a_i \\ a_j \end{pmatrix} = \mathbf{e}_i - \mathbf{e}_j, \quad \forall i, j \in \{1, \ldots, k\}.$$

Therefore, to the word $w = u_1 \cdots u_n v_n \cdots v_1$ there corresponds the sequence of $n+1$ points in $\mathbb{Z}^k$

$$p_0 = \mathbf{0}, \ p_1 = h_k \begin{pmatrix} u_1 \\ v_1 \end{pmatrix}, \ p_2 = h_k \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} + h_k \begin{pmatrix} u_2 \\ v_2 \end{pmatrix}, \ldots, \ p_n = \sum_{j=1}^{n} h_k \begin{pmatrix} u_j \\ v_j \end{pmatrix},$$

where $\mathbf{0}$ denotes the origin $(0, 0, \ldots, 0)$. By the definition of h_k, note that all these points lie in the subspace $\mathbf{H}_k$ of $\mathbb{Z}^k$ satisfying the equation

$$x_1 + \cdots + x_k = 0.$$

Definition 13. *A* path *of length n in $\mathbf{H}_k$ is a sequence $p_0, \ldots, p_n$ of points in $\mathbf{H}_k$ such that, for all $j \ge 1$, $p_j - p_{j-1}$ belongs to $\{\mathbf{e}_i - \mathbf{e}_j \mid 1 \le i, j \le k, i \ne j\}$.*

A k-colored path of length n in $\mathbf{H}_k$ is a sequence $p_0, c_0, p_1, c_1, \ldots, p_{n-1}, c_{n-1}, p_n$ alternating points in $\mathbf{H}_k$ and elements belonging to $A \cup \{\varepsilon\}$ in such a way that, if $p_j \ne p_{j+1}$, then $p_{j+1} - p_j$ belongs to $\{\mathbf{e}_i - \mathbf{e}_j \mid 1 \le i, j \le k, i \ne j\}$ and $c_j = \varepsilon$, otherwise c_j belongs to A and can be interpreted as the color assigned to a loop on p_j. Note that paths are special cases of k-colored paths.

For the rest of this paper we will only consider paths that start at the origin.

Remark 14. *For $k = 3$, $\mathbf{H}_3$ corresponds to the so-called* triangular lattice *(sometimes called hexagonal lattice) because a point $\mathbf{x}$ has exactly six neighbors. The set of neighbors of $\mathbf{x}$ is denoted by*

$$N(\mathbf{x}) := \mathbf{x} + \{\mathbf{e}_1 - \mathbf{e}_2, \mathbf{e}_1 - \mathbf{e}_3, \mathbf{e}_2 - \mathbf{e}_1, \mathbf{e}_2 - \mathbf{e}_3, \mathbf{e}_3 - \mathbf{e}_1, \mathbf{e}_3 - \mathbf{e}_2\}.$$

Consider the word $w = 23321211$ over the alphabet $\{1, 2, 3\}$. The prefix of length 4 of $m(w)$ is

$$\begin{pmatrix} 2 & 3 & 3 & 2 \\ 1 & 1 & 2 & 1 \end{pmatrix}$$

and corresponds to the sequence of moves $p_1 - 0 = \mathbf{e}_2 - \mathbf{e}_1$, $p_2 - p_1 = \mathbf{e}_3 - \mathbf{e}_1$, $p_3 - p_2 = \mathbf{e}_3 - \mathbf{e}_2$ and $p_4 - p_3 = \mathbf{e}_2 - \mathbf{e}_1$ and the path represented in Fig. 2. The second path in Fig. 2 is colored and has four loops with labels $2, 1, 3$ and 1 respectively. It corresponds to $w' = 2321323113121211$. The prefix of length 8 of $m(w')$ is

$$\begin{pmatrix} 2 & 3 & 2 & 1 & 3 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 & 2 & 1 & 3 & 1 \end{pmatrix}.$$

Observe that in this prefix we have an occurrence of a repeated symbol in positions $3, 4$ and $7, 8$ corresponding to the four loops in the path.
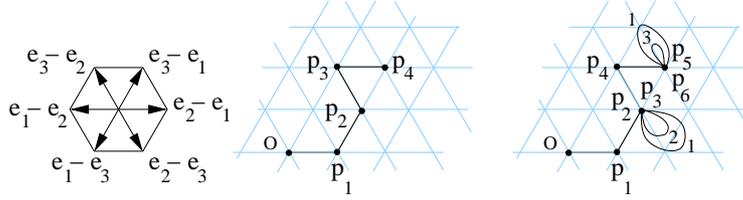
Fig. 2. In the triangular lattice, a path and a 3-colored path.

The word $w = u_1 \cdots u_n v_n \cdots v_1$ is said to be *simple* if, for all $i \in \{1, \ldots, n\}$, $u_i \neq v_i$. In this case, in the sequence of points $p_0 = \mathbf{0}, p_1, \ldots, p_n$ corresponding to w, for all $j < n$, we have $p_j \neq p_{j+1}$. Therefore simple words w of length $2n$ correspond to paths of length n in $\mathbf{H}_k$. Such paths are denoted by $\mathfrak{p}(w)$. For a non-simple word w of length $2n$ there is a corresponding k-colored path $\mathfrak{p}(w)$ of length n in $\mathbf{H}_k$ (where at least one loop $p_j = p_{j+1}$ occurs along the path). Conversely, for each k-colored path of length n in $\mathbf{H}_k$, there is a corresponding word of length $2n$.

Proposition 15. *A word w over $\{a_1, \ldots, a_k\}$ of even length $2n$ is abelian unbordered if and only if the k-colored path $\mathfrak{p}(w)$ in $\mathbf{H}_k$ of length n does not return to the origin. Moreover, a simple word w over $\{a_1, \ldots, a_k\}$ of even length $2n$ is abelian unbordered if and only if $\mathfrak{p}(w)$ is a path in $\mathbf{H}_k$ of length n without loops that does not return to the origin.*

Proof. Let $w = u_1 \cdots u_n v_n \cdots v_1$. Observe that starting from the origin, the path $p_0 = \mathbf{0}, c_0, p_1, \ldots, p_{n-1}, c_{n-1}, p_n$ returns to the origin if and only it there exists $j \leq n$ such that

$$\sum_{i=1}^{j} h_k \binom{u_i}{v_i} = 0. \qquad \qquad \square$$

If $w = u_1 \cdots u_n \alpha v_n \cdots v_1$ is a word of odd length, we can first consider the prefix of length n of $m(w)$ and then add an extra loop of color α to the end of the corresponding path $\mathfrak{p}(w)$. As for Proposition 12, we get the following.

Proposition 16. *A word w over $\{a_1, \ldots, a_k\}$ of odd length $2n + 1$ is abelian unbordered if and only if the k-colored path $\mathfrak{p}(w)$ of length $n + 1$ in $\mathbf{H}_k$ does not return to the origin. In particular, such a path ends with a loop whose color is the one corresponding to the central letter of w.*

Remark 17. *The numbers of abelian unbordered words of length n over a 3-letter alphabet, for $1 \leq n \leq 10$, are: 3, 6, 18, 48, 144, 402, 1206, 3438, 10314, 29754 and for simple abelian unbordered words, we get 3, 6, 18, 30, 90, 168, 504, 954, 2862, 5508. As we can observe, over a 3-letter alphabet, the number of abelian unbordered words (resp. simple abelian unbordered words) of length $2n + 1$ is three times*

8   *D. Goč, N. Rampersad, M. Rigo, P. Salimov*

n	1	2	3	4	5	6	7	8	9	10
a.u.w of length n	3	6	18	48	144	402	1206	3438	10314	29754
simple a.u.w of length n	3	6	18	30	90	168	504	954	2862	5508

Table 1. Number of (simple) abelian unbordered words over a 3-letter alphabet.

the number of abelian unbordered words (resp. simple abelian unbordered words) of length $2n$ because there are three available choices for the central letter. This observation extends trivially to an arbitrary alphabet.

From the discussion above and taking only entries of even index in the previous table, we also get the number of paths (resp. paths without loops) of length n in the triangular lattice $\mathbf{H}_3$ that do not return to the origin. We denote this quantity by $p_3(n)$ (resp. $s_3(n)$). The first few values of $p_3(n)$, $n \geq 1$, are 6, 48, 402, 3438, 29754, 259464, 2274462 and the first few values of $s_3(n)$ are 6, 30, 168, 954, 5508, 32016, 187200. The next statement means that one only needs to compute the sequence $(s_k(n))_{n\geq 1}$ to get $(p_k(n))_{n\geq 1}$ and thus the number of abelian unbordered words of length n.

Lemma 18. *We have*

$$p_k(n) = \sum_{i=1}^{n} s_k(i)\, k^{n-i} \binom{n-1}{n-i} .$$

Proof. By a (k-colored) path, we mean a path in $\mathbf{H}_k$ that does not return to the origin. Each such k-colored path of length n has a unique underlying path of length i, for some $i \in \{1, \ldots, n\}$. To get a k-colored path of length n, $n-i$ loops are added to this underlying path. Each loop can be placed independently at any point of the path, except the origin, and can be colored independently in one of k colors. So, the total number of ways to extend such a path of length i to a k-colored path of length n is $k^{n-i}\binom{n-1}{n-i}$. □

4.1. *Computation of $(s_3(n))_{n\geq 0}$ and then $(s_k(n))_{n\geq 0}$*

We show how to get a recurrence relation to compute the number $s_3(n)$, i.e., the number of paths in the triangular lattice $\mathbf{H}_3 = (V, E)$ that do not return to the origin; here V (resp. E) is the set of vertices (resp. edges) of $\mathbf{H}_3$. Consider the map

$$e : V \to \mathbb{N}, \quad \mathbf{x} \mapsto \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{0}, \\ 0 & \text{otherwise.} \end{cases}$$

If $f : V \to \mathbb{N}$ is a map, we denote by $\mathcal{S}f : V \to \mathbb{N}$ the map defined by

$$(\mathcal{S}f)(\mathbf{x}) = \sum_{\mathbf{y} \in N(\mathbf{x})} f(\mathbf{y})$$
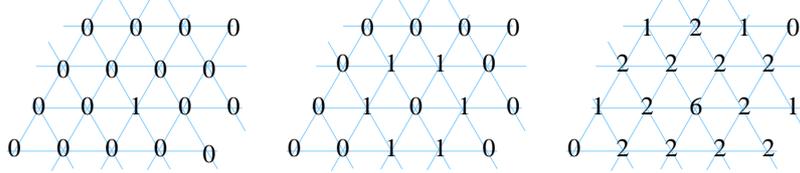
Fig. 3. values of e, $\mathcal{S}e$ and $\mathcal{S}^2 e$ around $\mathbf{0}$.

where $N(\mathbf{x})$ is the set of neighbors of $\mathbf{x}$. In particular, if $f, g : V \to \mathbb{N}$ are maps, then $\mathcal{S}(f + g) = \mathcal{S}f + \mathcal{S}g$. A simple induction argument gives the following result.

Lemma 19. *With the above notation, $(\mathcal{S}^n e)(\mathbf{x})$ is equal to the number of paths of length n that end at $\mathbf{x}$.*

The values of the maps e, $\mathcal{S}e$ and $\mathcal{S}^2 e$ around $\mathbf{0}$ are given in Figure 3. Let $r_{3,n} : V \to \mathbb{N}$ be defined as follows: $r_{3,n}(\mathbf{x})$ is the number of paths of length n that end at $\mathbf{x}$ and never return to the origin. Then

$$s_3(n) = \sum_{\mathbf{x} \in V} r_{3,n}(\mathbf{x})$$

where the sum is finite, since $r_{3,n}(\mathbf{x}) \neq 0$ implies that $\mathbf{x}$ is at distance at most n from the origin. If a map $f : V \to \mathbb{N}$ is constant on $N(\mathbf{0})$ (as is the case for $\mathcal{S}^n e$), then $\gamma(f)$ is a shorthand for $f(\mathbf{y})$ for any $\mathbf{y} \in N(\mathbf{0})$. By the symmetry of $\mathbf{H}_3$, we see that $r_{3,n}$ is constant on $N(\mathbf{0})$. Note that

$$s_3(n + 1) = 6 s_3(n) - \sum_{\mathbf{x} \in N(\mathbf{0})} r_{3,n}(\mathbf{x}) = 6 s_3(n) - 6 \gamma(r_{3,n})$$

because all paths except the ones that end in vertices adjacent to $\mathbf{0}$ have 6 prolongations, and the excluded ones have 5 possible prolongations. The same argument can be applied to maps: $r_{3,n+1} = \mathcal{S}r_{3,n} - 6 \gamma(r_{3,n}) e$ and, applied inductively, this leads to the following relation for $r_{3,n+1}$:

$$r_{3,n+1} = \mathcal{S}^{n+1} e - \sum_{i=0}^{n} 6 \gamma(r_{3,i}) \mathcal{S}^{n-i} e . \tag{1}$$

The sequence $((\mathcal{S}^n e)(\mathbf{0}))_{n \geq 0}$ counting the paths of length n starting and ending at $\mathbf{0}$ is well-known (A002898 gives the number of n-step closed paths on the hexagonal lattice). For instance, we have

$$(\mathcal{S}^n e)(\mathbf{0}) = \sum_{\ell=0}^{n} (-2)^{n-\ell} \binom{n}{\ell} \sum_{j=0}^{\ell} \binom{\ell}{j}^3 \tag{2}$$

and its first values are 1, 0, 6, 12, 90, 360, 2040, 10080, 54810, 290640,... Due to the 6-fold symmetry of the maps around the origin, note that

$$\gamma(\mathcal{S}^n e) = \frac{(\mathcal{S}^{n+1} e)(\mathbf{0})}{6}. \tag{3}$$

Taking into account both (1) and (3), for all $n \geq 0$, we have

$$\gamma(r_{3,n+1}) = \frac{(\mathcal{S}^{n+2}e)(\mathbf{0})}{6} - \sum_{i=0}^{n} \gamma(r_{3,i})(\mathcal{S}^{n-i+1}e)(\mathbf{0})$$

and $\gamma(r_{3,0}) = 0$. As a conclusion, using the sequence $((\mathcal{S}^n e)(\mathbf{0}))_{n \geq 0}$, we can compute inductively $(\gamma(r_{3,n}))_{n \geq 0}$ and therefore $(s_3(n))_{n \geq 0}$. Using the above formula, the first values of $(\gamma(r_{3,n}))_{n \geq 0}$ are 0, 1, 2, 9, 36, 172, 816, 4101, 20840, 108558, 572028,... Knowing that $s_3(0) = 1$ is enough to compute

$$s_3(1) = 6(s_3(0) - \gamma(r_{3,0})) = 6, \ \ s_3(2) = 6(s_3(1) - \gamma(r_{3,1})) = 6(6-1) = 30, \ldots$$

Let $k \geq 3$. We now turn to the general case to compute $(s_k(n))_{n \geq 0}$. Consider the homomorphism of groups χ between $(\mathbf{H}_k, +)$ and $(\mathbb{Z}((z_1, \ldots, z_{k-1})), \cdot)$ defined by the images of a basis of $\mathbf{H}_k$, $\chi : \mathbf{e}_1 - \mathbf{e}_k \mapsto z_1, \ \mathbf{e}_2 - \mathbf{e}_k \mapsto z_2, \ldots, \ \mathbf{e}_{k-1} - \mathbf{e}_k \mapsto z_{k-1}$.

$s_3(n)$	$s_4(n)$	$s_5(n)$	$s_6(n)$
1	1	1	1
6	12	20	30
30	132	380	870
168	1536	7480	25860
954	18036	148140	771930
5508	213264	2944320	23090220
32016	2530464	58625000	691372500
187200	30097152	1168618000	20713217400
1098594	358552116	23312236300	620781579450

Table 2. The first values of $s_3(n)$, $s_4(n)$, $s_5(n)$ and $s_6(n)$.

In particular, $\chi(-\mathbf{v}) = 1/\chi(\mathbf{v})$ and $\chi(\mathbf{v} + \mathbf{v}') = \chi(\mathbf{v}).\chi(\mathbf{v}')$. Any element of $\mathbf{H}_k$ can be decomposed as a linear combination with integer coefficients of $\mathbf{e}_1 - \mathbf{e}_k, \mathbf{e}_2 - \mathbf{e}_k, \ldots, \mathbf{e}_{k-1} - \mathbf{e}_k$ and corresponds by χ to a unique Laurent polynomial in $z_1, \ldots, z_{k-1}$. Each vertex $\mathbf{x}$ in $\mathbf{H}_k$ has exactly $k(k-1)$ neighbors:

$$\mathbf{x} + \{\mathbf{e}_i - \mathbf{e}_j \mid 1 \leq i, j \leq k, \ i \neq j\}$$

and these $k(k-1)$ translations of $\mathbf{x}$ are coded through χ by the terms

$$\left\{ z_i + \frac{1}{z_i} \mid 1 \leq i \leq k-1 \right\} \cup \left\{ \frac{z_i}{z_j} \mid 1 \leq i, j \leq k-1, \ i \neq j \right\}.$$

Now consider the Laurent polynomial corresponding to these elementary translations:

$$T = \sum_{i=1}^{k-1} \left(z_i + \frac{1}{z_i} \right) + \sum_{i \neq j} \frac{z_i}{z_j} = \left(1 + \sum_{i=1}^{k-1} z_i \right) \left(1 + \sum_{i=1}^{k-1} \frac{1}{z_i} \right) - k.$$

Let $\mathbf{x} \in \mathbf{H}_k$ and $(j_1, \ldots, j_{k-1}) \in \mathbb{Z}^{k-1}$ be such that $\chi(\mathbf{x}) = z_1^{j_1} \cdots z_{k-1}^{j_{k-1}}$. The number of paths of length n from the origin to $\mathbf{x}$ in the lattice $\mathbf{H}_k$ is given by the coefficient

of $z_1^{j_1} \cdots z_{k-1}^{j_{k-1}}$ in T^n. In particular, the constant term gives exactly the number of paths of length n returning to the origin. Furthermore, for $k = 3$ one can derive (2).

Example 20. *For $k = 4$, the number of paths of length n in $\mathbf{H}_4$ starting and ending at the origin is Sloane's sequence* A002899 *and is given by*

$$\sum_{\ell=0}^{n} (-4)^{n-\ell} \binom{n}{\ell} \sum_{j=0}^{\ell} \binom{\ell}{j}^2 \binom{2\ell - 2j}{\ell - j} \binom{2j}{j}.$$

For $k = 5$: the first few values are $1, 0, 20, 120, 1860, 23280, 342200, 5115600, \ldots$ and for $k = 6$: $1, 0, 30, 240, 4770, 82080, 1650900, 34524000, 758894850, \ldots$.

Being able to compute $(\mathcal{S}^n e)(\mathbf{0})$ for the lattice $\mathbf{H}_k$, we can proceed exactly as for the computation of $s_3(n)$ and get, for all $n \geq 0$,

$$\gamma(r_{k,n+1}) = \frac{(\mathcal{S}^{n+2} e)(\mathbf{0})}{k(k-1)} - \sum_{i=0}^{n} \gamma(r_{k,i})(\mathcal{S}^{n-i+1} e)(\mathbf{0})$$

with $\gamma(r_{k,0}) = 0$ and finally, $s_k(n+1) = k(k-1)(s_k(n) - \gamma(r_{k,n}))$.

5. About the Thue–Morse Word

Currie and Saari [5] proved that if $n \not\equiv 1 \pmod 6$, then the Thue–Morse word $\mathbf{t}$ has an unbordered factor of length n, but they left it open to decide for which lengths congruent to 1 modulo 6 does this property hold. Then Goč, Henshall and Shallit [8] showed that $\mathbf{t}$ has an unbordered factor of length n if and only if $(n)_2 \notin 1(01^*0)^* 10^* 1$, where $(n)_2$ denotes the base 2 expansion of n. Here we characterize the lengths of the abelian unbordered factors occurring in $\mathbf{t}$ as follows.

Theorem 21. *Let*

$$L = \{0^*(n)_2 : \mathbf{t} \text{ has an abelian unbordered factor of length } n\}.$$

Then L is accepted by the automaton in Figure 4, where all states are accepting except the four grey ones.

Generally, abelian properties of k-automatic sequences are not suited to be expressed in the extended Presburger arithmetic $\langle \mathbb{N}, +, V_k \rangle$ (see [13]). Nevertheless, we can take advantage of the fact that the Thue–Morse word is abelian periodic of period 2 and apply Büchi's theorem [2] with a technique similar to [3, 8]. We take verbatim the statement of Büchi's theorem as formulated by Charlier, Rampersad and Shallit in [3], which states that the k-automatic sequences are exactly the sequences definable in the first order structure $\langle \mathbb{N}, +, V_k \rangle$.

Theorem 22. [3] *If we can express a property of a k-automatic sequence $\mathbf{x}$ using quantifiers, logical operations, integer variables, the operations of addition, subtraction, indexing into $\mathbf{x}$, and comparison of integers or elements of $\mathbf{x}$, then this property is decidable.*

12 *D. Goč, N. Rampersad, M. Rigo, P. Salimov*
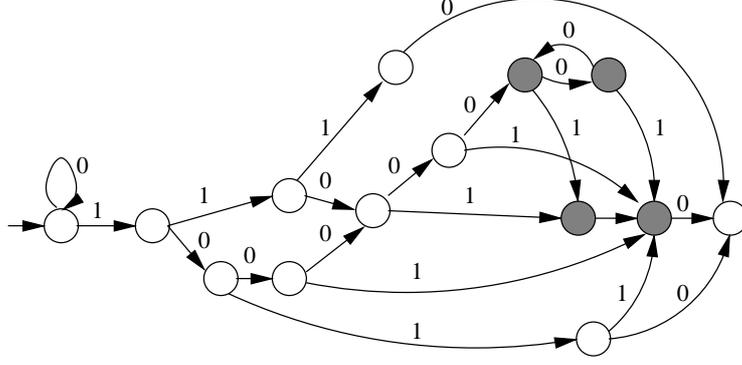
Fig. 4. A finite automaton for abelian unbordered factors in the Thue–Morse word.

The technique we are now describing can obviously be adapted to any k-automatic abelian periodic word. We will give in (4) below a first order formula $\varphi(n)$ in $\langle \mathbb{N}, +, V_2 \rangle$ that is satisfied if and only if an abelian unbordered factor of length n occurs in the Thue–Morse word $\mathbf{t}$. General procedures to obtain a finite automaton recognizing the base 2 expansions of the integers belonging to the set $\{n \in \mathbb{N} \mid \langle \mathbb{N}, +, V_2 \rangle \models \varphi(n)\}$ do exist (see for instance [2]). Hence a certified regular expression for the base 2 expansion of the elements in the above set will follow. Note that, since $\mathbf{t}$ is 2-automatic, we can define in $\langle \mathbb{N}, +, V_2 \rangle$ a unary function that maps i to $\mathbf{t}(i)$. Such a formula is again described in [2]. Predicates $e(n)$ and $o(n)$ are simply shorthands to characterize even and odd integers, $e(n) \equiv (\exists x)(n = x + x)$, $o(n) = \neg e(n)$. We define a predicate $B(i, n, k)$ which is true if and only if the Thue–Morse word has an abelian bordered factor of length n occurring at i with a border of length k. Since the Thue–Morse word $\mathbf{t}$ is a concatenation of ab and ba, discussing only the parity of the position i, the length n of the factor and the length k of the border, the predicate $B(i, n, k)$ is defined by the disjunction of the following terms $(e(i) \wedge e(n) \wedge e(k))$, $(e(i) \wedge e(n) \wedge o(k) \wedge \mathbf{t}(i + k - 1) = \mathbf{t}(i + n - k))$, $(e(i) \wedge o(n) \wedge e(k) \wedge \mathbf{t}(i + n - k) \neq \mathbf{t}(i + n - 1))$, $(e(i) \wedge o(n) \wedge o(k) \wedge \mathbf{t}(i + k - 1) = \mathbf{t}(i + n - 1))$, $(o(i) \wedge e(n) \wedge o(k) \wedge \mathbf{t}(i) = \mathbf{t}(i + n - 1))$, $(o(i) \wedge o(n) \wedge e(k) \wedge \mathbf{t}(i) \neq \mathbf{t}(i + k - 1))$, $(o(i) \wedge o(n) \wedge o(k) \wedge \mathbf{t}(i) = \mathbf{t}(i + n - k))$ and $(o(i) \wedge e(n) \wedge e(k) \wedge [(\mathbf{t}(i) = \mathbf{t}(i + n - k) \wedge \mathbf{t}(i + k - 1) = \mathbf{t}(i + n - 1)) \vee (\mathbf{t}(i) = \mathbf{t}(i + n - 1) \wedge \mathbf{t}(i + k - 1) = \mathbf{t}(i + n - k))])$. As an example, if i is even, n and k are odd, we have the situation depicted in Figure 5. In that case, since all blocks ab and ba are abelian equivalent, one has just to check equality of two symbols in adequate positions corresponding to the parameters.

Now the Thue–Morse word has an abelian unbordered factor of length n if and only if the following formula holds true

$$\varphi(n) \equiv (\exists i)(\forall k)(k \geq 1 \wedge 2k \leq n) \to \neg B(i, n, k). \tag{4}$$
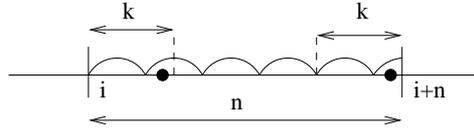
Fig. 5. A factor occurring in the Thue–Morse word.

A detailed description of how an automaton is created out of a predicate is given in [7]. We give a sketch of the method here. First, the predicate is re-written as a logical expression on primitives that are encodable as automata. For example, the automaton in Figure 6 evaluates the predicate $e(n)$. Likewise, the automaton in Figure 7 (left) encodes all ordered pairs (i, j) such that $\mathbf{t}(i) = \mathbf{t}(j)$. (In this automaton the pairs are encoded by first padding $(i)_2$ and $(j)_2$ with zeroes to be the same length and then listing the corresponding pairs of digits starting with the most significant pair.) In a similar manner, we can obtain an automaton for expressions like '$\mathbf{t}(i) \neq \mathbf{t}(i+k-1)$)'. Applying the logical connectives $\wedge$, $\vee$ and $\neg$ corresponds to language intersection, union and complement operations on automata respectively. Finally, the existential quantifier, $(\exists i)$, is implemented by dropping the corresponding digit from the transition labels of the automaton. In Figure 7 (right) we can see the result of applying such an operation to the automaton previously given in Figure 7 (left). (The universal quantifier is simply rewritten as $\neg(\exists i)\neg$ and evaluated as usual.)
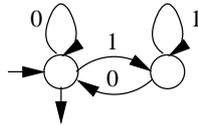
Fig. 6. An automaton accepting $(n)_2$, the binary expansion of n, such that n is even. (The digits of n are read starting with the most significant digit first.)
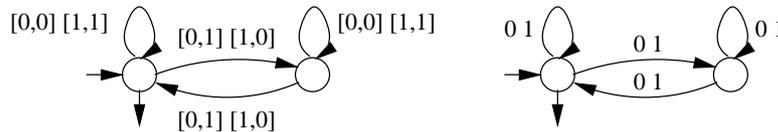
Fig. 7. (left) An automaton accepting pairs $(i, j)_2$ such that $\mathbf{t}(i) = \mathbf{t}(j)$. (right) An automaton accepting $(i)_2$ such that $(\exists j)\mathbf{t}(i) = \mathbf{t}(j)$. (This automaton is not minimized.)

To compute the automaton in Figure 4 we split up our task into three parts. First, we computed the automata for $e(n)$ and $o(n)$. Both of these operations took

14   *D. Goč, N. Rampersad, M. Rigo, P. Salimov*

about 0.025 seconds to compute and both can be expressed by automata with 2 states.

Second, we computed the automaton for $B(i, n, k)$. The resulting automaton had 108 states and took a bit over a second to compute. Due to its size, it cannot be displayed here.

Finally, using the automaton from the previous step we computed the automaton encoding $\phi(n)$ evaluating the expression in (4). We give the program output of this computation:

```
      [> 1 k] (6 => 1 states) in 0.012s
      [> 2*k n] (6 => 3 states) in 0.010s
     [\or ] (4 => 4 states) in 0.018s
      [\machine ] (108 => 108 states) in 0.043s
     [\not ] (109 => 107 states) in 0.062s
    [\or ] (215 => 215 states) in 0.113s
   [\not ] (216 => 215 states) in 0.113s
  [\exists k] (215 => 23 states) in 0.115s
 [\not ] (24 => 22 states) in 0.021s
[\exists i] (22 => 11 states) in 0.009s
[total] (11 states) in 0.589s
```

This output shows that the computation took a total of 0.589 seconds and the largest intermediate step was an automaton with 215 states.

References

[1] M. Barnabei, F. Bonetti, M. Silimbani, Restricted involutions and Motzkin paths, *Adv. in Appl. Math.* **47** (2011), 102–115.

[2] V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and p-recognizable sets of integers, *Bull. Belg. Math. Soc* **1** (1994), 191–238.

[3] E. Charlier, N. Rampersad, J. Shallit, Enumeration and decidable properties of automatic sequences, *Int. J. Found. Comput. Sci.*, **23** (2012), 1035–1066.

[4] W.-F. Chuan, Unbordered factors of the characteristic sequences of irrational numbers, *Theoret. Comput. Sci.* **205** (1998), 337–344.

[5] J. D. Currie, K. Saari, Least periods of factors of infinite words, *RAIRO Inform. Théor. App.* **43** (2009), 165–178.

[6] A. Ehrenfeucht, D. M. Silberger, Periodicity and unbordered segments of words, *Disc. Math.* **26** (1979), 101–109.

[7] D. Goč, Automatic Sequences and Decidable Properties: Implementation and Applications, Master's thesis, University of Waterloo, (2013).

[8] D. Goč, D. Henshall, J. Shallit, Automatic theorem-proving in combinatorics on words, in CIAA 2012, Lect. Notes in Comp. Science **7381** (2012), 180–191.

[9] T. Harju, D. Nowotka, Periodicity and Unbordered Words: A Proof of Duval's Conjecture, *J. ACM* **54** (2007).

[10] S. Holub, K. Saari, On highly palindromic words, *Disc. Appl. Math.* **157** (2009), 953–959.

[11] N. Rampersad, M. Rigo, P. Salimov, On the number of abelian bordered words, in DLT 2013, Lect. Notes in Comp. Science **7907** (2013), 420–432.

[12] A. Sapounakis, P. Tsikouras, On *k*-colored Motzkin words, *J. Integer Seq.* **7** (2004).

[13] L. Schaeffer, Deciding Properties of Automatic Sequences, Master's Thesis, University of Waterloo, 2013.

[14] Neil J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, *The OEIS Foundation Inc.*, http://oeis.org/.