Some Results on Words in Two Dimensions

Formal Languages & Automata Theory Seminar

Taylor J. Smith

Joint work with G. Gamard, G. Richomme, and J. Shallit

School of Computing
Queen's University
Kingston, Ontario, Canada

October 30, 2017

Table of Contents

Table of Contents

Background

- Combinatorics on words is a well-studied subfield of theoretical computer science, with its origins in the early 20th century.
- Many results in the one-dimensional case have appeared.
- However, the two-dimensional case is not as popular, even though many of the one-dimensional results seem naturally extendible to higher dimensions.
- In this presentation, we investigate various two-dimensional generalizations of some well-known properties of words.

- A **two-dimensional word**

$$A = \begin{bmatrix} a_{0,0} & \cdots & a_{0,n-1} \\ \vdots & \ddots & \vdots \\ a_{m-1,0} & \cdots & a_{m-1,n-1} \end{bmatrix}$$

  is a map from $\{0, 1, \ldots, m-1\} \times \{0, 1, \ldots, n-1\}$ to an alphabet Σ.
  - Also called an **array**, a **picture**, and a **figure** in the literature.
- The **set of two-dimensional words** $\Sigma^{m \times n}$ contains all two-dimensional words of dimension $m \times n$ over Σ.
  - We also have the sets Σ^{**} (all two-dimensional words over Σ) and Σ^{++} (all *nonempty* two-dimensional words over Σ).

Preliminaries

- A pair of two-dimensional words A and B may be **concatenated** in
  - the horizontal direction, denoted $A \ominus B$; or
  - the vertical direction, denoted $A \varobar B$.

- A pair of two-dimensional words A and B may be **concatenated** in
  - the horizontal direction, denoted $A \ominus B$; or
  - the vertical direction, denoted $A \oplus B$.

Example

Given

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}, \; B = \begin{bmatrix} 7 & 8 \end{bmatrix}, \text{ and } C = \begin{bmatrix} 3 \\ 6 \end{bmatrix},$$

we have that

$$A \ominus B = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \text{ and } A \oplus C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}.$$

Preliminaries

- Two-dimensional words may have **powers**, **prefixes**, and **suffixes**.
  - A prefix/suffix is **nontrivial** if it is nonempty.
  - A prefix/suffix is **proper** if it is not equal to the word itself.

- Two-dimensional words may have **powers**, **prefixes**, and **suffixes**.
  - A prefix/suffix is **nontrivial** if it is nonempty.
  - A prefix/suffix is **proper** if it is not equal to the word itself.

Example

Given $A = \begin{bmatrix} 4 & 6 \end{bmatrix}$, the 2×3 power of A is

$$A^{2\times 3} = \begin{bmatrix} 4 & 6 & 4 & 6 & 4 & 6 \\ 4 & 6 & 4 & 6 & 4 & 6 \end{bmatrix}.$$

$A^{2\times 3}$ has, among others, the prefix/suffix

$$B = \begin{bmatrix} 4 & 6 \\ 4 & 6 \end{bmatrix}.$$

- A two-dimensional word A is **primitive** if it cannot be written as a power; that is, $A \neq B^{p \times q}$ for some $B \in \Sigma^{++}$ with either $p \geq 2$ or $q \geq 2$.

- A two-dimensional word A is **primitive** if it cannot be written as a power; that is, $A \neq B^{p \times q}$ for some $B \in \Sigma^{++}$ with either $p \geq 2$ or $q \geq 2$.

Example

The two-dimensional word $B = \begin{bmatrix} 2 & 4 \end{bmatrix}$ is primitive.
The two-dimensional word

$$A = \begin{bmatrix} 2 & 4 \\ 2 & 4 \end{bmatrix}$$

is not primitive, since we can write $A = B^{2 \times 1}$.

- A two-dimensional word A is **bordered** if we can write

$$A = (Q \varobar R \varobar Q) \ominus (S \varobar T \varobar S) \ominus (Q \varobar R \varobar Q)$$

for $Q \in \Sigma^{++}$ and $R, S, T \in \Sigma^{**}$.

- A two-dimensional word A is **bordered** if we can write

$$A = (Q \varobar R \varobar Q) \ominus (S \varobar T \varobar S) \ominus (Q \varobar R \varobar Q)$$

for $Q \in \Sigma^{++}$ and $R, S, T \in \Sigma^{**}$.

Example

$$A = \begin{bmatrix} 7 & 4 & 1 & 7 & 4 \\ 6 & 8 & 0 & 6 & 8 \\ 3 & 2 & 9 & 3 & 2 \\ 7 & 4 & 1 & 7 & 4 \\ 6 & 8 & 0 & 6 & 8 \end{bmatrix}$$

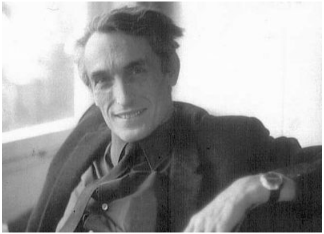
We see immediately that A is bordered.

Table of Contents

Background

- The **Lyndon-Schützenberger theorems** define a set of conditions for
  1. when a word has identical nontrivial proper prefixes and suffixes; and
  2. when the concatenation of two words x and y commutes; that is, when $xy = yx$.

R. C. Lyndon



M.-P. Schützenberger

1D First Lyndon-Schützenberger Theorem

Theorem

Let $y \in \Sigma^+$. Then the following are equivalent:

(1) *There exists $p \in \Sigma^+$ such that p is both a proper prefix and suffix of y;*

(2) *There exist $u \in \Sigma^+$, $v \in \Sigma^*$, and an integer $e \geq 1$ such that $y = (uv)^e u = u(vu)^e$.*

1D First Lyndon-Schützenberger Theorem

Theorem

Let $y \in \Sigma^+$. Then the following are equivalent:

(1) *There exists $p \in \Sigma^+$ such that p is both a proper prefix and suffix of y;*

(2) *There exist $u \in \Sigma^+$, $v \in \Sigma^*$, and an integer $e \geq 1$ such that $y = (uv)^e u = u(vu)^e$;*

(3) ***There exist $s \in \boldsymbol{\Sigma}^+$ and $t \in \boldsymbol{\Sigma}^*$ such that $y = sts$;***

(4) ***There exist $q \in \boldsymbol{\Sigma}^+$ and $r \in \boldsymbol{\Sigma}^*$ such that qr is a proper prefix of y and $qry = yrq$;***

(6) ***There exist a proper prefix $x \in \boldsymbol{\Sigma}^+$ of y, $w \in \boldsymbol{\Sigma}^*$, and an integer $i \geq 2$ such that $yw = x^i$.***

1D First Lyndon-Schützenberger Theorem

Theorem

Let $y \in \Sigma^+$. Then the following are equivalent:

(1) *There exists $p \in \Sigma^+$ such that p is both a proper prefix and suffix of y;*

(2) *There exist $u \in \Sigma^+$, $v \in \Sigma^*$, and an integer $e \geq 1$ such that $y = (uv)^e u = u(vu)^e$;*

(3) ***There exist $s \in \Sigma^+$ and $t \in \Sigma^*$ such that $y = sts$;***

(4) ***There exist $q \in \Sigma^+$ and $r \in \Sigma^*$ such that qr is a proper prefix of y and $qry = yrq$;***

(6) ***There exist a proper prefix $x \in \Sigma^+$ of y, $w \in \Sigma^*$, and an integer $i \geq 2$ such that $yw = x^i$.***

Remark

There exist conditions (5) and (7) which are analogous to conditions (4) and (6) for suffixes.

1D Second Lyndon-Schützenberger Theorem

Theorem

Let $x, y \in \Sigma^+$. Then the following are equivalent:

(1) $xy = yx$;

(2) *There exist $z \in \Sigma^+$ and integers $k, l > 0$ such that $x = z^k$ and $y = z^l$;*

(3) *There exist integers $i, j > 0$ such that $x^i = y^j$.*

Theorem

Let $x, y \in \Sigma^+$. Then the following are equivalent:

(1) $xy = yx$;

(2) *There exist* $z \in \Sigma^+$ *and integers* $k, l > 0$ *such that* $x = z^k$ *and* $y = z^l$;

(3) *There exist integers* $i, j > 0$ *such that* $x^i = y^j$;

(4) ***There exist integers*** $\boldsymbol{r, s > 0}$ ***such that*** $\boldsymbol{x^r y^s = y^s x^r}$;

(5) $\boldsymbol{x\{x, y\}^* \cap y\{x, y\}^* \neq \emptyset}$.

1D Second Lyndon-Schützenberger Theorem

Theorem

Let $x, y \in \Sigma^+$. Then the following are equivalent:

(1) $xy = yx$;

(2) *There exist* $z \in \Sigma^+$ *and integers* $k, l > 0$ *such that* $x = z^k$ *and* $y = z^l$;

(3) *There exist integers* $i, j > 0$ *such that* $x^i = y^j$;

(4) ***There exist integers*** $\boldsymbol{r, s > 0}$ ***such that*** $\boldsymbol{x^r y^s = y^s x^r}$;

(5) $\boldsymbol{x\{x, y\}^* \cap y\{x, y\}^* \neq \emptyset}$.

Remark

Condition (5) is essentially the **defect theorem** from the field of coding theory.

2D First Lyndon-Schützenberger Theorem

- We can extend the first Lyndon-Schützenberger theorem to two dimensions by
  - considering two-dimensional overlapping words; or
  - considering two-dimensional bordered words.
- The overlapping extension is not very interesting.
  - Simply apply the 1D version of the theorem to each row/column of the pair of two-dimensional words.
- We will focus on the bordered extension.

2D First Lyndon-Schützenberger Theorem

Theorem

*Let $A \in \Sigma^{m \times n}$ be a nonempty two-dimensional **bordered** word. Then the following are equivalent:*

(1) *There exist $P_1, P_2 \in \Sigma^{++}$ such that P_1 is a proper prefix/suffix of A horizontally and P_2 is a proper prefix/suffix of A vertically;*

(2) *There exist $U_1, U_2 \in \Sigma^{++}$, $V_1, V_2 \in \Sigma^{**}$, and integers $e, f \geq 1$ such that $A = (U_1 \ominus V_1)^e \ominus U_1 = (U_2 \oplus V_2)^f \oplus U_2$;*

(3) *There exist $S_1, S_2 \in \Sigma^{++}$ and $T_1, T_2 \in \Sigma^{**}$ such that $A = S_1 \ominus T_1 \ominus S_1 = S_2 \oplus T_2 \oplus S_2$;*

Theorem (Cont.)

Let $A \in \Sigma^{m \times n}$ *be a nonempty two-dimensional* ***bordered*** *word. Then the following are equivalent:*

(4) *There exist* $U_1, U_2 \in \Sigma^{++}$ *and* $V_1, V_2 \in \Sigma^{**}$ *such that* $U_1 \ominus V_1 \ominus A = A \ominus V_1 \ominus U_1$ *and* $U_2 \oplus V_2 \oplus A = A \oplus V_2 \oplus U_2$*;*

(5) *There exist* $X_1, X_2 \in \Sigma^{++}$*, which are proper prefixes of* A *horizontally and vertically, respectively;* $Z_1, Z_2 \in \Sigma^{**}$*; and integers* $i_1, i_2 \geq 2$ *such that* $A \ominus Z_1 = X_1^{i_1 \times 1}$ *and* $A \oplus Z_2 = X_2^{1 \times i_2}$*;*

(6) *There exist* $R_1, R_2 \in \Sigma^{++}$*, which are proper suffixes of* A *horizontally and vertically, respectively;* $W_1, W_2 \in \Sigma^{**}$*; and integers* $j_1, j_2 \geq 2$ *such that* $W_1 \ominus A = R_1^{j_1 \times 1}$ *and* $W_2 \oplus A = R_2^{1 \times j_2}$*.*

Theorem

Let A and B be nonempty two-dimensional words. Then the following are equivalent:

(1) *There exist positive integers* p_1, p_2, q_1, q_2 *such that* $A^{p_1 \times q_1} = B^{p_2 \times q_2}$.

(2) *There exist* $C \in \Sigma^{++}$ *and positive integers* r_1, r_2, s_1, s_2 *such that* $A = C^{r_1 \times s_1}$ *and* $B = C^{r_2 \times s_2}$.

(3) *There exist positive integers* t_1, t_2, u_1, u_2 *such that* $A^{t_1 \times u_1} \circ B^{t_2 \times u_2} = B^{t_2 \times u_2} \circ A^{t_1 \times u_1}$ *where* $\circ \in \{\varobar, \ominus\}$.

Table of Contents

Primitive Enumeration

- Over an alphabet of size k, there are

$$\psi_k(n) = \sum_{d|n} \mu(d) k^{n/d}$$

1D primitive words of length n, where $\mu(d)$ is the **Möbius function**, defined by

$$\mu(n) = \begin{cases} 1, & \text{if } n \text{ has an even number of prime divisors;} \\ -1, & \text{if } n \text{ has an odd number of prime divisors; and} \\ 0, & \text{if } n \text{ is divisible by a square} > 1. \end{cases}$$

Example

Enumerating all primitive words of length 4 over a binary alphabet:

$$\begin{aligned}\psi_2(4) &= \sum_{d|4} \mu(d) 2^{4/d} \\ &= \mu(1)2^{4/1} + \mu(2)2^{4/2} + \mu(4)2^{4/4} \\ &= (1)(2^4) + (-1)(2^2) + (0)(2^1) \\ &= 16 \text{ total words } - \underbrace{4 \text{ non-primitive words}}_{\text{copies of } 00,01,10,11}\end{aligned}$$

Indeed, the 12 primitive words are 0001, 0010, 0011, 0100, 0110, 0111, 1000, 1001, 1011, 1100, 1101, and 1110.

- We can produce an analogous 2D formula that enumerates all two-dimensional primitive words of size $m \times n$.
- Before we continue, we require the following corollary of the 2D second Lyndon-Schützenberger theorem.

Corollary

Given $A \in \Sigma^{++}$, there exist a unique primitive $C \in \Sigma^{++}$ and positive integers i and j such that $A = C^{i \times j}$.

Theorem

Let $\psi_k(m, n)$ denote the number of two-dimensional primitive words of dimension $m \times n$ over a k-letter alphabet. Then

$$\psi_k(m, n) = \sum_{d_1 | m} \sum_{d_2 | n} \mu(d_1)\mu(d_2) k^{mn/(d_1 d_2)}.$$

Primitive Verification

- The literature features a good deal of previous work on pattern matching in two-dimensional words.
- However, none of this work is directly related to the matters of primitivity or periodicity.
- It would be desirable to have an (efficient) algorithm to check the primitivity of a two-dimensional word.

Primitive Verification

- Could we take the elements of the two-dimensional word in row-major/column-major order, then check if this resulting word is primitive?
- No, since this method does not work in some cases.

Primitive Verification

- Could we take the elements of the two-dimensional word in row-major/column-major order, then check if this resulting word is primitive?
- No, since this method does not work in some cases.

Example

The two-dimensional word $A = \begin{bmatrix} \mathtt{a} & \mathtt{a} \\ \mathtt{b} & \mathtt{b} \end{bmatrix}$ is not 2D primitive.

Its row-majorized word $A_{\mathrm{RM}} = [\mathtt{aa}][\mathtt{bb}]$ is 1D primitive.

Example

The two-dimensional word $A = \begin{bmatrix} \mathtt{a} & \mathtt{b} & \mathtt{a} \\ \mathtt{b} & \mathtt{a} & \mathtt{b} \end{bmatrix}$ is 2D primitive.

Its row-majorized word $A_{\mathrm{RM}} = [\mathtt{aba}][\mathtt{bab}]$ is not 1D primitive.

Primitive Verification

- Before we continue, we make the following observations.

Remark

- A word w is primitive if and only if w is not a subword of the word $w_F w_L$, where w_F is w with the first symbol removed and w_L is w with the last symbol removed.
- We can check this in linear time by using, for example, the Knuth-Morris-Pratt string-matching algorithm.
- There exists an algorithm $\textsc{1DPrimitiveRoot}(w)$ to obtain the primitive root of some word w.

- Before we continue, we require the following lemma.

Lemma

Let $A \in \Sigma^{m \times n}$. Let the primitive root of row i of A be r_i and the primitive root of column j of A be c_j. Then the primitive root of A has dimension $p \times q$, where

$$p = \mathrm{lcm}(|c_0|, |c_1|, \ldots, |c_{n-1}|)$$

and

$$q = \mathrm{lcm}(|r_0|, |r_1|, \ldots, |r_{m-1}|).$$

Primitive Verification

Theorem

It is possible to check whether a $m \times n$ two-dimensional word is primitive and to compute the primitive root in $O(mn)$ time, for fixed alphabet size.

Primitive Verification

Algorithm 1: Computing the primitive root of A

procedure $\textsc{2DPrimitiveRoot}(A)$
    for $0 \leq i < m$ **do**
        $r_i \leftarrow \textsc{1DPrimitiveRoot}(A[i, 0..n-1])$
    $q \leftarrow \mathsf{lcm}(|r_0|, |r_1|, \ldots, |r_{m-1}|)$
    for $0 \leq j < n$ **do**
        $c_j \leftarrow \textsc{1DPrimitiveRoot}(A[0..m-1, j])$
    $p \leftarrow \mathsf{lcm}(|c_0|, |c_1|, \ldots, |c_{n-1}|)$
    for $0 \leq i < p$ **do**
        for $0 \leq j < q$ **do**
            $C[i, j] \leftarrow A[i, j]$
    return (C, p, q)

Table of Contents

Bordered Enumeration

- The number of one-dimensional unbordered words of length n over an alphabet of size k satisfies

$$u_k(n) = \begin{cases} k, & \text{if } n = 1; \\ k(k-1), & \text{if } n = 2; \\ k \cdot u_k(n-1), & \text{if } n \geq 3 \text{ is odd}; \\ k \cdot u_k(n-1) - u_k(n/2), & \text{if } n \geq 4 \text{ is even}. \end{cases}$$

- The number of bordered words of length n is therefore $b_k(n) = k^n - u_k(n)$.
- How can we enumerate the number of two-dimensional unbordered words of size mn, $U_k(m, n)$?

- We say that a one-dimensional word w has period p if $w[i] = w[i+p]$ for all i.

Lemma

Let $1 \leq p < n$. A one-dimensional word w of length n has period p if and only if w has a border of length $n - p$.

Corollary

If a one-dimensional word has a border of length $> \lfloor n/2 \rfloor$, then it also has a shorter border.

Bordered Enumeration

Technique 1

- Use the inclusion-exclusion principle.
- Take a two-dimensional word A and consider each column of A to be a "symbol".
- If A is bordered, then each "symbol" is bordered.
- We use our lemma to determine the possible one-dimensional border lengths.

Example

Consider one-dimensional words of length 3. These words can only have period length 2. Given such a word, specifying 2 symbols in that word fixes the remaining symbol.

Removing this symbol from the word and considering each possible pair of remaining symbols as being members of an alphabet of 4 "symbols", we get

$$\begin{aligned} U_2(3, n) &= 2^{3n} - b_{2^2}(n) \\ &= 2^{3n} - b_4(n), \end{aligned}$$

where $m = 3$ and $n > 1$.

Queen's University

Technique 2

- Use polynomials.
- Find the most general word w of length m having all periods from a set of periods P.
- Consider all nonempty subsets S of P.
- Starting with $P(x) = 0$, add the term $(-1)^{|S|}x^{d(w)}$, where $d(w)$ denotes the number of distinct symbols in w.
  - This is another application of the inclusion-exclusion principle, but with a different approach.

Example

Let $m = 5$. Then $P = \{3, 4\}$.

- For $S_1 = \{3\}$, the most general word of length 5 with period 3 is 12312.
- For $S_2 = \{4\}$, the most general word of length 5 with period 4 is 12341.
- For $S_3 = \{3, 4\}$, the most general word of length 5 with periods 3 and 4 is 11211.

This gives $P(x) = -x^3 - x^4 + x^2$, so

$$\begin{aligned} U_2(5, n) &= 2^{5n} - b_{2^3}(n) - b_{2^4}(n) + b_{2^2}(n) \\ &= 2^{5n} - b_8(n) - b_{16}(n) + b_4(n). \end{aligned}$$

Bordered Verification

- It would again be desirable to have an (efficient) algorithm to check whether a given two-dimensional word is bordered.
- Recall the following results:

Lemma

Let $1 \leq p < n$. A one-dimensional word w of length n has period p if and only if w has a border of length $n - p$.

Corollary

A one-dimensional word w of length n has no period shorter than n if and only if w is unbordered.

Bordered Verification

- Before we continue, we make the following observations.

Remark

- There exists an algorithm $\textsc{1DPeriod}(w)$ to obtain the periods of a one-dimensional word w.
- This algorithm returns the periods as a bit vector P where the ith bit of the vector is 1 if a period of length i exists in the word and 0 otherwise.
- By our observation, this algorithm need only search for periods p of length $\lceil n/2 \rceil \leq p \leq n - 1$.

Bordered Verification

Theorem

It is possible to check whether a $m \times n$ two-dimensional word is bordered and compute the dimension of the largest border in $O(mn)$ time, for fixed alphabet size.

Bordered Verification

Algorithm 2: Computing the primitive root of A

procedure $\text{2DBorder}(A, m, n)$

for $0 \le i < m$ **do**

$P_i \leftarrow \text{1DPeriod}(A[i,\ 0..n-1])$

$P \leftarrow P \cap P_i$

if $P = \emptyset$ **then**

return "unbordered"

$d \leftarrow$ smallest period common to all P_i

for $0 \le j < n$ **do**

$Q_j \leftarrow \text{1DPeriod}(A[0..m-1,\ j])$

$Q \leftarrow Q \cap Q_j$

if $Q = \emptyset$ **then**

return "unbordered"

$e \leftarrow$ smallest period common to all Q_j

return $(m - e, n - d)$

Table of Contents

Conclusions

- Properties of two-dimensional words is an area ripe for investigation.
- We saw generalizations of the one-dimensional Lyndon-Schützenberger theorems and extensions of the theorems to two-dimensions.
- We showed methods of enumerating and verifying primitive words and bordered words in two dimensions.
- The algorithms to perform this verification are very efficient. (Linear time!)

Future Work

- Can we generalize properties of words (e.g., overlaps, borders) to words of dimension greater than 2?
- Is there a better method for enumerating all two-dimensional unbordered words of dimension $m \times n$ over a k-letter alphabet?

References

[1] G. Gamard, G. Richomme, J. Shallit, and T. J. Smith. Periodicity in rectangular arrays. *Inf. Proc. Lett.*, 118:58–63, 2017.

[2] R. C. Lyndon and M.-P. Schützenberger. The equation $a^M = b^N c^P$ in a free group. *Mich. Math. J.*, 9(4):289–298, 1962.

[3] T. J. Smith. Properties of two-dimensional words. Master's thesis, University of Waterloo, 2017.