Preliminaries on Partial Words

Francine Blanchet-Sadri

24th April 2008

This material is based upon work supported by the National Science Foundation under Grants CCR–9700228, CCF–0207673 and DMS–0452020.

The following link

`http://www.uncg.edu/mat/reu/resources`

contains useful information on relevant papers and recommended literature related to the tutorial.

1 Preliminaries on Partial Words

1.1 Alphabets, letters, and words

Let A be a nonempty finite set of symbols, which we call an alphabet. An element $a \in A$ is called a letter. A word over the alphabet A is a finite sequence of elements of A.

The empty word consists of no letters and is denoted by ε.

The set of all words over A is denoted by A^* and is equipped with the associative operation defined by the concatenation of two sequences. The empty word is the neutral element for concatenation, as

$$u\varepsilon = \varepsilon u = u$$

The set $A^+ = A^* \setminus \{\varepsilon\}$ is equipped with the structure of a semigroup and is called the free semigroup over A. The set A^* is equipped with the structure of a monoid and is called the free monoid over A.

1.2 Partial functions and partial words

A word of length n over an alphabet A can be defined by a total function

$$u : \{0, \ldots, n-1\} \to A$$

and is usually represented as

$$u = a_0 a_1 \ldots a_{n-1} \text{ with } a_i \in A$$

A partial word (or, pword) of length n over A is a partial function

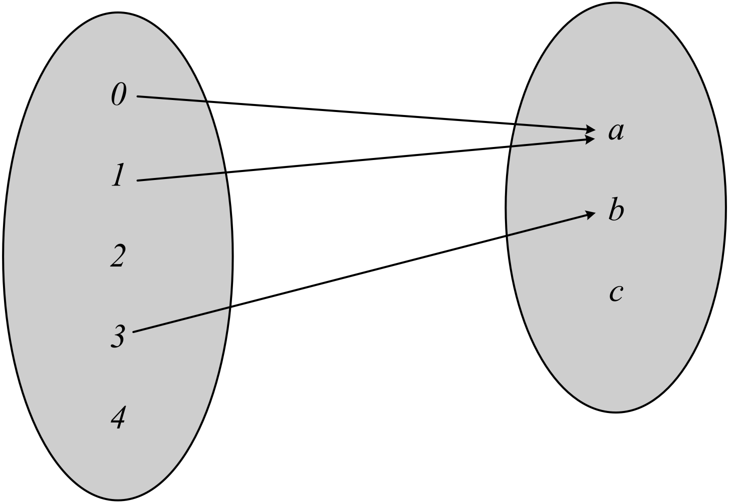
$$u : \{0, \ldots, n-1\} \to A$$

For $0 \leq i < n$,

- if $u(i)$ is defined, then $i \in D(u)$
- if $u(i)$ is not defined, then $i \in H(u)$

Every total word (or full word) is itself a partial word with an empty set of holes.

For any partial word u over A, $|u|$ or n denotes its length.


0
1
2
3
4
a
b
c

If u is a partial word of length n over A, then the companion of u, denoted by $u_\diamond$, is the total function

$$u_\diamond : \{0, \ldots, n-1\} \rightarrow A \cup \{\diamond\}$$

defined by

$$u_\diamond(i) = \begin{cases} u(i) & \text{if } i \in D(u) \\ \diamond & \text{otherwise} \end{cases}$$

$u_\diamond = abb\diamond b\diamond cb$ is the companion of the partial word u of length 8 where $D(u) = \{0, 1, 2, 4, 6, 7\}$ and $H(u) = \{3, 5\}$

The bijectivity of the map $u \mapsto u_\diamond$ allows us to define for partial words concepts such as concatenation, power, reversal, etc in a trivial way. More specifically, for partial words u, v:

- The concatenation of u and v, uv, is defined by $(uv)_\diamond = u_\diamond v_\diamond$,
- The i-power of u, u^i, is defined by $(u^i)_\diamond = (u_\diamond)^i$ where $(u_\diamond)^0 = \varepsilon$, and $(u_\diamond)^{i+1} = (u_\diamond)^i u_\diamond$,
- The reversal of u, $\text{rev}(u)$, is defined by $(\text{rev}(u))_\diamond = \text{rev}(u_\diamond)$ where $\text{rev}(u_\diamond)$ is $u_\diamond$ written backwards.

1.3 Periodicity

A period of a partial word u over A is a positive integer p such that

$$u(i) = u(j) \text{ whenever } i, j \in D(u) \text{ and } i \equiv j \bmod p$$

In such a case, we call u p-periodic.

$p(u)$ will denote the minimal period of u and $\mathcal{P}(u)$ the set of all periods of u

$u = a\diamond\diamond a\diamond b$ is 6-periodic, 4-periodic, and 3-periodic, and $p(u) = 3$

A weak period of a partial word u over A is a positive integer p such that

$$u(i) = u(i+p) \text{ whenever } i, i+p \in D(u)$$

In such a case, we call u weakly p-periodic.

We denote the set of all weak periods of u by $\mathcal{P}'(u)$ and the minimal weak period of u by $p'(u)$.

If $u = a\diamond\diamond a\diamond b$, then
$\mathcal{P}(u) = \{3, 4, 6\}$
$\mathcal{P}'(u) = \{1, 3, 4, 6\}$
$p(u) = 3$ and $p'(u) = 1$

1.4 FACTORIZATIONS OF PARTIAL WORDS

A factorization of a partial word u is any sequence $u_1, u_2, \ldots, u_i$ of pwords such that $u = u_1 u_2 \ldots u_i$. We write this factorization as $(u_1, u_2, \ldots, u_i)$.

The following are two factorizations of $u = abc\diamond ab$:

$$(ab, c\diamond, a, b)$$

$$(a, bc\diamond, ab)$$

A partial word u is a factor of a partial word v if there exist pwords x, y (possibly equal to ε) such that $v = xuy$. The factor u is proper if $u \neq \varepsilon$ and $u \neq v$. The partial word u is a prefix (respectively, suffix) of v if $x = \varepsilon$ (respectively, $y = \varepsilon$).

$u[i..j)$ is the factor of u starting at i and ending at $j-1$
$u[0..i)$ is the prefix of u of length i
$u[j..|u|)$ is the suffix of u of length $|u| - j$

Prefixes of $v = abc\diamond ab$ are ε, a, ab, abc, $abc\diamond$, $abc\diamond a$, $abc\diamond ab$.
Suffixes of $v = abc\diamond ab$ are ε, b, ab, $\diamond ab$, $c\diamond ab$, $bc\diamond ab$, and $abc\diamond ab$.

For partial words u and v, the unique maximal common prefix of u and v is denoted by $\text{pre}(u, v)$.

The common prefixes of $u = a\diamond bcb$ and $v = a\diamond bbab$ are $\varepsilon, a, a\diamond, a\diamond b$, the latter being $\text{pre}(u, v)$.

For a set X of partial words, we denote by $P(X)$ the set of prefixes of elements in X and by $S(X)$ the set of suffixes of elements in X:

$$P(X) = \{u \mid \text{there exists } x \text{ such that } ux \in X\}$$
$$S(X) = \{u \mid \text{there exists } x \text{ such that } xu \in X\}$$

$P(\{u\})$ (respectively, $S(\{u\})$) will be abbreviated by $P(u)$ (respectively, $S(u)$)

For a set X of partial words, we use the notation $\|X\|$ for the cardinality of X.

1.6 Containment and Compatiblity

If u and v are two partial words of equal length, then u is contained in v, denoted by $u \subset v$, if all elements in $D(u)$ are in $D(v)$ and $u(i) = v(i)$ for all $i \in D(u)$.

$$\begin{array}{rcccc} u = & a & \diamond & b & \diamond \\ v_1 = & a & \diamond & \diamond & b \end{array}$$

$$u \not\subset v_1$$

$$\begin{array}{rcccc} u = & a & \diamond & b & \diamond \\ v_2 = & a & \diamond & a & b \end{array}$$

$$u \not\subset v_2$$

$$\begin{array}{rcccc} u = & a & \diamond & b & \diamond \\ v_3 = & a & \diamond & b & b \end{array}$$

$$u \subset v_3$$

A partial word u is primitive if there exists no word v such that $u \subset v^i$ with $i \geq 2$.

$u = a\diamond ab$ is not primitive, because $u \subset (ab)^2$. However, $a\diamond bb$ is primitive

The partial words u and v are compatible, denoted by $u \uparrow v$, if there exists a partial word w such that $u \subset w$ and $v \subset w$.

$$
\begin{array}{lcccccc}
x = & a & \diamond & b & \diamond & a & \diamond \\
y = & a & \diamond & \diamond & c & b & b
\end{array}
$$

$$x \not\uparrow y$$

$$
\begin{array}{lcccccc}
u = & a & \diamond & b & b & c & \diamond \\
v = & \diamond & b & b & \diamond & c & \diamond
\end{array}
$$

$$u \uparrow v$$

Let u and v be partial words such that $u \uparrow v$. The least upper bound of u and v is the partial word $u \vee v$, where

$$u \subset u \vee v \text{ and } v \subset u \vee v, \text{ and}$$
$$D(u \vee v) = D(u) \cup D(v)$$

$$\begin{array}{cccccccc} u & = & a & b & a & \diamond & \diamond & a \\ v & = & a & \diamond & \diamond & b & \diamond & a \\ \hline u \vee v & = & a & b & a & b & \diamond & a \end{array}$$

For a set X of partial words, we denote by $C(X)$ the set of all partial words compatible with elements of X. More specifically,

$$C(X) = \{u \mid \text{there exists } v \in X \text{ such that } u \uparrow v\}$$

We denote $C(\{u\})$ simply by $C(u)$.

RULES

- Multiplication: If $u \uparrow v$ and $x \uparrow y$, then $ux \uparrow vy$.
- Simplification: If $ux \uparrow vy$ and $|u| = |v|$, then $u \uparrow v$ and $x \uparrow y$.
- Weakening: If $u \uparrow v$ and $w \subset u$, then $w \uparrow v$.

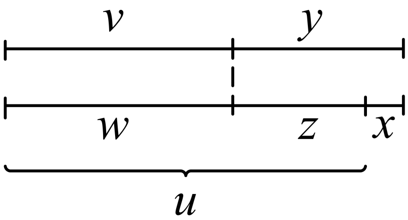
Lemma

Let u, v, x, y be partial words such that $ux \uparrow vy$.

- *If $|u| \geq |v|$, then there exist pwords w, z such that $u = wz$, $v \uparrow w$, and $y \uparrow zx$.*
- *If $|u| \leq |v|$, then there exist pwords w, z such that $v = wz$, $u \uparrow w$, and $x \uparrow zy$.*

J. Berstel and L. Boasson, Partial words and a theorem of Fine and Wilf, Theoretical Computer Science 218 (1999) 135–141.

Proof. We prove the first statement (the second one is similar). We use the following figure to illustrate our ideas:



If $|u| \geq |v|$, then set $u = wz$ with $|v| = |w|$. Then $wzx = ux \uparrow vy$ and the simplification rule gives the result. $\square$

2 Combinatorial Properties of Partial Words

http://www.uncg.edu/mat/research/equations

Lemma

Let x, y, z ($x \neq \varepsilon$ and $y \neq \varepsilon$) be words such that $xz = zy$. Then $x = uv$, $y = vu$, and $z = (uv)^n u$ for some words u, v and integer $n \geq 0$.

$x = abcda$, $y = daabc$, and $z = abc$

$xz = zy$, because

$(abcda)(abc) = (abc)(daabc)$

$u = abc$, $v = da$, and $n = 0$

Theorem

Let x, y, z be partial words with x, y nonempty. If $xz \uparrow zy$ and $xz \vee zy$ is $|x|$-periodic, then there exist words u, v such that $x \subset uv$, $y \subset vu$, and $z \subset (uv)^n u$ for some integer $n \geq 0$.

Let $x = \diamond ba$, $y = \diamond b \diamond$, and $z = b \diamond ab \diamond\diamond\diamond\diamond$. Then we have

$$\begin{array}{lcccccccccccc} xz & = & \diamond & b & a & b & \diamond & a & b & \diamond & \diamond & \diamond & \diamond \\ zy & = & b & \diamond & a & b & \diamond & \diamond & \diamond & \diamond & \diamond & b & \diamond \\ xz \vee zy & = & b & b & a & b & \diamond & a & b & \diamond & \diamond & b & \diamond \end{array}$$

It is clear that $xz \uparrow zy$ and $xz \vee zy$ is $|x|$-periodic. Putting $u = bb$ and $v = a$, we can verify that the conclusion does indeed hold.

F. Blanchet-Sadri and D.K. Luhmann, Conjugacy on partial words, Theoretical Computer Science 289 (2002) 297–312.

Corollary

Let x, y be nonempty partial words, and let z be a full word. If $xz \uparrow zy$, then there exist words u, v such that $x \subset uv$, $y \subset vu$, and $z \subset (uv)^n u$ for some integer $n \geq 0$.

Note that the above Corollary does not necessarily hold if z is not full even if x, y are full. The partial words $x = a, y = b$, and $z = \diamond bb$ provide a counterexample.

F. Blanchet-Sadri and D.K. Luhmann, Conjugacy on partial words, Theoretical Computer Science 289 (2002) 297–312.

Theorem

Let x, y and z be partial words such that $|x| = |y| > 0$. Then $xz \uparrow zy$ if and only if xzy is weakly $|x|$-periodic.

Proof. Let m be defined as $\lfloor \frac{|z|}{|x|} \rfloor$ and n as $|z|$ mod $|x|$. Then let $x = u_0 v_0, y = v_{m+1} u_{m+2}$ and $z = u_1 v_1 u_2 v_2 \ldots u_m v_m u_{m+1}$ where each u_i has length n and each v_i has length $|x| - n$.

u_0	v_0	u_1	v_1	$\ldots$	u_{m-1}	v_{m-1}	u_m	v_m	u_{m+1}
u_1	v_1	u_2	v_2	$\ldots$	u_m	v_m	u_{m+1}	v_{m+1}	u_{m+2}

Assume $xz \uparrow zy$. Therefore for all i such that $0 \leq i \leq m + 1$, $u_i \uparrow u_{i+1}$ and for all j such that $0 \leq j \leq m$, $v_j \uparrow v_{j+1}$. Thus $xz \uparrow zy$ implies that xzy is weakly $|x|$-periodic. Conversely, assume xzy is weakly $|x|$-periodic. This implies that $u_i v_i \uparrow u_{i+1} v_{i+1}$ for all i such that $0 \leq i \leq m$. Note that $u_{m+1} v_{m+1} u_{m+2}$ being weakly $|x|$-periodic, as a result $u_{m+1} \uparrow u_{m+2}$. This shows that $xz \uparrow zy$. □

F. Blanchet-Sadri, Dakota D. Blair and Rebeca V. Lewis, Equations on partial words, (http://www.uncg.edu/mat/research/equations).

Theorem

Let x, y and z be partial words such that $|x| = |y| > 0$. Then the following hold:

1. *If $xz \uparrow zy$, then xz and zy are weakly $|x|$-periodic.*
2. *If xz and zy are weakly $|x|$-periodic and $\lfloor \frac{|z|}{|x|} \rfloor > 0$, then $xz \uparrow zy$.*

The assumption $\lfloor \frac{|z|}{|x|} \rfloor > 0$ is necessary. To see this, consider $x = aa$, $y = ba$ and $z = a$. Here, xz and zy are weakly $|x|$-periodic, but $xz \not\uparrow zy$.

F. Blanchet-Sadri, Dakota D. Blair and Rebeca V. Lewis, Equations on partial words, (`http://www.uncg.edu/mat/research/equations`).

Let $x = ab\diamond d\diamond f$, $y = \diamond\diamond\diamond bc\diamond$, and $z = abcdefab\diamond defabcdefabcdefabcdefab\diamond d$. The figure displays the compatibility relation $xz \uparrow zy$ and highlights factorizations of x, y and z:

```
↑ ab^d ^f abcd ef ab^d ef abcd ef abcd ef abcd ef ab^d
  abcd ef ab^d ef abcd ef abcd ef abcd ef ab^d ^^ ^bc^
```

The concatenation xzy is seen to be weakly $|x|$-periodic:

a	b	$\diamond$	d	$\diamond$	f
a	b	c	d	e	f
a	b	$\diamond$	d	e	f
a	b	c	d	e	f
a	b	c	d	e	f
a	b	c	d	e	f
a	b	$\diamond$	d	$\diamond$	$\diamond$
$\diamond$	b	c	$\diamond$		

2.2 Commutativity

Theorem

Let x and y be nonempty words. Then $xy = yx$ if and only if there exists a word z such that $x = z^m$ and $y = z^n$ for some integers m, n.

For nonempty partial words x and y, if there exist a word z and integers m, n such that $x \subset z^m$ and $y \subset z^n$, then

$$xy \subset z^{m+n}$$
$$yx \subset z^{m+n}$$

and $xy \uparrow yx$. In addition, the converse holds as well, provided the partial word xy has at most one hole.

Theorem

Let x and y be nonempty partial words such that xy has at most one hole. If $xy \uparrow yx$, then there exists a word z such that $x \subset z^m$ and $y \subset z^n$ for some integers m, n.

J. Berstel and L. Boasson, Partial words and a theorem of Fine and Wilf, Theoretical Computer Science 218 (1999) 135–141.

The converse is not true in general:

$$x = \diamond bb \text{ and } y = abb\diamond$$

$$xy = \diamond bbabb\diamond \uparrow abb\diamond\diamond bb = yx$$

Our extension of commutativity is based on the concept of xy being $(|x|, |y|)$-SPECIAL where $|x| \leq |y|$.

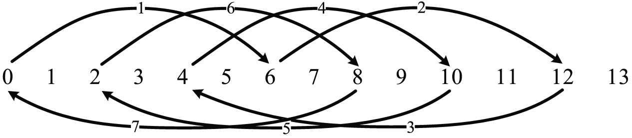
Let k, l be positive integers satisfying $k \leq l$. For $0 \leq i < k + l$,

$$\mathrm{seq}_{k,l}(i) = (i_0, i_1, i_2, \ldots, i_n, i_{n+1})$$

where

- $i_0 = i = i_{n+1}$
- for $1 \leq j \leq n$, $i_j \neq i$
- for $1 \leq j \leq n + 1$

$$i_j = \begin{cases} i_{j-1} + k & \text{if } i_{j-1} < l \\ i_{j-1} - l & \text{otherwise} \end{cases}$$



$$\mathrm{seq}_{6,8}(0) = (0, 6, 12, 4, 10, 2, 8, 0)$$

Let k, l be positive integers satisfying $k \leq l$ and let z be a partial word of length $k + l$. We say that z is (k, l)-special if there exists $0 \leq i < \text{gcd}(k, l)$ such that $\text{seq}_{k,l}(i) = (i_0, i_1, i_2, \ldots, i_n, i_{n+1})$ contains (at least) two positions that are holes of z while $z(i_0)z(i_1)z(i_2)\ldots z(i_{n+1})$ is not 1-periodic.

Let $z = cbca\diamond\diamond cbc\diamond caca$, and let $k = 6$ and $l = 8$ so $|z| = k + l$. We wish to determine if z is $(6, 8)$-special. First, we find $\text{seq}_{6,8}(0) = (0, 6, 12, 4, 10, 2, 8, 0)$ and

$$\begin{array}{cccccccc} z(0) & z(6) & z(12) & z(4) & z(10) & z(2) & z(8) & z(0) \\ c & c & c & \diamond & c & c & c & c \end{array}$$

This sequence does not satisfy the definition, and so we must continue with calculating $\text{seq}_{6,8}(1) = (1, 7, 13, 5, 11, 3, 9, 1)$. The corresponding letter sequence is

$$\begin{array}{cccccccc} z(1) & z(7) & z(13) & z(5) & z(11) & z(3) & z(9) & z(1) \\ b & b & a & \diamond & a & a & \diamond & b \end{array}$$

Here we have two positions in the sequence which are holes, and the sequence is not 1-periodic. Hence, z is $(6, 8)$-special.

Theorem

Let x, y be nonempty partial words such that $|x| \leq |y|$. If $xy \uparrow yx$ and xy is not $(|x|, |y|)$-special, then there exists a word z such that $x \subset z^m$ and $y \subset z^n$ for some integers m, n.

F. Blanchet-Sadri and Arundhati R. Anavekar, Testing primitivity on partial words,
(`http://www.uncg.edu/mat/primitive`).

Proof (sketch). Since $xy \uparrow yx$, there exists a word u such that $xy \subset u$ and $yx \subset u$. Put $|x| = k$ and $|y| = l$. Put $l = mk + r$ where $0 \leq r < k$. Either $r = 0$ or $r > 0$, and for each possibility the proof is split into three cases that refer to a given position i of u. Case 1 refers to $0 \leq i < k$, Case 2 to $k \leq i < l$, and Case 3 to $l \leq i < l + k$ (Cases 1 and 3 are symmetric as is seen by putting $i = l + j$ where $0 \leq j < k$). The following diagram pictures the containments $xy \subset u$ and $yx \subset u$:

xy	$x(0)$	$\ldots$	$x(k-1)$	$y(0)$	$\ldots$	$y(l-k-1)$	$y(l-k)$	$\ldots$	$y(l-1)$
yx	$y(0)$	$\ldots$	$y(k-1)$	$y(k)$	$\ldots$	$y(l-1)$	$x(0)$	$\ldots$	$x(k-1)$
u	$u(0)$	$\ldots$	$u(k-1)$	$u(k)$	$\ldots$	$u(l-1)$	$u(l)$	$\ldots$	$u(l+k-1)$

We prove the result for Case 1 under the assumptions that $r > 0$ and $i < r$. Here

- $x(i) \subset u(i)$ and $y(i) \subset u(i)$
- $y(i) \subset u(i+k)$ and $y(i+k) \subset u(i+k)$
- $y(i+k) \subset u(i+2k)$ and $y(i+2k) \subset u(i+2k)$
- $y(i+2k) \subset u(i+3k)$ and $y(i+3k) \subset u(i+3k)$
- $\vdots$
- $y(i+(m-1)k) \subset u(i+mk)$ and $y(i+mk) \subset u(i+mk)$
- $y(i+mk) \subset u(i+(m+1)k)$ and $x(i+k-r) \subset u(i+(m+1)k)$
- $x(i+k-r) \subset u(i+k-r)$ and $y(i+k-r) \subset u(i+k-r)$
- $y(i+k-r) \subset u(i+2k-r)$ and $y(i+2k-r) \subset u(i+2k-r)$
- $\vdots$

Let $x(i)y(i)y(i+k)\dots y(i+mk)x(i+k-r)\dots x(i) = v_i$. We claim that v_i is 1-periodic, say with letter a_i in $A \cup \{\diamond\}$. The claim follows from the above containments in case v_i has less than two holes. For the case where v_i has at least two holes, the claim follows since xy is not (k,l)-special. It turns out that $a_j = a_{j+r} = \cdots$ for $0 \leq j < r$. Let $z = a_0a_1\dots a_{r-1}$. If r divides k, then $x \subset z^{k/r}$ and $y \subset z^{(mk/r)+1}$. If r does not divide k, then z is 1-periodic with letter a say. In this case, $x \subset a^k$ and $y \subset a^l$.
□

Given $x = ab\diamond a\diamond\diamond a\diamond b$ and $y = a\diamond babba\diamond\diamond a\diamond b$, the alignment of xy and yx may be observed with the depiction in the figure. We can check that $xy \uparrow yx$ and also that xy is not $(|x|, |y|)$-special. Here $x \subset (abb)^3$ and $y \subset (abb)^4$.

```
↑  ab^ a^^ a^b a^b abb a^^ a^b
   a^b abb a^^ a^b ab^ a^^ a^b
```

Reference

F. Blanchet-Sadri, *Algorithmic Combinatorics on Partial Words*, Chapman & Hall/CRC Press, Boca Raton, FL, 2008.