Recurrent Partial Words

F. Blanchet-Sadri [1] A. Chakarov [2] L. Manuelli [3]
J. Schwartz [3] S. Stich [3]

[1]University of North Carolina at Greensboro

[2]University of Colorado at Boulder

[3]Princeton University

WORDS 2011

This material is based upon work supported by the National Science Foundation under Grant No. DMS–0754154. The Department of Defense is also gratefully acknowledged.

Outline

1. Preliminaries
2. Recurrent partial words
3. Completions of infinite partial words
4. Conclusion

1. Preliminaries

- A finite partial word of length n over an alphabet A is a function $w : \{0, \ldots, n-1\} \to A \cup \{\diamond\} = A_\diamond$.
- An infinite partial word over A is a function $w : \mathbb{N} \to A_\diamond$.
- In both the finite and infinite cases, if $w(i) \neq \diamond$, then i is defined in w, and if $w(i) = \diamond$, then i is a hole in w.
- If w has no holes, then w is a full word.
- A completion $\hat{w}$ is a “filling in” of w's holes with letters from A.

$w = abb\diamond b\diamond cb$ is a partial word of length 8 with holes at positions 3 and 5; $\hat{w} = abbabbcb$ is one of w's completions

Compatibility

The partial words u and v are compatible, denoted by $u \uparrow v$, if there exist completions $\hat{u}, \hat{v}$ such that $\hat{u} = \hat{v}$.

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

Periodicity

- A finite partial word w over A is called p-periodic, if p is a positive integer such that $w(i) = w(j)$ whenever i and j are defined in w and satisfy $i \equiv j \bmod p$. We say that w is periodic if it is p-periodic for some p.
- An infinite partial word w over A is called periodic if there exists a positive integer p (called a period of w) and letters $a_0, a_1, \ldots, a_{p-1} \in A$ such that for all $i \in \mathbb{N}$ and $j \in \{0, \ldots, p-1\}$, $i \equiv j \bmod p$ implies $w(i) \uparrow a_j$.

Shift function

If w is an infinite partial word, then we define the shift $\sigma_p(w)$ by

$$(\sigma_p(w))(i) = w(i + p)$$

Ultimate periodicity

- An infinite partial word w is called ultimately periodic if there exist a finite partial word u and an infinite periodic partial word v (both over A) such that $w = uv$.

- If w is a full ultimately periodic word, then $w = xy^\omega = xyyy\cdots$ for some finite words x, y with $y \neq \varepsilon$ called a period of w (we also call the length $|y|$ a period).

  If $|x|$ and $|y|$ are as small as possible, then y is called the minimal period of w.

Factor and subword

- A finite partial word u is a factor of the partial word w if u is a block of consecutive symbols of w.

  $\diamond a \diamond$ is a factor of $aa\diamond a\diamond b$

- A finite full word u is a subword of the partial word w, denoted $u \lhd w$, if u is a block of consecutive symbols of some completion of w.

  aaa, aab, baa, bab are the subwords of $aa\diamond a\diamond b$ corresponding to the factor $\diamond a\diamond$

Subword complexity

The subword complexity of a partial word w over a given alphabet is the function that assigns to each integer n, $0 \leq n \leq |w|$, the number $p_w(n)$ of distinct subwords of w of length n.

If $w = ba\diamond ab$, then $p_w(3) = 5$ since aaa, aab, aba, baa and bab are the subwords of length 3 of w.

Ferenczi's necessary conditions

Theorem

The following are necessary conditions for a function p_w from $\mathbb{N}$ to $\mathbb{N}$ to be the subword complexity function of an infinite partial word w over a finite alphabet A:

1. *p_w is non-decreasing;*
2. *$p_w(m+n) \leq p_w(m)p_w(n)$ for all m, n;*
3. *whenever $p_w(n) \leq n$ or $p_w(n+1) = p_w(n)$ for some n, then p_w is bounded;*
4. *if A has k letters, then $p_w(n) \leq k^n$ for all n; if $p_w(n_0) < k^{n_0}$ for some n_0, then there exists a real number $\kappa < k$ such that $p_w(n) \leq \kappa^n$ for all n sufficiently large.*

S. Ferenczi, Complexity of sequences and dynamical systems, *Discrete Mathematics* **206** (1999) 145–154.

2. Recurrent partial words

- An infinite partial word w is recurrent if every $u \in \mathrm{Sub}_w(n)$ occurs infinitely often in w.
- An infinite partial word w is uniformly recurrent, if for every $u \in \mathrm{Sub}_w(n)$, there exists $m \in \mathbb{N}$ such that every factor of length m of w has u as a subword, that is, $u \lhd w[0..m-1]$, $u \lhd w[1..m]$, ...

Clearly, a uniformly recurrent partial word is recurrent.

Equivalent formulations of recurrence

Proposition

Let w be an infinite partial word. The following are equivalent:

1. *The partial word w is recurrent;*
2. *Every subword compatible with a finite prefix of w occurs at least twice;*
3. *Every subword of w occurs at least twice.*

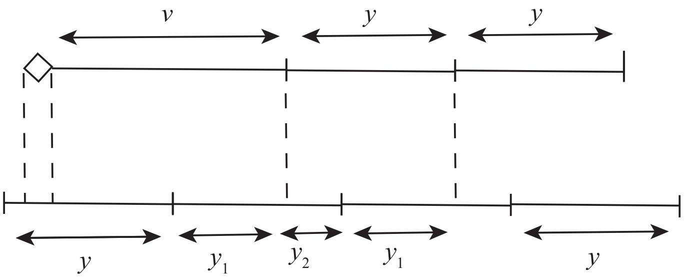
Theorem

If w is an infinite recurrent partial word with a positive but finite number of holes, then w is not ultimately periodic.

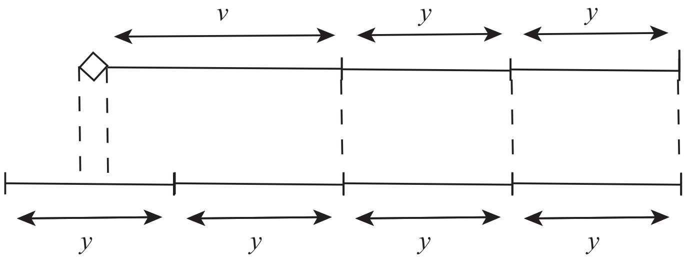
Proof.

- Suppose $w = xyyy\cdots$ where y is a finite full word such that $|y|$ is the minimal period of w.
- Let j be the position of the last hole in x. Let $z = ax[j+1..|x|)y^n = avy^n$ where $n \geq |y|$ and $a \neq y(j')$, where $j' = |y| - 1 - |v| \bmod |y|$.
- Since w is recurrent and z is a subword of w, z occurs in $u = y^\omega$. Thus, there exists $i \in \{0, \ldots, |y| - 1\}$ such that $u(i)\cdots u(i + |z| - 1) = z$. Since $y(i) = a \neq y(j')$, $i \neq j'$.
- Set $i' = (i + |v| + 1) \bmod |y|$, $y_1 = y(0)\cdots y(i' - 1)$, and $y_2 = y(i')\cdots y(|y| - 1)$. We get $y = y_1y_2 = y_2y_1$, and so y_1 and y_2 are powers of a common word y', $1 \leq |y'| < |y|$.
- However, $u = y^\omega = (y^{|y'|})^\omega = ((y')^{|y|})^\omega = (y')^\omega$ is $|y'|$-periodic, which contradicts the minimality of period $|y|$.

Proof (continued)



Proof (continued)



□

Gap function

To extend the above theorem to the case where w has infinitely many holes we need a definition.

- Let $H(n) - 1$ be the position of the nth hole in an infinite partial word w (we also say that $H(n)$ is the hole function of w).
- Let $h(n) = H(n) - H(n-1)$, for $n \geq 2$, be defined as the gap function of w.

$$\diamond\diamond a\diamond a\diamond aaa\diamond aaaaa\diamond aaaaaaaaaaa\diamond aaaaaaaaaaaaaaaaaaaa\diamond\cdots$$

has holes at positions $H(n) - 1 = \lceil 2^{4(n-1)/5} \rceil - 1$ and the distance between the 5th and 6th holes is $h(6) = H(6) - H(5) = 16 - 10 = 6$

The proof for the case where w has an eventually increasing gap function is analogous to the proof when w has only finitely many holes.

Corollary

Let w be a recurrent partial word with infinitely many holes for which there exists $N > 0$ such that $h(n) < h(n+1)$ for all $n \geq N$. Then w is not ultimately periodic.

We need the eventually increasing gap function restriction. Consider for example $w = ab\diamond^{\omega}$, which is ultimately periodic *and* recurrent.

The proof for the case where w has an eventually increasing gap function is analogous to the proof when w has only finitely many holes.

Corollary

Let w be a recurrent partial word with infinitely many holes for which there exists $N > 0$ such that $h(n) < h(n+1)$ for all $n \geq N$. Then w is not ultimately periodic.

We need the eventually increasing gap function restriction. Consider for example $w = ab\diamond^{\omega}$, which is ultimately periodic *and* recurrent.

Recurrence function

Let w be an infinite partial word. We define $R_w(n)$, the recurrence function of w, to be the smallest integer m such that every factor of length m of w contains at least one occurrence of every subword of length n of w.

Theorem

Let w be a uniformly recurrent infinite full word. Then the following hold:

1. $R_w(n+1) > R_w(n)$ *for all* $n \geq 0$;
2. $R_w(n) \geq p_w(n) + n - 1$ *for all* $n \geq 0$;
3. $R_w(n) \geq 2n$ *for all* $n \geq 0$ *if and only if* w *is not of the form* x^ω *for any non-empty finite word* x.

J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, 2003.

Basic properties of the recurrence function

Theorem

Let w be a uniformly recurrent infinite partial word. Then the following hold:

1. $R_w(n+1) > R_w(n)$ *for all* $n \geq 0$*;*
2. *If for each* $n > 0$ *there exists an index* i *such that* $w[i..i+n)$ *is a full word, then* $R_w(n) \geq p_w(n) + n - 1$ *for all* $n \geq 0$*;*
3. *If* w *has a positive finite number of holes or an eventually increasing gap function, then* $R_w(n) \geq 2n$ *for all* $n \geq 0$*.*

Uniformly recurrent partial words with finitely many holes cannot achieve maximal complexity.

Theorem

Let w be a uniformly recurrent infinite partial word with finitely many holes. Then there exists N such that $p_w(n) < k^n$ for all $n \geq N$, where k is the alphabet size.

The same is true for uniformly recurrent partial words with eventually increasing gap functions. We need this restriction since $w = \diamond^\omega$ is uniformly recurrent and achieves maximal complexity.

Corollary

Let w be a uniformly recurrent infinite partial word with eventually increasing gap function. Then there exists $N > 0$ such that $p_w(n) < k^n$ for all $n \geq N$, where k is the alphabet size.

Uniformly recurrent partial words with finitely many holes cannot achieve maximal complexity.

Theorem

Let w be a uniformly recurrent infinite partial word with finitely many holes. Then there exists N such that $p_w(n) < k^n$ for all $n \geq N$, where k is the alphabet size.

The same is true for uniformly recurrent partial words with eventually increasing gap functions. We need this restriction since $w = \diamond^\omega$ is uniformly recurrent and achieves maximal complexity.

Corollary

Let w be a uniformly recurrent infinite partial word with eventually increasing gap function. Then there exists $N > 0$ such that $p_w(n) < k^n$ for all $n \geq N$, where k is the alphabet size.

When we assume the usual restrictions on w, we find a strong relationship between w and its various completions $\hat{w}$.

Proposition

Let w be an infinite partial word having a finite number of holes or an eventually increasing gap function. Then w is recurrent if and only if every completion $\hat{w}$ is recurrent.

3. Completions of infinite partial words

We will consider the relationship between the complexity of an infinite partial word w and the complexities of its various completions $\hat{w}$.

- When does there exist a completion $\hat{w}$ attaining maximal complexity, i.e. attaining the complexity of w?
- How *close* can the complexity of a completion $\hat{w}$ come to the complexity of w?
- How close is *too* close?

It turns out that these questions are intimately related to the notion of recurrence.

Theorem

Let w be an infinite recurrent partial word. Then there exists a completion $\hat{w}$ of w such that $Sub(w) = Sub(\hat{w})$.

Proof.

- The set $\mathrm{Sub}(w)$ is countable, so choose some enumeration of its elements $x_0, x_1, x_2, \ldots$.
- Choose n_0 so that $x_0 \lhd w[0..n_0]$.
- Since x_1 occurs infinitely often in w, we can find some $n_1 > n_0$ so that $x_1 \lhd w(n_0..n_1]$.
- Similarly we can find some $n_2 > n_1$ so that $x_2 \lhd w(n_1..n_2]$ and so on for each x_i.
- We complete $w[0..n_0]$ so that it contains x_0 as a subword, $w(n_0..n_1]$ so that it contains x_1, and so on to get $\hat{w}$.
- By construction $\mathrm{Sub}(w) \subset \mathrm{Sub}(\hat{w})$ and we have $\mathrm{Sub}(\hat{w}) \subset \mathrm{Sub}(w)$.

□

- The condition that w be recurrent is sufficient for there to exist a completion $\hat{w}$ with $\mathrm{Sub}(\hat{w}) = \mathrm{Sub}(w)$.
- In the case where w has infinitely many holes, this turns out also to be necessary.

Theorem

Let w be a partial word with infinitely many holes. Then w is recurrent if and only if there exists a completion $\hat{w}$ such that $Sub(w) = Sub(\hat{w})$.

Proof.

- We have already shown the direction where we assume w to be recurrent.
- So suppose there exists a completion $\hat{w}$ such that $\mathrm{Sub}(w) = \mathrm{Sub}(\hat{w})$.
- We show that the prefix of length $H(n) - 1$ of $\hat{w}$ occurs twice for every $n \geq 1$.
- Choose $a \in A$ such that $a \neq \hat{w}(H(n) - 1)$. Then $v = \hat{w}[0..H(n) - 1)a \in \mathrm{Sub}(w) = \mathrm{Sub}(\hat{w})$. Hence v must occur in $\hat{w}$ but cannot occur as a prefix. Thus there exists $i > 0$ such that $\hat{w}[i..i + H(n)) = v$. But then $\hat{w}[i..i + H(n) - 1) = \hat{w}[0..H(n) - 1)$.
- Thus every prefix of $\hat{w}$ occurs twice and thus $\hat{w}$ is recurrent and since $\mathrm{Sub}(w) = \mathrm{Sub}(\hat{w})$, w is recurrent as well.

$\square$

- The condition that w has infinitely many holes is really needed in the previous theorem.
- Consider $w = \diamond a^\omega$ and $\hat{w} = ba^\omega$. Then $\mathrm{Sub}(w) = \mathrm{Sub}(\hat{w})$ but w is not recurrent since b occurs only once.
- Note however that $\sigma(w)$ is recurrent.
- This holds more generally but we need to introduce a new definition.

An infinite partial word w is ultimately recurrent if there exists an integer $p \geq 0$ such that $\sigma_p(w)$ is recurrent.

Corollary

Let w be an infinite partial word with at least one hole. If there exists a completion $\hat{w}$ of w such that $Sub(w) = Sub(\hat{w})$, then w is ultimately recurrent. In fact $\sigma_{H(1)}(w)$ is recurrent, where $H(n)$ is the hole function.

- $\mathrm{RSub}_w(n)$ denotes the set of recurrent subwords of length n of a partial word w.
- $\mathrm{RSub}(w) = \bigcup_{n\geq 1} \mathrm{RSub}_w(n)$.
- $r_w(n) = |\mathrm{RSub}_w(n)|$.
- $d_w(n) = p_w(n) - r_w(n)$ counts the number of non-recurrent subwords of length n.

The $d_w(n)$ function, which is non-decreasing, is important when studying ultimate recurrence.

Ultimate recurrence

The following proposition captures the fact that in an ultimately recurrent partial word with finitely many holes almost every subword is recurrent.

Proposition

Let w be an infinite partial word with finitely many holes. Then w is ultimately recurrent if and only if $d_w(n)$ is bounded.

The case when w has infinitely many holes is markedly different. In particular $d_w(n)$ cannot be positive and bounded.

Proposition

Let w be a partial word with infinitely many holes. Then $d_w(n)$ is either identically zero or unbounded.

Ultimate recurrence

The following proposition captures the fact that in an ultimately recurrent partial word with finitely many holes almost every subword is recurrent.

Proposition

Let w be an infinite partial word with finitely many holes. Then w is ultimately recurrent if and only if $d_w(n)$ is bounded.

The case when w has infinitely many holes is markedly different. In particular $d_w(n)$ cannot be positive and bounded.

Proposition

Let w be a partial word with infinitely many holes. Then $d_w(n)$ is either identically zero or unbounded.

If w is ultimately recurrent, then intuitively we expect w to have a large proportion of recurrent subwords.

Proposition

Let w be an ultimately recurrent infinite partial word. Then there exists a constant C such that $r_w(n) \leq p_w(n) \leq Cr_w(n)$ for all n sufficiently large. In other words, $p_w(n) = \Theta(r_w(n))$.

Since we can always find a completion that contains all the recurrent subwords, we have the following.

Corollary

Let w be an ultimately recurrent infinite partial word. Then there exists a completion $\hat{w}$ such that $p_w(n) = \Theta(p_{\hat{w}}(n))$.

If w is ultimately recurrent, then intuitively we expect w to have a large proportion of recurrent subwords.

Proposition

Let w be an ultimately recurrent infinite partial word. Then there exists a constant C such that $r_w(n) \leq p_w(n) \leq Cr_w(n)$ for all n sufficiently large. In other words, $p_w(n) = \Theta(r_w(n))$.

Since we can always find a completion that contains all the recurrent subwords, we have the following.

Corollary

Let w be an ultimately recurrent infinite partial word. Then there exists a completion $\hat{w}$ such that $p_w(n) = \Theta(p_{\hat{w}}(n))$.

The converse of the previous proposition does not hold. There exist partial words with infinitely many holes such that

- $p_w(n)$ is linear;
- $r_w(n)$ is linear;
- w is *not* ultimately recurrent.

We simply can consider words w that consist entirely of a's and $\diamond$'s, with hole function $H(n) = \lceil \alpha^n \rceil$ where $\alpha > 2$ is a real number.

It is easy to check that $r_w(n) = n + 1$ in this example.

Suppose w is a partial word with infinitely many holes.

- If $\hat{w}$ is a completion of w, then $p_{\hat{w}}(n) \leq p_w(n)$.
- If $p_{\hat{w}}(n) = p_w(n)$, then w is recurrent.
- Next best we can hope for is "off by a constant" complexity, i.e. $p_w(n) \leq p_{\hat{w}}(n) + C$ for all $n > 0$ and some constant C.
- This cannot happen in general, if $p_w(n) \leq p_{\hat{w}}(n) + C$ for all $n > 0$ and some constant C, then $p_w(n) = p_{\hat{w}}(n)$ and w must be recurrent.

Proposition

Let w be a partial word with infinitely many holes. If $\hat{w}$ is a completion of w such that $p_w(n) \leq p_{\hat{w}}(n) + C$ for all $n > 0$ and some constant C, then $Sub(w) = Sub(\hat{w})$ and thus $p_w(n) = p_{\hat{w}}(n)$.

In fact we can generalize the proposition, needing to stay close only for arbitrarily large n.

Corollary

Let w be a partial word with infinitely many holes. Suppose there exists a constant C such that for each $N > 0$ there exists a completion $\hat{w}$ such that $p_w(n) \leq p_{\hat{w}}(n) + C$ for all $n \geq N$. Then $p_w(n) = p_{\hat{w}}(n)$ and w is recurrent.

Proposition

Let w be a partial word with infinitely many holes. If $\hat{w}$ is a completion of w such that $p_w(n) \leq p_{\hat{w}}(n) + C$ for all $n > 0$ and some constant C, then $Sub(w) = Sub(\hat{w})$ and thus $p_w(n) = p_{\hat{w}}(n)$.

In fact we can generalize the proposition, needing to stay close only for arbitrarily large n.

Corollary

Let w be a partial word with infinitely many holes. Suppose there exists a constant C such that for each $N > 0$ there exists a completion $\hat{w}$ such that $p_w(n) \leq p_{\hat{w}}(n) + C$ for all $n \geq N$. Then $p_w(n) = p_{\hat{w}}(n)$ and w is recurrent.

- We can do better than the previous corollary.
- The proof relies on the fact that the holes introduce "variety". Thus if $p_{\hat{w}}(n)$ is *too* close to $p_w(n)$, then w is recurrent.

Proposition

Let w be an infinite partial word with hole function $H(n)$ and let φ be an increasing function. If for each $N > 0$ there exists a completion $\hat{w}$ such that $p_w(n) \leq p_{\hat{w}}(n) + \varphi(n)$ for all $n \geq N$ and $\lim_{n\to\infty} \frac{\varphi(H(n))}{k^n} = 0$, then $p_w(n) = p_{\hat{w}}(n)$ and w is recurrent.

- Another question is how $p_{\hat{w}}(n)$ relates to $r_w(n)$.
- If no completion has a complexity *too* much greater than $r_w(n)$, then w must be ultimately recurrent.

Theorem

Let w be an infinite partial word. Then w is ultimately recurrent if and only if for each completion $\hat{w}$ there exists a constant C such that $p_{\hat{w}}(n) \leq r_w(n) + C$ for all $n > 0$.

If w is ultimately recurrent, then the same C works for all completions, in other words the bound is uniform across completions.

4. Conclusion

- Completions can achieve complexities equal (or “close”) to that of the original partial word if and only if the word is recurrent or ultimately recurrent.
- How close $p_{\hat{w}}(n)$ can be to $p_w(n)$ without w being recurrent depends on the density of the holes in w.
- The slower $H(n)$ grows the farther away a non-maximal completion complexity $p_{\hat{w}}(n)$ must be from $p_w(n)$.
- There does not, in general, appear to be a relation between $r_w(n)$ and $p_w(n)$.
- For each $0 < \delta < 1$, we can find a partial word w with infinitely many holes such that

$$\lim_{n\to\infty} \frac{r_w(n)}{p_w(n)} = \delta$$