If L is ANY set then SUBSEQ(L) is Regular
Exposition by William Gasarch (gasarch@cs.umd.edu)

1 Introduction

Definition 1.1 Let Σ be a finite alphabet.

1. Let $w \in \Sigma^*$. $SUBSEQ(w)$ is the set of all strings you get by replacing some of the symbols in w with the empty string.
2. Let $L \subseteq \Sigma^*$. $SUBSEQ(L) = \bigcup_{w \in L} SUBSEQ(w)$.

The following are easy to show:

1. If L is regular than $SUBSEQ(L)$ is regular.
2. If L is context free than $SUBSEQ(L)$ is context free.
3. If L is c.e. than $SUBSEQ(L)$ is c.e.

Note that one of the obvious suspects is missing. Is the following true:

If L is decidable then $SUBSEQ(L)$ is decidable.

We will show something far stronger. We will show that

If L is ANY subset of Σ^* WHATSOEVER then $SUBSEQ(L)$ is regular.

Higman [2] first proved this theorem. His proof is the one we give here; however, he used different terminology.

The proofs that if L is regular (context free, c.e.) then $SUBSEQ(L)$ is regular (context free, c.e.) are constructive. That is, given the DFA (CFG, TM) for L you could produce the DFA (CFG, TM) for $SUBSEQ(L)$. (In the case of c.e. you are given M such that $L = DOM(M)$ and you can produce a TM M' such that $SUBSEQ(L) = DOM(M')$). The proof that if L is any language whatsoever then $SUBSEQ(L)$ is regular will be nonconstructive. We will discuss this later.

Definition 1.2 A set together with an ordering $(X, \preceq)$ is a *well quasi ordering* (wqo) if for any sequence $x_1, x_2, \ldots$ there exists i, j such that $i < j$ and $x_i \preceq x_j$. We call this i, j an *uptick*

Note 1.3 If $(X, \preceq)$ is a wqo then its both well founded and has no infinite antichains.

Lemma 1.4 *Let $(X, \preceq)$ be a wqo. For any sequence $x_1, x_2, \ldots$ there exists an infinite ascending subsequence.*

Proof: Let $x_1, x_2, \ldots,$ be an infinite sequence. Define the following coloring:
$COL(i, j) =$

- UP if $x_i \preceq x_j$.
- DOWN if $x_j \prec x_j$.
- INC if x_i and x_j are incomparable.

By Ramsey's theorem there is either an infinite homog UP-set, an infinite homog DOWN-set or an infinite homog INC-set. We show the last two cannot occur.

If there is an infinite homog DOWN-set then take that infinite subsequence. That subsequence violates the definition of well quasi ordering.

If there is an infinite homog INC-set then take that infinite subsequence. That subsequence violates the definition of well quasi ordering. ▮

We now redefine wqo.

Definition 1.5 A set together with an ordering $(X, \preceq)$. is a *well quasi ordering* (wqo) if one of the following equivalent conditions holds.

- For any sequence $x_1, x_2, \ldots$ there exists i, j such that $i < j$ and $x_i \preceq x_j$.
- For any sequence $x_1, x_2, \ldots$ there exists an *infinite* ascending subsequence.

Definition 1.6 If $(X, \preceq_1)$ and $(Y, \preceq_2)$ are wqo then we define $\preceq$ on $X \times Y$ as $(x, y) \preceq (x', y')$ if $x \preceq_1 y$ and $x' \preceq_2 y'$.

Lemma 1.7 *If $(X, \preceq_1)$ and $(Y, \preceq_2)$ are wqo then $(X \times Y, \preceq)$ is a wqo ($\preceq$ defined as in the above definition).*

Proof: Let $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots$ be an infinite sequence of elements from $A \times B$.

Define the following coloring:
$COL(i, j) =$

- UP-UP if $x_i \preceq x_j$ and $y_i \preceq y_j$.
- UP-DOWN if $x_i \preceq x_j$ and $y_j \preceq y_i$.
- UP-INC if $x_i \preceq x_j$ and y_j, y_i are incomparable.
- DOWN-UP, DOWN-DOWN, DOWN-INC, INC-UP, INC-DOWN, INC-INC are defined similarly.

By Ramsey's theorem there is a homog set in one of those colors. If the color has a DOWN in it then there is an infinite descending sequence within either $x_1, x_2, \ldots$, or $y_1, y_2, \ldots$ which violates either X or Y being a wqo. If the color has an INC in it then there is an infinite antichain within either $x_1, x_2, \ldots$, or $y_1, y_2, \ldots$ which violates either X or Y being a wqo. Hence the color must be UP-UP. This shows that there is an infinite ascending sequence. ▮

2 Subsets of Well Quasi Orders that are Closed Downward

Lemma 2.1 *Let $(X, \preceq)$ be a countable wqo and let $Y \subseteq X$. Assume that Y is closed downward under $\preceq$. Then there exists a finite set of elements $\{z_1, \ldots, z_k\} \subseteq X - Y$ such that*

$$y \in Y \text{ iff } (\forall i)[z_i \not\preceq y].$$

(The set $\{z_1, \ldots, z_k\}$ is called an obstruction set.)

Proof: Let OBS be the set of elements z such that

1. $z \notin Y$.
2. Every $y \preceq z$ is in Y.

Claim 1: OBS is finite
Proof: We first show that every $z, z' \in OBS$ are incomparable. Assume, by way of contradiction, that $z \preceq z'$. Then $z \in Y$ by part 2 of the definition of OBS. But if $z \in Y$ then $z \notin OBS$. Contradiction.

Assume that OBS is infinite. Then the elements of OBS (in any order) form an infinite anti-chain. This violates the property of $\preceq$ being a wqo. Contradiction.
End of Proof

Let $OBS = \{z_1, z_2, \ldots\}$. The order I put the elements in is arbitrary.
Claim 2: For all y:

$$y \in Y \text{ iff } (\forall i)[z_i \not\preceq y].$$

Proof of Claim 2:
We prove the contrapositive

$$y \notin Y \text{ iff } (\exists i)[z_i \preceq y].$$

Assume $y \notin Y$. If $y \in OBS$ then we are done. If $y \notin OBS$ then, by the definition of OBS there must be some z such that $z \notin Y$ and $z \prec y$. If $z \in OBS$ then we are done. If not then repeat the process with z. The process cannot go on forever since then we would have an infinite descending sequence, violating the wqo property. Hence, after a finite number of steps, we arive at an element of OBS. Therefore there is a $z \in OBS$ with $z \preceq y$.

Assume $(\exists i)[z_i \preceq y]$. Since Y is closed downward under $\preceq$ and $z_i \notin Y$, this implies that $y \notin Y$.
▮

3 $(\Sigma^*, \preceq_{\text{subseq}})$ is a Well Quasi Ordering

Definition 3.1 *The subsequence order*, which we denote $\preceq_{\text{subseq}}$, is defined as $x \preceq_{\text{subseq}}$‘$y$ if x is a subsequence of y.

IDEA: We will show that $(\Sigma^*, \preceq_{\text{subseq}})$ is a wqo. Note that if $A \subseteq \Sigma^*$ then $SUBSEQ(A)$ is closed under$\preceq_{\text{subseq}}$. Hence by the Lemma 2.1 there exists strings $z_1, \ldots, z_n$ such that
$x \in SUBSEQ(A)$ iff $(\forall i)[z_i \not\preceq x]$

For fixed z the set $\{x \mid z \not\preceq x\}$ is regular. Hence $SUBSEQ(A)$ is the intersection of a finite number of regular sets and is hence regular.

Theorem 3.2 $(\Sigma^*, \preceq)$ *is a wqo.*

Proof: Assume not. Then there exists (perhaps many) sequences $x_1, x_2, \ldots$ such that for all $i < j$, $x_i \not\preceq x_j$. We call such these *bad sequences.*

Look at ALL of the bad sequences. Look at ALL of the first elements of those bad sequences. Let y_1 be the *shortest* such element (if there is a tie then pick one of them arbitrarily).

Assume that $y_1, y_2, \ldots, y_n$ have been picked. Look at ALL of the bad sequences that begin $y_1, \ldots, y_n$ (there will be at least one). Look at ALL of the $n+1$st elements of those sequences. Let y_{n+1} be the shortest such element (if there is a tie then pick one of them arbitrarily). We have a sequence

$$y_1, y_2, \ldots$$

This is refered to as a *minimal bad sequence.*

Let $y_i = y_i'\sigma_i$ where $\sigma_i \in \Sigma$. (note that none of the y_i are empty since if they were they would not be part of any bad sequence).

Let $Y = \{y_1', y_2', \ldots\}$.

Claim: Y is a wqo.

Proof of Claim:

Assume not. Then there is a bad sequence $y_{k_1}', y_{k_2}', \ldots$ We know that $y_{k_i} = y_{k_i}'\sigma_{k_i}$. Lets say the bad sequence is

$$y_{84}', y_{12}', y_4', y_{1001}', y_{32}', \ldots \text{(no pattern is intended).}$$

Lets say that y_1', y_2', y_3' never appear. So y_4' is the least indexed element. We will remove all the elements before y_4'. Hence we can assume that the sequence starts with y_4'.

More generally, we will start the sequence at the least indexed element. We just assume this, so we assume that $k_1 \le \{k_2, k_3, \ldots\}$. Consider the following sequence:

$$y_1, y_2, \ldots, y_{k_1-1}, y_{k_1}', y_{k_2}', \ldots.$$

We show this is a BAD sequence.

There cannot be an $i < j \le k_1 - 1$ such that $y_i \preceq y_j$ since that would mean that $y_1, y_2, \ldots$ is not a bad sequence.

There cannot be an $i < j$ with $y_{k_i}' \preceq y_{k_j}'$ since that would mean that $y_{k_1}', y_{k_2}', \ldots$ is not a bad sequence.

And now for the interesting case. There cannot be an $i \le k_1 - 1$ and a k_j such that $y_i \preceq y_{k_j}'$. If we had this then we would have

$$y_i \preceq y_{k_j}' \preceq y_{k_j}' \sigma k_j = y_{k_j}.$$

But we made sure that $i < k_j$, so this would imply that $y_1, y_2, \ldots$ is not a bad sequence.

OKAY, so this is a bad sequence. So what? Well look— its a bad sequence that begins $y_1, y_2, \ldots, y_{k_1-1}$ but its k_1th element is y_{k_1}' which is SHORTER than y_{k_1}. This contradicts $y_1, y_2, \ldots,$ being a MINIMAL bad sequence.

End of Proof of Claim

So we know that Y is a wqo. We also know that Σ with any ordering is a wqo. By Lemma 1.7 $Y \times \Sigma$ is a wqo.

Look at the sequence

$$(y'_1, \sigma_1), (y'_2, \sigma_2), \ldots.$$

where $y_i = y'_i \sigma_i$.

Since Y is a wqo there exists $i < j$ such that

$$(y'_i, \sigma_i) \preceq_{\text{subseq}} (y'_j, \sigma_j), \ldots.$$

Clearly $y_i \preceq_{\text{subseq}} y_j$. ▮

4 Main Result

Theorem 4.1 *Let Σ be a finite alphabet. If $L \subseteq \Sigma^*$ then $SUBSEQ(L)$ is regular.*

Proof: Let $L \subseteq \Sigma^*$. The set $SUBSEQ(L)$ is closed under the $\preceq_{\text{subseq}}$ ordering. By Theorem 3.2 $\preceq_{\text{subseq}}$ is a wqo. By Lemma 2.1 $SUBSEQ(L)$ has a finite obstruction set. From this it is easy to show that $SUBSEQ(L)$ is regular. ▮

5 Nonconstructive?

One can ask: Given a DFA, CFG, P-machine, NP-machine, TM (decidable), TM (c.e.) for a language L, can one actually obtain a DFA for $SUBSEQ(L)$. For that matter, can you obtain a CFG, etc for $SUBSEQ(L)$.

	$SUBSEQ(REG)$	$SUBSEQ(CFGP$	$SUBSEQ(P)$	$SUBSEQ(DEC)$	$SUBSEQ(C.E.)$
REG	CON	CON	CON	CON	CON
CFG	CON	CON	CON	CON	CON
P	$NONCON$	$NONCON$	$NONCON$	$NONCON$	CON
NP	$NONCON$	$NONCON$	$NONCON$	$NONCON$	CON
DEC	$NONCON$	$NONCON$	$NONCON$	$NONCON$	CON
$C.E.$	$NONCON$	$NONCON$	$NONCON$	$NONCON$	CON

Gasarch, Fenner, Postow [1] showed all of the NONCON results. Leeuwen [3] showed that, given a CFG for L, you can obtain a DFA for $SUBSEQ(L)$ (it also appears in [1] which is online). All the rest of the results are easy.

References

[1] S. Fenner, W. Gasarch, and B. Postow. The complexity of finding SUBSEQ(A). *Theory of Computing Systems*, 45(3):577–612, October 2009. 10.1007/s00224-008-9111-4.

[2] A. G. Higman. Ordering by divisibility in abstract algebra. *Proceedings of the London Mathematical Society*, 3:326–336, 1952.

[3] J. van Leeuwen. Effective constructions in well-partially-ordered free monoids. *Discrete Mathematics*, 21:237–252, 1978.