Downward Closure of Any Language is Regular

2025-06-06

Let Σ be a finite alphabet, and let $L \subseteq \Sigma^*$ be an arbitrary language. We show that the language consisting of *subsequences* of strings in L is regular. The next section formally defines what we are trying to prove. The material is essentially [1].

1 Introduction

We will treat subsequences using orderings. Define the partial order $\leqslant$ in Σ^* where

$$x \leqslant y \text{ iff } x \text{ is a subsequence of } y.$$

Definition 1A (Closures). For language $L \subseteq \Sigma^*$, let $L\uparrow$ denote the **upward closure**, and $L\downarrow$ denote the **downward closure**, define as

$$L\uparrow = \{y \in \Sigma^* : y \geqslant x \text{ for some } x \in L\}$$
$$L\downarrow = \{y \in \Sigma^* : y \leqslant x \text{ for some } x \in L\}$$

By abuse of notation, we also define $x\uparrow \overset{\text{def}}{=} \{y \in \Sigma^* : y \geqslant x\}$ and $x\downarrow \overset{\text{def}}{=} \{y \in \Sigma^* : y \leqslant x\}$.

The main theorem can now be stated as follows:

Theorem 1A. For any $L \subseteq \Sigma^*$, the languages $L\uparrow$ and $L\downarrow$ are regular.

2 Proof of Theorem 1A

To prove Theorem 1A, we start with a proposition and some lemmas.

Proposition 2A. If $L \subseteq \Sigma^*$ is regular, then so are $L\uparrow$ and $L\downarrow$.

Proof. Let M be an DFA that accepts L. We construct NFAs by adding transitions to M:

- To accept $L\uparrow$, for each $a \in \Sigma$, add a new a-transition from every state to itself.
- To accept $L\downarrow$, for each transition, say, p to q, in M, add a new ε-transition from p to q.

□

Definition 2A. A subset $A \subseteq \Sigma^*$ is an **antichain** if its elements are pairwise incomparable. In other words, $x \not\leqslant y$ and $y \not\leqslant x$ for all $x, y \in A$.

Lemma 2A. Every antichain of Σ^* is finite.

Lemma 2B. For every language $L \subseteq \Sigma^*$, there exist finite $F, G \subseteq \Sigma^*$ such that

$$L\uparrow = F\uparrow \text{ and } L\downarrow = \Sigma^* \setminus G\uparrow.$$

We show that Lemma 2A $\Longrightarrow$ Lemma 2B $\Longrightarrow$ Theorem 1A. The proof of Lemma 2A is the most complicated, so we leave it at the end. To conclude this section, we prove Theorem 1A from Lemma 2B.

Theorem. $L\uparrow$ and $L\downarrow$ are regular for any $L \subseteq \Sigma^*$.

Proof. Let $L\uparrow = F\uparrow$ and $L\downarrow = \Sigma^* \setminus G\uparrow$ for finite $F, G \subseteq \Sigma^*$ as per Lemma 2B. Since $F\uparrow$ and $G\uparrow$ are both regular (Proposition 2A), so are $L\uparrow$ and $L\downarrow$. □

3 Proof of Lemma 2B

Let $L \subseteq \Sigma^*$, the proof of Lemma 2B consists of two parts.

Lemma. There exists a finite $F \subseteq \Sigma^*$ such that $L\uparrow = F\uparrow$.

Proof. Let $F \subseteq L$ be the set of minimal elements of L:

$$F = \{x \in L : \nexists y \in L \text{ s.t. } y < x\}.$$

It follows that F is finite (by Lemma 2A), and $F\uparrow = L\uparrow$. □

Lemma. There exists a finite $G \subseteq \Sigma^*$ such that $L\downarrow = \Sigma^* \setminus G\uparrow$.

Proof. Let $B = \Sigma^* \setminus L\downarrow$, then $B \subseteq B\uparrow$ by definition. Suppose, for contradiction, $B\uparrow \nsubseteq B$, i.e. there exists $x \in B\uparrow \cap L\downarrow$. Since $x \in B\uparrow$, let $y \in B$ such that $y \leqslant x$. However, because x is also in $L\downarrow$, we have that $y \in (L\downarrow)\downarrow = L\downarrow$, contradicting that $y \in B$.

Thus $B = B\uparrow$. Pick a finite $G \subseteq \Sigma^*$ such that $G\uparrow = B\uparrow = \Sigma^* \setminus L\downarrow$, it follows that $L\downarrow = \Sigma^* \setminus G\uparrow$ as wanted. □

4 Proof of Lemma 2A

We are left to show that for a language L, the set of minimal elements is finite, or in general, every antichain of Σ^* is finite under the ordering $\leqslant$ (for subsequences). We start with a proposition, and prove the lemma by induction on alphabet size.

Definition 4A. A subset $C \subseteq \Sigma^*$ is a **chain** iff all elements of C are pairwise comparable.

Proposition 4A. If Lemma 2A holds, then every infinite subset of Σ^* contains an infinite chain.

Proof. By contradiction, suppose $C \subseteq \Sigma^*$ is infinite, but every chain of C is finite. Therefore, C has infinitely many chains. By Lemma 2A, C only has finitely many maximal elements, so there exists an element $x \in C$ that is the maximal element of infinitely many chains. Thus, infinitely many and hence arbitrarily long strings of Σ^* precede x, i.e. are subsequences of x: a contradiction. $\square$

Finally, we are ready to prove the lemma.

Lemma. Every antichain of Σ^* is finite.
Proof. By induction on $|\Sigma|$. Note that an alphabet of size 1 is trivial.

Assume the lemma holds for alphabet size $n-1$, but $A = \{y_i \in \Sigma^* : i > 1\}$ is an infinite antichain for $|\Sigma| = n$. There exists some shortest string x such that $x \not\leqslant y_i$ for all $i > 1$; if not, $\Sigma^* = A\!\downarrow\, \subseteq A$, and thus $A = \Sigma^*$. Furthermore, choose A such that x is of minimum length. Note that $x \neq \varepsilon$.

Let $\ell = |x|$ and write

$$x = a_1 a_2 \cdots a_\ell$$

for each $a_k \in \Sigma$. Notice that, if $\ell = 1$, then each $y_i \in (\Sigma \setminus \{a_1\})^*$, which contradicts the induction hypothesis.

By the choice of x (being the shortest), we have that $a_1 \cdots a_{\ell-1} \leqslant y_i$ for all but finitely many i, i.e. there exists $N \geqslant 1$ such that for all $i \geqslant N$, $a_1 \cdots a_{\ell-1} \leqslant y_i$. Without loss of generality, we can throw away the first N strings from A, and assume it holds for all i (which is still infinitely many). Therefore, for each i, there exists $y_{i_1}, y_{i_2}, \cdots, y_{i_\ell}$ such that

$$y_i = y_{i_1} a_1 y_{i_2} a_2 \cdots y_{i_{\ell-1}} a_{\ell-1} y_{i_\ell}.$$

where $y_{i_j} \in (\Sigma \setminus \{a_j\})^*$ for each $j < \ell$ (e.g. by choosing the shortest y_{i_1}, then the shortest y_{i_2}, and so on). It is also the case that $y_{i_\ell} \in (\Sigma \setminus \{a_\ell\})^*$ because otherwise $x \leqslant y_i$.

We proceed to throw away more strings from A. Formally, we construct a decreasing sequence of infinite index sets $N_1 \supseteq N_2 \supseteq \cdots \supseteq N_\ell$ such that for every $1 \leqslant j \leqslant \ell$ and $p, q \in N_j$, we have that $y_{p_j} \leqslant y_{q_j}$ whenever $p \leqslant q$.

Proof of Lemma 2A

Let $N_0 = \{i \in \mathbb{N} : i \geqslant 1\}$. Given N_{j-1}, define

$$A_j = \{y_{i_j} : i \in N_{j-1}\}.$$

If A_j is finite, then for some fixed string w, the index set $\{i \in N_{j-1} : y_{i_j} = w\} = N_j$ is infinite. If not, $A_j \subseteq (\Sigma \setminus \{a_j\})^*$ contains an infinite chain (by induction hypothesis)

$$y_{s_{1j}} \leqslant y_{s_{2j}} \leqslant \cdots$$

and it suffices to let N_j be an infinite increasing subsequence of $s_1, s_2, \cdots$.

Lastly, for $p < q$ belonging to N_ℓ, we have that $p, q \in N_j$ for all $1 \leqslant j < \ell$ as well. So $y_{p_j} \leqslant y_{q_j}$ for all $1 \leqslant j \leqslant \ell$, and

$$\begin{aligned} y_p &= y_{p_1} a_1 y_{p_2} a_2 \cdots y_{p_{\ell-1}} a_{\ell-1} y_{p_\ell} \\ &\leqslant y_{q_1} a_1 y_{q_2} a_2 \cdots y_{q_{\ell-1}} a_{\ell-1} y_{q_\ell} = y_q, \end{aligned}$$

contradicting that A (containing y_p and y_q) is an antichain. □

Bibliography

[1] L. H. Haines, “On free monoids partially ordered by embedding,” *Journal of Combinatorial Theory*, vol. 6, no. 1, pp. 94–98, Jan. 1969, doi: 10.1016/S0021-9800(69)80111-0.