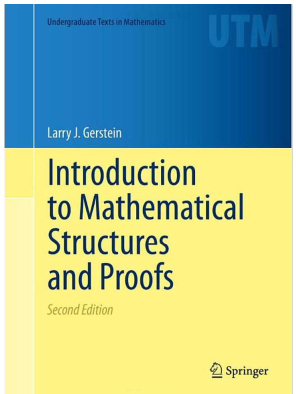
Mathematical Reasoning

Chapter 4. Finite and Infinite Sets

4.5. Languages and Finite Automata—Proofs of Theorems



Theorem 4.49

Theorem 4.49. If Σ is an alphabet then Σ^* is countably infinite.

Partial Proof. Let $\Sigma = \{s_1, s_2, \ldots, s_t\}$. Then Σ^* contains the countably infinite subset of words $S = \{s_1, s_1s_1, s_1s_1s_1, \ldots\}$. So there is an injection from this countably infinite set to Σ^* (the inclusion mapping), and so by Definition 4.27 we have $\#S \leq \#(\Sigma^*)$, or $\aleph_0 \leq \#(\Sigma^*)$. Now we put the *lexicographic ordering* on Σ by declaring that s_i precedes s_j if and only if $i < j$. This gives us a strategy to list the words in Σ^* based on their lengths. We start with $\varnothing$ (a word of length 0), then list the words of length 1 in lexicographic order, then list the words of length 2 lexicographically based on the first letter and then the second letter, and so forth. (This is a bit informal and so puts the "Partial" in "Partial Proof.")

Theorem 4.49 (continued)

Theorem 4.49. If Σ is an alphabet then Σ^* is countably infinite.

Partial Proof (continued). We then have:

$$\varepsilon, \underbrace{s_1, s_2, \ldots, s_t}_{\text{length 1}}, \underbrace{s_1s_1, s_1s_2, \ldots, s_1s_t, \ldots, s_ts_1, s_t, s_2, \ldots, s_ts_t}_{\text{length 2}},$$

$$\underbrace{s_1s_1s_1, s_1s_1s_2, \ldots, s_ts_ts_t}_{\text{length 3}}, \ldots.$$

Since this every would of any (finite) length is in this list exactly once, then we can set up a bijection from $\mathbb{N}$ to Σ^*. Therefore $\Sigma^* \approx \mathbb{N}$ and Σ^* is countable, as claimed. □

Theorem 4.51

Theorem 4.51. Let Σ be any alphabet. Then there are uncountably many languages over Σ.

Proof. Since a language L is just a subset of Σ^*, then the set of all languages L over σ is the power set of Σ^*, $P(\Sigma^*)$. By Cantor's Theorem (I) (Theorem 4.31), we have $\#(\Sigma^*) < \#P(\Sigma^*)$. But $\#(\Sigma^*)$ is countably infinite by Theorem 4.49, $\#(\Sigma^*) = \aleph_0$, and so the set of all languages $P(\Sigma^*)$ is uncountable, as claimed. □

Lemma 4.60. Pumping Lemma

Lemma 4.60. Pumping Lemma.
Let Σ be an alphabet, and let $L \subseteq \Sigma^*$ be an infinite regular language. Then there are strings x, y, z in Σ^*, with $y \neq \varepsilon$, such that for every $n \in \mathbb{N}$ the string xy^nz belongs to L.

Proof. By the definition of regular, there is a finite automaton M such that $L = L(M)$. Suppose M has exactly k states. Since L is infinite, there is a word $w \in L = L(M)$ of length greater than k. When M reads the symbols of w, it transitions from state to state going through the length of w (which is greater than k) states. So by the Pigeonhole Principle (Corollary 4.9), there must be some state q that M enters twice.

Lemma 4.60. Pumping Lemma (continued)

Lemma 4.60. Pumping Lemma.
Let Σ be an alphabet, and let $L \subseteq \Sigma^*$ be an infinite regular language. Then there are strings x, y, z in Σ^*, with $y \neq \varepsilon$, such that for every $n \in \mathbb{N}$ the string xy^nz belongs to L.

Proof (continued). So there are strings x, y, z, with $y \neq \varnothing$, such that concatenation gives $w = xyz$ and such that M is in state q before and after reading y. Then M accepts the word $xyyz$ (because M is in state q before reading y the first time, back in state q after reading y, and back to state q after reading y the second time; then M is in a final state after reading x since $w = xyz \in L(M)$). Similarly, by the Principle of Mathematical Induction (Theorem 2.66) we have that M accepts the word xy^nz for all $n \in \mathbb{N}$, as claimed. □