Higman's Lemma: Proof

Khushraj

October 15, 2025

Statement of Higman's Lemma

Let X be a finite alphabet. The set X^* of all finite words over X is partially ordered by the *subsequence relation*: $u \sqsubseteq v$ if u is a subsequence of v (obtained by deleting some symbols).
Given a relation $\leq$ between elements of set X, we define the embedding order (or relation) wrt $\leq$ (denoted as $\sqsubseteq_{\leq}$) between elements of set X^* as follows. $a_1 a_2 \dots a_n \sqsubseteq b_1 b_2 \dots b_m$ iff there exists $i_1, i_2, \dots i_n$ s.t. $1 \leq i_1 < i_2 < \dots i_n \leq m$ and $a_1 \leq b_{i_1}$, $a_2 \leq b_{i_2}$, ..., $a_n \leq b_{i_n}$. Notice that $\sqsubseteq_{=}$ is the usual subsequence relation.
Higman's Lemma: If $(X, \leq)$ is well-quasi-ordered (WQO), then $(X^*, \sqsubseteq_{\leq})$ is WQO under the embedding order (subsequence with $\leq$ on symbols).

Key Concepts

- *Quasi-order:* Reflexive and transitive relation.
- *Well-quasi-order (WQO):* No infinite strictly descending chains and no infinite antichains.
- *Bad sequence:* Infinite sequence where no earlier element embeds into a later one (i.e., $\neg$ WQO).
- *Embedding order $\sqsubseteq_{\leq}$:* $u = u_1, u_2, \ldots, u_n$ embeds into v if for some subsequence $v' = v'_1, v'_2, \ldots v'_n$ of v with each symbol $\leq$ the corresponding one in $u_i \leq v'_i$ for all $i \in [1, n]$. .

Proof: Setup (Contradiction)

Assume X^* is not WQO: There exists a bad infinite sequence $W = (w_1, w_2, \dots)$ with $w_i \not\sqsubseteq_{\leq} w_j$ for all $i < j$.
Among all bad sequences, choose W *minimal* lexicographically by word lengths:

- w_1 has minimal possible length to start a bad sequence.
- w_2 minimal to extend (w_1), etc.

W cannot start with the empty word (embeds everywhere). Thus, for each i, $w_i = a_i z_i$ where $a_i \in X$, $z_i \in X^*$ $(|z_i| < |w_i|)$.

Subsequence of Leading Symbols

Consider $(a_1, a_2, \dots) \in X \times X \times \dots$. Since X is WQO, it has an infinite non-decreasing subsequence $a_{i_1} \leq a_{i_2} \leq \dots$ with $i_1 < i_2 < \dots$.
Form $W' = (w_1, \dots, w_{i_1 - 1}, z_{i_1}, z_{i_2}, z_{i_3}, \dots)$.
W' is "smaller" than W in minimality order:

- Agrees up to $i_1 - 1$.
- At i_1: $|z_{i_1}| < |w_{i_1}|$.
- Later: $|z_{i_k}| < |w_{i_k}|$.

W' is Not Bad

By minimality of W, W' cannot be bad: $\exists j < k$ such that $u = W'_j \sqsubseteq_{\leq} v = W'_k$.
All possible cases lead to contradiction:

1. Case 1: Both u, v in initial $\{w_1, \ldots, w_{i_1-1}\}$: Then $w_j \sqsubseteq_{\leq} w_k$ in W ($j < k$), contradicting badness of W.
2. Case 2: $u = w_j$ ($j < i_1$), $v = z_{i_k}$ ($k \geq 1$): $u \sqsubseteq_{\leq} z_{i_k}$ implies $u \sqsubseteq_{\leq} a_{i_k} z_{i_k} = w_{i_k}$ (extend embedding), so $w_j \sqsubseteq_{\leq} w_{i_k}$ in W, contradiction.

Contradiction (Remaining Cases)

- $u = z_{i_j}$, $v = z_{i_k}$ $(1 \leq j < k)$: $u \sqsubseteq_{\leq} v$ implies $a_{i_j} z_{i_j} = w_{i_j} \sqsubseteq_{\leq} a_{i_k} z_{i_k} = w_{i_k}$ (prepend $a_{i_j} \leq a_{i_k}$), so $w_{i_j} \sqsubseteq_{\leq} w_{i_k}$ in W, contradiction.
- $u = z_{i_1}$, $v = w_m$ $(m > i_1 - 1)$???

All cases covered: Embedding in W' implies one in W, contradicting badness of W. Thus, no minimal bad sequence exists, so X^* is WQO. *This "minimal bad sequence" technique is key in WQO proofs (e.g., Kruskal's theorem).*