Language Operations and a Structure Theory of ω-Languages

August 2, 2012

Introduction: $\mathcal{P}(\Sigma^*) \to \mathcal{P}(\Sigma^\omega)$

We have the standard $\mathcal{P}(\Sigma^*) \to \mathcal{P}(\Sigma^\omega)$ language operators:

1. $\mathrm{ext}(L) := \{\alpha \in \Sigma^\omega \mid \exists n\colon \alpha[0,n] \in L\} = L \cdot \Sigma^\omega$
2. $\widehat{\mathrm{ext}}(L) := \{\alpha \in \Sigma^\omega \mid \forall n\colon \alpha[0,n] \in L\}$
3. $\lim(L) := \{\alpha \in \Sigma^\omega \mid \forall N\colon \exists n > N\colon \alpha[0,n] \in L\} =$ $\{\alpha \in \Sigma^\omega \mid \exists^\omega n\colon \alpha[0,n] \in L\}$
4. $\widehat{\lim}(L) := \{\alpha \in \Sigma^\omega \mid \exists N\colon \forall n > N\colon \alpha[0,n] \in L\}$

Introduction: $\mathcal{P}(\mathcal{P}(\Sigma^*)) \to \mathcal{P}(\mathcal{P}(\Sigma^*))$

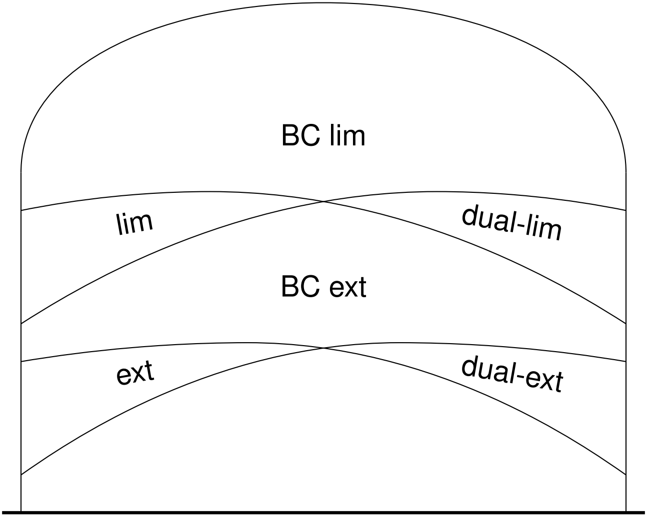
From these, define language class operators:

1. $\mathrm{ext}(\mathcal{L}) := \{\lim L \mid L \in \mathcal{L}\}$
2. $\widehat{\mathrm{ext}}(\mathcal{L}) := \left\{\widehat{\mathrm{ext}}\, L \;\middle|\; L \in \mathcal{L}\right\}$
3. $\lim(\mathcal{L}) := \{\lim L \mid L \in \mathcal{L}\}$
4. $\widehat{\lim}(\mathcal{L}) := \left\{\widehat{\lim}\, L \;\middle|\; L \in \mathcal{L}\right\}$

Boolean combinations:

1. $\mathrm{BC}\,\mathrm{ext}\,\mathcal{L} = \mathrm{BC}(\mathrm{ext}(\mathcal{L}))$
2. $\mathrm{BC}\lim\mathcal{L} = \mathrm{BC}(\lim(\mathcal{L}))$

$\mathcal{L}^*(\text{reg})$ inclusion diagram



Questions

- Instead of the class of regular $*$-languages, look at other $*$-language classes, e.g. starfree, LT, PT, or any arbitrary $*$-language class $\mathcal{L}$.
- For what $\mathcal{L}$ do we get the same relations as in the diagram? Are the inclusions still strict?

My diploma thesis:

- Chapter 3: general results on arbitrary $\mathcal{L}$, given some introduced properties on $\mathcal{L}$
- Chapter 4: concrete $*$-language classes

Properties on $\mathcal{L}$

1. $\mathcal{L}$ **closed under suffix-independence**: $L \in \mathcal{L} \Rightarrow L \cdot \Sigma^* \in \mathcal{L}$
   Examples: $\mathcal{L}^*(\mathrm{reg})$, $\mathcal{L}(\mathrm{starfree})$, $\mathcal{L}(\mathrm{PT}_n)$ (Lemma 4.10), $\mathcal{L}(\mathrm{LT})$
   Counter examples: $\mathcal{L}(\mathrm{finite})$, $\mathcal{L}(\mathrm{endwise})$, Example 3.4, Example 3.9
2. $\mathcal{L}$ **closed under union, intersection**
3. $\mathcal{L}$ **closed under negation**
4. $\mathcal{L}$ **closed under change of final states**: Let $\mathcal{A} = (Q, \Sigma, q_0, \delta, F)$ be a minimal deterministic automaton with $L^*(\mathcal{A}) \in \mathcal{L}$. Then, for all $F' \subseteq Q$, we have $L^*((Q, \Sigma, q_0, \delta, F')) \in \mathcal{L}$.
   Examples: $\mathcal{L}^*(\mathrm{reg})$, $\mathcal{L}(\mathrm{LT}_n)$, $\mathcal{L}(\mathrm{PT}_n)$, $\mathcal{L}(\mathrm{LT})$, $\mathcal{L}(\mathrm{PT})$, $\mathcal{L}(\mathrm{starfree})$ (Lemma 4.5)
5. $\mathcal{L}$ **closed under alphabet permutation**: For all permutations $\sigma : \Sigma \to \Sigma$ and $L \in \mathcal{L}$, we have $L_\sigma := \{\sigma(w) \mid w \in L\} \in \mathcal{L}$

General results: ext $\subseteq$ lim

- Lemma 3.3: $\mathcal{L}$ closed under suffix-independence $\Rightarrow$

$$\mathrm{ext}\,\mathcal{L} \subseteq \lim \cap \widehat{\lim}\,\mathcal{L}$$

  (but $\not\Leftarrow$, Example 3.4)
- Lemma 3.8: $\mathcal{L}$ closed under suffix-independence and negation $\Rightarrow$

$$\mathrm{ext} \cup \widehat{\mathrm{ext}}\,\mathcal{L} \subseteq \lim \cap \widehat{\lim}\,\mathcal{L}$$

Examples:

- $\mathcal{L}^*(\mathrm{reg})$, $\mathcal{L}(\mathrm{starfree})$, $\mathcal{L}(\mathrm{PT}_n)$, $\mathcal{L}(\mathrm{LT})$ are closed under suffix-independence and negation

$\mathrm{ext} \cup \widehat{\mathrm{ext}}\, \mathcal{L}^*(\mathrm{reg}) \subsetneqq \mathrm{BC}\, \mathrm{ext}\, \mathcal{L}^*(\mathrm{reg})$

- We have

$$\mathrm{ext} \cup \widehat{\mathrm{ext}}\, \mathcal{L} := (\mathrm{ext}\, \mathcal{L}) \cup (\widehat{\mathrm{ext}}\, \mathcal{L}) \subsetneqq \mathrm{BC}\, \mathrm{ext}\, \mathcal{L}$$

  for $\mathcal{L} = \mathcal{L}^*(\mathrm{reg})$.
- Separating languages: Let $\Sigma := \{a, b, c\}$.

$$L_a := \Sigma^* a \in \mathcal{L}, \quad L_b := \Sigma^* b \in \mathcal{L},$$

$$\tilde{L}_1 := \mathrm{ext}\, L_a \cap -\mathrm{ext}\, L_b, \quad \tilde{L}_2 := \lim L_a \cap -\lim L_b.$$

  Then

$$\tilde{L}_1 \notin \mathrm{ext} \cup \widehat{\mathrm{ext}}\, \mathcal{L} \quad \text{but} \quad \tilde{L}_1 \in \mathrm{BC}\, \mathrm{ext}\, \mathcal{L},$$
$$\tilde{L}_2 \notin \lim \cup \widehat{\lim}\, \mathcal{L} \quad \text{but} \quad \tilde{L}_2 \in \mathrm{BC} \lim \mathcal{L}.$$

- $L_a, L_b \in \mathcal{L}(\mathrm{starfree})$

General results: $\mathrm{ext} \cup \widehat{\mathrm{ext}} \subsetneq \mathrm{BC}\,\mathrm{ext}$

Definition 3.12. A language $L \subseteq \Sigma^*$ is called M-**invariant** for $M \subseteq \Sigma$ iff for all $w_1, w_2 \in \Sigma^*$, $a \in M$,

$$w_1 a w_2 \in L \quad \Rightarrow \quad w_1 M^* w_2 \subseteq L.$$

A language $L \subseteq \Sigma^*$ is called M-**relevant** iff L is not M-invariant and $\Sigma^* a \Sigma^* \cap L \neq \emptyset$ for every $a \in M$.

Theorem 3.15. Let $\mathcal{L}$ be closed under negation and under alphabet permutation. Let $\{a, b, c\} \subseteq \Sigma$. Let $L_a \in \mathcal{L}$ be $\{a\}$*-relevant* and $\{b, c\}$*-invariant*. Then

$$\mathrm{ext}\, L_a \notin \widehat{\mathrm{ext}}\, \mathcal{L}^*(\mathrm{reg}) \quad \Rightarrow \quad \mathrm{ext} \cup \widehat{\mathrm{ext}}\, \mathcal{L} \subsetneq \mathrm{BC}\,\mathrm{ext}\, \mathcal{L}$$

and

$$\lim L_a \notin \widehat{\lim}\, \mathcal{L}^*(\mathrm{reg}) \quad \Rightarrow \quad \lim \cup \widehat{\lim}\, \mathcal{L} \subsetneq \mathrm{BC}\lim \mathcal{L}.$$

General results

- **Theorem 3.19.** (Staiger-Wagner 1) $\mathcal{L}$ closed under change of final states. Then
$$\lim \cap \widehat{\lim}\, \mathcal{L} \subseteq \mathrm{BC}\,\mathrm{ext}\, \mathcal{L}.$$
- **Theorem 3.20.** (Staiger-Wagner 2) $\mathcal{L}$ closed under suffix-independence, negation, union and change of final states. Then
$$\mathrm{BC}\,\mathrm{ext}\, \mathcal{L} \subseteq \lim \cap \widehat{\lim}\, \mathcal{L}.$$
- **Theorem 3.22.** $\mathcal{L}$ closed under suffix-independence, negation, union, change of final states and alphabet permutation. Then we have
$$\mathrm{ext} \cap \widehat{\mathrm{ext}}\, \mathcal{L} \overset{(1.)}{\subseteq} \mathrm{ext} \cup \widehat{\mathrm{ext}}\, \mathcal{L} \overset{(2.)}{\subseteq} \mathrm{BC}\,\mathrm{ext}\, \mathcal{L} \overset{(3.)}{=}$$
$$\lim \cap \widehat{\lim}\, \mathcal{L} \overset{(4.)}{\subseteq} \lim \cup \widehat{\lim}\, \mathcal{L} \overset{(5.)}{\subseteq} \mathrm{BC}\lim \mathcal{L}.$$

With $\mathcal{L}$-ext-$\widehat{\mathrm{ext}}$-separating language L_a, the inclusions in (1) and (2) are strict. With $\mathcal{L}$-lim-$\widehat{\lim}$-separating language L'_a, the inclusions in (4) and (5) are strict.

Kleene closure

$$\mathrm{Kleene}'(\mathcal{L}) := \left\{ \bigcup_{i=1}^{n} U_i \cdot V_i^{\omega} \;\middle|\; U_i, V_i \subseteq \Sigma^*, U_i \cdot V_i^* \in \mathcal{L}, n \in \mathbb{N}_0 \right\}$$

- **Lemma 3.24.**
  - Generic power-set construction based on a non-det. UV^* automaton which results in a det. co-Büchi automaton for parts of UV^ω.
  - If the $*$-language given by the automata is in $\mathcal{L}$, we call $\mathcal{L}$ closed under change final states for all deterministic *simplified* automata.
  - With this, we get

$$\mathrm{Kleene}'\,\mathcal{L} \subseteq \mathrm{BC}\lim \mathcal{L}.$$

  - The idea in the proof can probably be generalized into a general constructive non-deterministic Büchi to deterministic Muller automaton conversion.
- **Lemma 3.25.** $\mathcal{L}$ closed under change of final states. Then

$$\lim \mathcal{L} \subseteq \mathrm{Kleene}'\,\mathcal{L}.$$

Congruence based classes $\mathcal{L}(R)$

Motivation: $\mathcal{L}(\mathrm{LT}_n)$ or $\mathcal{L}(\mathrm{PT}_n)$

Let $R \subseteq \Sigma^* \times \Sigma^*$ be a congruence relation.

$$\mathcal{L}^*(R) := \{L \subseteq \Sigma^* \mid L \text{ is finite union of } R\text{-equivalence-classes}\}\,.$$

There is a canonical deterministic automaton with states $S_R := \Sigma^*/R$. We call it the R-automaton.

- Lemma 3.28. $\mathcal{L}(R)$ is *closed under change of final states*.
- Lemma 3.28. $\mathcal{L}(R)$ is *closed under negation, union and intersection*.
- Example 3.29. *Closure under suffix-independence* doesn't directly follow from this.
- Lemma 3.30. $\mathcal{L}^\omega_E(\mathcal{A}_R) = \operatorname{ext} \mathcal{L}(R)$
- Lemma 3.31. $\mathcal{L}^\omega_{\text{Büchi}}(\mathcal{A}_R) = \lim \mathcal{L}(R)$
- Lemma 3.32. $\mathcal{L}^\omega_{\text{Muller}}(\mathcal{A}_R) = \mathrm{BC} \lim \mathcal{L}(R)$

General results: BC $\lim \mathcal{L}(R)$ in $\mathcal{L}^\omega(\mathrm{reg})$

- **Lemma 3.33.** BC $\lim \mathcal{L}(R) \cap \operatorname{ext} \mathcal{L}^*(\mathrm{reg}) \subseteq \operatorname{ext} \mathcal{L}(R)$
  Equality with $\operatorname{ext} \mathcal{L}(R) \subseteq$ BC $\lim \mathcal{L}(R)$.
- **Definition 3.41.** If there is a SCC $Q \subseteq S_R$ including two loops $P_1, P_2 \subseteq Q$, $P_1 \neq P_2$ with $P_1 \not\subseteq P_2$, $P_2 \not\subseteq P_1$, then call $\mathcal{L}(R)$ **postfix-loop-deterministic.**
  Examples:
  - $\mathcal{L}(\mathrm{PT}_n)$ for all n and $\mathcal{L}(\mathrm{LT}_1)$ are not postfix-loop-deterministic
  - $\mathcal{L}(\mathrm{LT}_n)$ for $n \geq 2$ is postfix-loop-deterministic (Lemma 4.14)
- **Theorem 3.44.** $\mathcal{L}(R)$ is not *postfix-loop-deterministic* $\Leftrightarrow$

$$\text{BC} \lim \mathcal{L}(R) \cap \lim \mathcal{L}^*(\mathrm{reg}) = \lim \mathcal{L}(R).$$

General results: $\mathcal{L}(R)$: Staiger-Wagner

- **Example 3.46.** There is $\mathcal{L}(R)$ infinity-postfix-independent and not postfix-loop-deterministic and

$$\operatorname{ext} \mathcal{L}(R) \not\subseteq \lim \mathcal{L}(R).$$

- **Theorem 3.47.** (Staiger-Wagner) $\mathcal{L}(R)$ not postfix-loop-deterministic. $\mathrm{BC}\operatorname{ext} \mathcal{L}(R) \subseteq \mathrm{BC}\lim \mathcal{L}(R)$. Then

$$\lim \cap \widehat{\lim}\, \mathcal{L}(R) = \mathrm{BC}\operatorname{ext} \mathcal{L}(R)$$

Concrete results

For $\mathcal{L} := \mathcal{L}(\mathrm{starfree})$, via Theorem 3.22, we get

$$\mathsf{ext} \cap \widehat{\mathsf{ext}}\,\mathcal{L} \subsetneqq \mathsf{ext} \cup \widehat{\mathsf{ext}}\,\mathcal{L} \subsetneqq \mathsf{BC}\,\mathsf{ext}\,\mathcal{L} =$$
$$\mathsf{lim} \cap \widehat{\mathsf{lim}}\,\mathcal{L} \subsetneqq \mathsf{lim} \cup \widehat{\mathsf{lim}}\,\mathcal{L} \subsetneqq \mathsf{BC}\,\mathsf{lim}\,\mathcal{L}.$$

For $\mathcal{L} := \mathcal{L}(\mathrm{LT})$, via Theorem 3.22, we get

$$\mathsf{ext} \cap \widehat{\mathsf{ext}}\,\mathcal{L} \subsetneqq \mathsf{ext} \cup \widehat{\mathsf{ext}}\,\mathcal{L} \subsetneqq \mathsf{BC}\,\mathsf{ext}\,\mathcal{L} =$$
$$\mathsf{lim} \cap \widehat{\mathsf{lim}}\,\mathcal{L} \subsetneqq \mathsf{lim} \cup \widehat{\mathsf{lim}}\,\mathcal{L} \subsetneqq \mathsf{BC}\,\mathsf{lim}\,\mathcal{L}.$$

For $\mathcal{L} := \mathcal{L}(\mathrm{PT})$, we get

$$\mathsf{ext} \cap \widehat{\mathsf{ext}}\,\mathcal{L} \subsetneqq \mathsf{ext} \cup \widehat{\mathsf{ext}}\,\mathcal{L} \subsetneqq \mathsf{BC}\,\mathsf{ext}\,\mathcal{L} =$$
$$\mathsf{lim} \cap \widehat{\mathsf{lim}}\,\mathcal{L} = \mathsf{lim} \cup \widehat{\mathsf{lim}}\,\mathcal{L} = \mathsf{BC}\,\mathsf{lim}\,\mathcal{L}.$$

Conclusion

- Closure under change of final state or variants of this closure was important in some proofs, e.g. Staiger-Wagner or Kleene closure.
- Another possible generalization: class of $\mathcal{L}$ automata (instead of single fixed R-automata as in $\mathcal{L}(R)$). e.g. $\bigcup_n \mathrm{PT}_n -$ automata.
- More concrete language classes can be studied. Supersets of the class of regular languages weren't studied at all here. Natural generalization would be to use pushdown automata in the proofs for the class of context free languages.