Handout 10

Myhill-Nerode

1 Rappel: Relations d'équivalence et congruences droites

Une relation $\sim$ sur un ensemble U est une relation d'équivalence si

1. $\sim$ est *réflexive*: pour tout $x \in U$: $x \sim x$;
2. $\sim$ est *symétrique*: pour tous $x, y \in U$: si $x \sim y$ alors $y \sim x$;
3. $\sim$ est *transitive*: pour tous $x, y, z \in U$: si $x \sim y$ et $y \sim z$ alors $x \sim z$.

Quand $\sim$ est une relation d'équivalence sur U et $x \in U$ alors la *classe d'équivalence de x* est $[x]_\sim = \{y \in U \mid x \sim y\}$. Deux classes d'équivalence $[x]_\sim$ et $[y]_\sim$ sont soit égales (quand $x \sim y$), soit disjointes (quand $x \not\sim y$). Il s'ensuit que U est *partionné* en classes d'équivalence par rapport à $\sim$, c'est-à-dire la famille de toutes les classes d'équivalence de $\sim$ est une partition de U.

Une famille C de parties de U est une *partition* de U quand

1. tous les éléments de C sont des parties non vides de U: $\emptyset \notin C$;
2. tous les éléments de C sont disjoints: pour tous $P_1, P_2 \in C$: $P_1 \neq P_2 \Rightarrow P_1 \cap P_2 = \emptyset$;
3. les éléments de C couvrent tout U: $\bigcup_{P \in C} P = U$.

Si C est une partition de U alors la relation $\sim_C$ sur U définie par $x \sim_C y$ si et seulement s'il existe un $P \in C$ tel que $x, y \in P$, est une relation d'équivalence.

L'*indice* d'une relation d'équivalence $\sim$ est le nombre de ses classes d'équivalence, ce nombre peut être fini ou infini.

Une relation d'équivalence $\sim$ sur Σ^* est une *congruence droite* si:

$$\forall x, y, z \in \Sigma^* : x \sim y \Rightarrow xz \sim yz$$

2 Équivalence induite par un langage

Soit $L \subseteq \Sigma^*$ un langage. La relation $\sim_L$ sur Σ^* est définie par

$$x \sim_L y \text{ ssi } \forall w \in \Sigma^* : xw \in L \Leftrightarrow yw \in L$$

Propriétés de cette relation:

1. $\sim_L$ est une relation d'équivalence.
2. $\sim_L$ est une congruence droite.
3. $x \sim_L y$ si et seulement si $x^{-1}L = y^{-1}L$.
4. toute classe d'équivalence $[x]_{\sim_L}$ est soit incluse dans L, soit disjointe de L.

3 Équivalence induite par un automate

Soit $A = (\Sigma, Q, q_0, F, \delta)$ un automate déterministe complet. La relation $\sim_A$ sur Σ^* est définie par

$$x \sim_A y \text{ ssi } \delta^*(q_0, x) = \delta^*(q_0, y)$$

Propriétés de cette relation:

- $\sim_A$ est une relation d'équivalence
- $\sim_A$ est une congruence droite
- si L est le langage reconnu par A, alors $\sim_A$ est un *raffinement* de $\sim_L$, c'est-à-dire:

  $$\forall x, y \in \Sigma^* : \text{si } x \sim_A y \text{ alors } x \sim_L y$$

  et on a donc que $indice(\sim_A) \geq indice(\sim_L)$.
- $|Q| \geq indice(\sim_A)$, donc $\sim_A$ est d'indice fini.

4 L'automate induit par une congruence droite $\sim$ d'indice fini

Soit L un langage tel que $\sim_L$ est une congruence droite d'indice fini. Alors l'*automate induit par* $\sim$ est $A_\sim = (\Sigma, Q, q_0, F, \delta)$ défini par:

- $Q = \{[w]_\sim \mid w \in \Sigma^*\}$
- $q_0 = [\epsilon]_\sim$
- $F = \{[w]_\sim \mid w \in L\}$
- $\delta([w]_\sim, a) = [wa]_\sim$

Cet automate est bien défini car $\sim$ est d'indice fini et une congruence droite. On a pour cet automate que $\delta^*(q_0, w) = [w]_\sim$.

5 Le théorème de Myhill-Nerode

Soit $L \subseteq \Sigma^*$ un langage.

1. L est rationnel si et seulement si $\sim_L$ est d'indice fini.
2. Si L est rationnel alors l'indice de $\sim_L$ est égal au nombre d'états du plus petit automate déterministe complet qui reconnaît L.