The Büchi-Elgot-Traktenbrot theorem for regular languages over finite words

Fix Σ - finite alphabet.

$MSOL_\Sigma$:

$$\varphi, \psi ::= Q_a(x) \mid x=y \mid s(x,y) \mid x<y \mid x\in Y$$
$$\varphi \vee \psi \mid \neg\varphi \mid \exists x.\varphi(x) \mid \exists Y \varphi(Y)$$

where $a \in \Sigma$, x,y are First-order vars, Y is a Set Variable

Definition

Let $w \in \Sigma^*$, $|w| = t$
Let $FV = \{x_1, .. x_n\}$ and $SV = \{Y_1, .. Y_m\}$, $V = FV \cup SV$.

Let $I = \langle I_f, I_s \rangle$ be an interpretation of FV and SV on the set of positions $[1,n] = \{1, \dots n\}$:

$$I_f : FV \to [1,n]$$
$$I_s : SV \to 2^{[1,n]}$$

Let $\Sigma_V = \Sigma \times F$ where $F = \{ f \mid f : V \to \{0,1\} \}$

Then the encoding function maps w and $I = \langle I_f, I_s \rangle$ to a word over Σ_V :

$$\cancel{enc(w, I) : (w, I)}$$

$$enc(w, I) = \binom{a_1}{f_1}\binom{a_2}{f_2} \cdots \binom{a_t}{f_t}$$

where $w = a_1 \dots a_n$ and
$f_j(x_i) = 1$ iff $I_f(x_i) = 1$
$f_j(Y_i) = 1$ iff $I_s(Y_i) = 1$ $\quad\square$

~~The Büchi-Elgot-Traktenbrot theorem for regular languages over finite words.~~ ①

<u>Definition</u> Let $\mathrm{Valid}_V \subseteq \Sigma_V^*$ be the set

$$\mathrm{Valid}_V = \left\{ \begin{pmatrix} a_1 \\ f_1 \end{pmatrix}\begin{pmatrix} a_2 \\ f_2 \end{pmatrix}\cdots\begin{pmatrix} a_t \\ f_t \end{pmatrix} \;\middle|\; \begin{array}{l} \text{for each } x_i \in FV, \text{ there is a} \\ \text{unique } j \text{ such that } f_j(x_i) = 1 \end{array} \right\}$$

<u>Prop</u> : enc is a 1-1 correspondence between the sets

$$\{ (w, I) \mid I \text{ is an interpretation of } V \text{ on } w \}$$

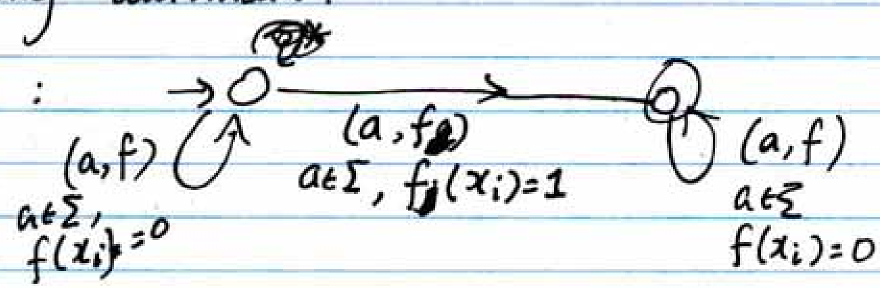
and

$$\mathrm{Valid}_V .$$

<u>Prop</u>. Valid_V is regular

<u>Proof</u>. An automaton can check if for each x_i, there is one and only one f_j such that $f_j(x_i) = 1$. Can be constructed by intersecting the following automata:

$\mathcal{A}_i$:



□

<u>Definition</u>. For any $\varphi(x_1, \ldots x_n, Y_1, \ldots Y_m)$ in MSO_Σ let $\mathrm{Lang}_\varphi = \{ \mathrm{enc}(w, I_V) \mid w \models_I \varphi \}$ where $V = \{x_1, \ldots x_n, Y_1, \ldots Y_m\}$.

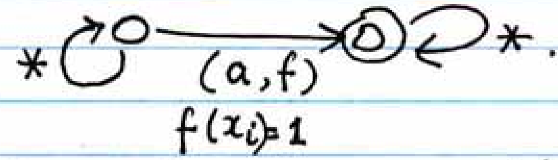
(Note that $\mathrm{Lang}_\varphi \subseteq \Sigma_V^*$)

<u>Lemma</u> For every $\varphi(x_1, .. x_n, Y_1, .. Y_m)$ in MSO_Σ, $Lang_\varphi$ is regular

<u>Proof</u> : Induction on structure of formula φ.

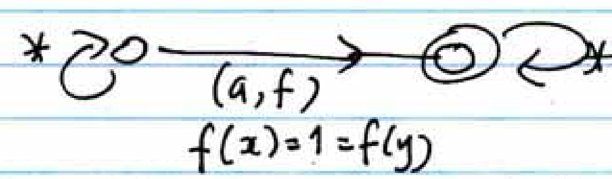
<u>Base cases</u>

- $Q_a(x_i)$: Automaton can check if whenever $f_j(x_i) = 1$, $a_i = a$:



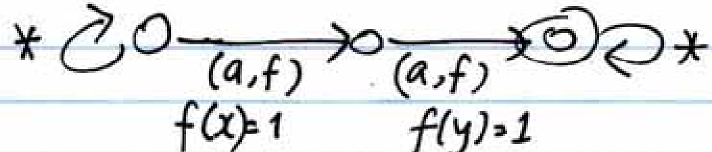
Take intersection with $Valid_V$.

- $x = y$: Automaton can check that whenever $f_j(x) = 1$, $f_j(y) = 1$ as well.



Take intersection with $Valid_V$.

- $s(x,y)$: Automaton can check whether there are consecutive positions j and $j+1$ such that $f_j(x) = 1$ and $f_{j+1}(y) = 1$.

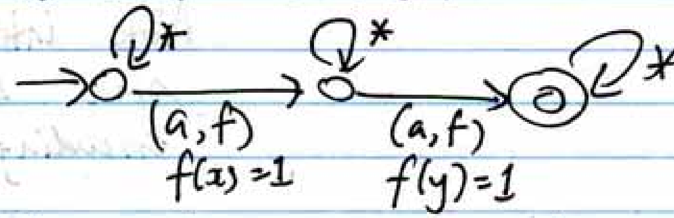


Take intersection with $Valid_V$.

- $x < y$

(Note that we can skip this as MSOL can express $<$ using successor relation S)

Automaton will check:



Take intersection with V.

Induction step.

- $\varphi \vee \psi$

Assume Lang$_\varphi$ and Lang$_\psi$ have been shown to be regular.

Also assume they are over the same variables
[If not, change them so they are.

For example, if Lang$_\varphi$ is not over valuations for x, change ~~and~~ automaton A_φ accepting Lang$_\varphi$ ~~so~~ by replacing every edge

$$q \xrightarrow{a,f} q'$$

by

$$q \xrightarrow{a,f'} q'$$

such that
$f'(x_i) = f(x_i)$
for each $x_i \neq x$.].

Then $\text{Lang}_{\varphi \vee \psi} = \text{Lang}_\varphi \cup \text{Lang}_\psi$

Hence regular.

∘ $\neg\varphi$ $\quad$ $Lang(\neg\varphi) = \overline{Lang(\varphi)} \cap Valid_v$

Hence $Lang(\neg\varphi)$ is regular.

Note intersection with $Valid_v$ is neccesary since $\overline{Lang(\varphi)}$ may have invalid encodings.

∘ $\exists x. \varphi(x)$ $\quad$ Note that $Lang(\exists x. \varphi(x))$ contains precisely the words $\binom{a_1}{f_1} \cdots \binom{a_t}{f_t}$ ~~Lang~~ such that there is some word $\binom{a_1}{g_1} \cdots \binom{a_t}{g_t}$ in $Lang(\varphi)$ ~~such that~~ such that

g_j is same as f_j except that g_j gives a mapping for x as well.

Hence we can take A_φ, the automaton accepting $Lang(\varphi)$, and replace every edge

$$q \circ \xrightarrow[a,g]{} \circ q'$$

by

$$q \circ \xrightarrow[a,f]{} \circ q'$$

where $f = g\restriction_{\text{free variables of } \exists x.\varphi(x)}$, ~~is.~~

i.e. f is same as g except the valuation for x is "thrown away".

~~$\exists X\ \varphi(X)$~~

- $\exists Y.\ \varphi(Y)$ — Same automaton operation as for $\exists x.\varphi(x)$.

$\square$

<u>Corollary</u>. Let φ be an MSO_Σ sentence.

Then $\{ w \in \Sigma^* \mid w \models \varphi \}$ is regular

<u>Proof</u>. Take automaton A_φ accepting $Lang_\varphi$.

A_φ is over (a, f) where f is an empty function. Replace $q \xrightarrow[a,f]{} q'$ by $q \xrightarrow[a]{} q'$

Resulting automaton accepts required language.

<u>Theorem</u> Let $L \subseteq \Sigma^*$

L is regular iff there is an MSO formula φ s.t.

$\{ w \in \Sigma^* \mid w \models \varphi \} = L$.

<u>Corollary</u> MSO_Σ has a decidable satisfiability problem:

<u>Proof</u>. Given $\varphi \in MSO_\Sigma$,

construct A_φ s.t.

$$\mathcal{L}(A_\varphi) = \{ w \in \Sigma^* \mid w \models \varphi \}$$

Then φ is satisfiable iff $\mathcal{L}(A_\varphi) \neq \emptyset$.

Check emptiness of A_φ. $\square$