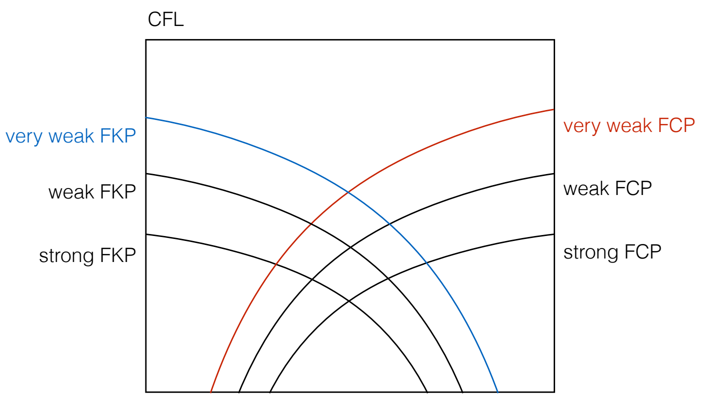
The Strong, Weak, and Very Weak Finite Context and Kernel Properties

Makoto Kanazawa
National Institute of Informatics and SOKENDAI

Ryo Yoshinaka
Tohoku University



I'll spend most of my time filling you in on the background.
This talk is about six subclasses of the context-free languages, two of which turn out to be identical.
The classification has to do with work on distributional learning of context-free languages.
I'll first introduce the strongest of the two types of properties of CFGs.

Strings and Contexts

$L \subseteq \Sigma^*$: fixed language

The "acceptance" relation $R_L \subseteq \Sigma^* \times (\Sigma^* \times \Sigma^*)$ between strings and contexts defined by

$$x\ R_L\ (u,v) \Leftrightarrow uxv \in L$$

induces a **Galois connection**

$$(\mathcal{P}(\Sigma^*), \subseteq) \longleftrightarrow (\mathcal{P}(\Sigma^* \times \Sigma^*), \supseteq)$$

$$X \subseteq \Sigma^* \mapsto X^{\triangleright} = \{ (u,v) \mid uXv \subseteq L \}$$

$$W^{\triangleleft} = \{ x \mid uxv \in L \text{ for all } (u,v) \in W \} \leftarrow W \subseteq \Sigma^* \times \Sigma^*$$

All of these properties are relative to a fixed language L, which is supposed to be the target of learning.
Pairs of strings are called "contexts".
R_L: the relation of a context "accepting" a string.

$(\cdot)^{\triangleright\triangleleft}$ is a closure operator.

- $X \subseteq X^{\triangleright\triangleleft}$
- $X_1 \subseteq X_2 \Rightarrow X_1^{\triangleright\triangleleft} \subseteq X_2^{\triangleright\triangleleft}$
- $X^{\triangleright\triangleleft\triangleright\triangleleft} = X^{\triangleright\triangleleft}$

$W^{\triangleleft}$ is a closed set (i.e., $W^{\triangleleft} = W^{\triangleleft\triangleright\triangleleft}$) for all $W \subseteq \Sigma^* \times \Sigma^*$.

L is a closed set.

$\because L = \{(\varepsilon,\varepsilon)\}^{\triangleleft}$

$(XY)^{\triangleright\triangleleft} = (X^{\triangleright\triangleleft}Y^{\triangleright\triangleleft})^{\triangleright\triangleleft}$

CFG

$G = (N, \Sigma, P, I)$

I: set of initial nonterminals

$L(G,A) = \{ x \in \Sigma^* \mid A \Rightarrow_G^* x \}$

$L(G) = \bigcup_{A \in I} L(G,A)$

I use a definition of CFGs slightly different from the standard one in that multiple initial nonterminals are allowed.

Strong FCP/FKP

A CFG $G = (N, \Sigma, P, I)$ has

- the **strong finite context property** if for each $A \in N$, there is some finite $C_A \subseteq \Sigma^* \times \Sigma^*$ such that

$$L(G,A) = C_A^{\triangleleft},$$

- the **strong finite kernel property** if for each $A \in N$, there is some finite $K_A \subseteq \Sigma^*$ such that

$$L(G,A) = K_A^{\triangleright\triangleleft}.$$

$\triangleright$, $\triangleleft$: relative to L(G)

The Dyck Language

$$S \to \varepsilon \mid aSbS$$

$$L(G,S) = D_1^* = \{(\varepsilon,\varepsilon)\}^{\triangleleft} = \{\varepsilon\}^{\triangleright\triangleleft}$$

I'll go over a few examples.
The standard grammar for the (one-sided) Dyck language satisfies both the strong FCP and the strong FKP.

Regular Sets

$M = (Q, \Sigma, \delta, q_I, F)$ DFA for L

$G = (Q, \Sigma, P, I)$

$$P = \{\, q \to ar \mid \delta(q,a) = r \,\} \cup \{\, q \to \varepsilon \mid q \in F \,\}$$

$$I = \{\, q_I \,\}$$

$$\hat{\delta}(q_I,u) = r \Rightarrow L(G,r) = \{\, x \mid \hat{\delta}(r,x) \in F \,\} = \{(u,\varepsilon)\}^{\triangleleft}$$

The right-linear grammar for a regular set corresponding to a DFA has the strong FCP.

Regular Sets

$\Sigma^*/{\equiv_L}$: syntactic monoid of L

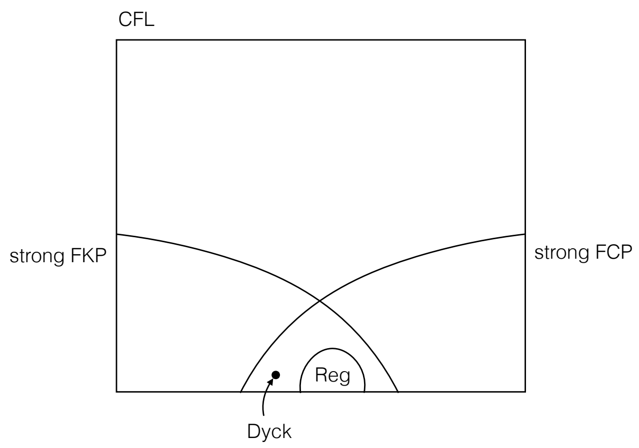
$G = (\Sigma^*/{\equiv_L}, \Sigma, P, I)$

$$P = \{\, [uv]_{\equiv_L} \to [u]_{\equiv_L}\, [v]_{\equiv_L} \mid u,v \in \Sigma^* \,\} \cup \{\, [a]_{\equiv_L} \to a \mid a \in \Sigma \cup \{\varepsilon\} \,\} \cup \{\, [u]_{\equiv_L} \to [v]_{\equiv_L} \mid [u]_{\equiv_L} \supseteq [v]_{\equiv_L} \,\}$$

$$I = \{\, [x]_{\equiv_L} \mid x \in L \,\}$$

$$L(G,[x]_{\equiv_L}) = [x]_{\equiv_L}^{\triangleright\triangleleft} = \{x\}^{\triangleright\triangleleft}$$

The CFG corresponding to the syntactic monoid of a regular set has the strong FKP.



$L = \{ w \in \{a,b\}^* \mid |w|_a \neq |w|_b \}$

Applying the pumping lemma to a long string in a^+, we get:

$S \Rightarrow^* a^i\ A\ a^j$,

$A \Rightarrow^+ a^k\ A\ a^l$ with $k+l > 0$,

$A \Rightarrow^* a^m$.

$$\{ a^{m+n(k+l)} \mid n \geq 0 \} \subseteq L(G,A).$$

$$a^{i+nk}\ L(G,A)\ a^{j+nl} \subseteq L \text{ for all } n \geq 0,$$

$$\{ b^{i+j+n(k+l)} \mid n \geq 0 \} \cap L(G,A) = \varnothing.$$

An example that satisfies neither the strong FCP nor the strong FKP.
L(G,A) includes an infinite subset of a^+, and there is an infinite set of contexts consisting of strings in a^+ that accept all strings in L(G,A).
The complement of L(G,A) includes an infinite subset of b^+.

$L = \{ w \in \{a,b\}^* \mid |w|_a \neq |w|_b \}$

- For any finite $K \subseteq \{a,b\}^*$, $K^{\triangleright\triangleleft} \cap a^*$ is finite.

  Let $m = \max \{ |w|_a - |w|_b \mid w \in K \}$.

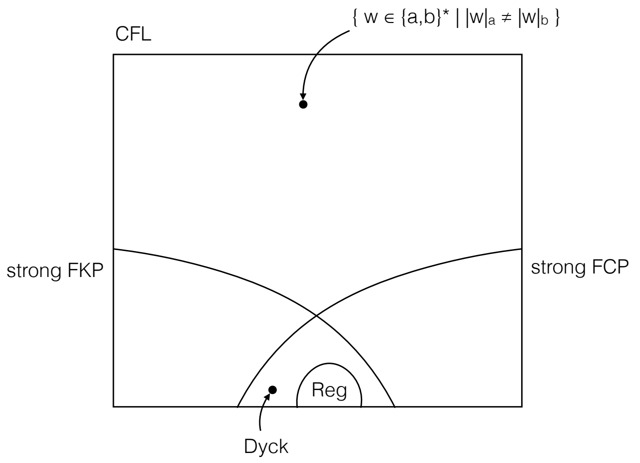
  $$K^{\triangleright} \supseteq \{ (\varepsilon, b^n) \mid n > m \},$$

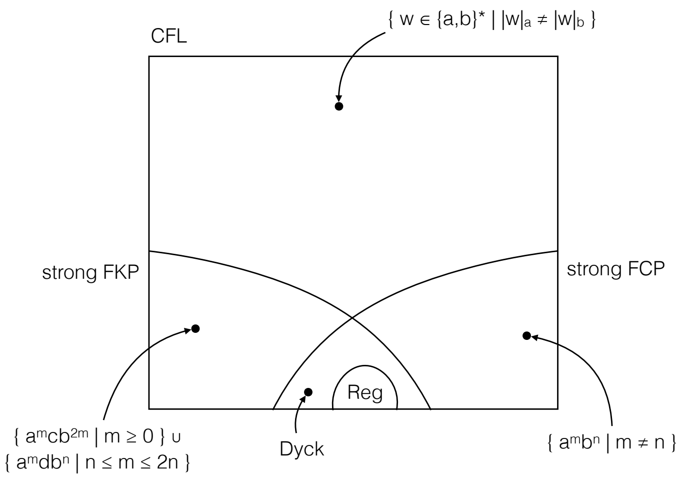
  $$K^{\triangleright\triangleleft} \cap a^* \subseteq \{ a^n \mid n \leq m \}.$$

  So $\{ a^{m+n(k+l)} \mid n \geq 0 \} \subseteq L(G,A) \neq K^{\triangleright\triangleleft}$.

- For any finite $C \subseteq \{a,b\}^* \times \{a,b\}^*$, $C^{\triangleleft} \cap b^*$ is co-finite.

  Since $\{ b^{i+j+n(k+l)} \mid n \geq 0 \} \cap L(G,A) = \varnothing$, $L(G,A) \neq C^{\triangleleft}$.





It's easy to see that the other two regions are also nonempty.

Learning from Positive Data and Membership Queries

positive presentation of L

$t_1, t_2, t_3, \ldots$ $\qquad$ $L = \{ t_n \mid n \geq 1 \}$

At stage n:

- positive data: $D_n = \{t_1,\ldots,t_n\}$
- membership queries

  $w \in L?$

  with $|w|$ bounded by a polynomial in the size of D_n

polynomial update time

The strong FCP and the strong FKP were introduced in the context of Alex Clark's work on "distributional learning" of context-free languages.

Dual Learner

- Assume the target language has a CFG with the strong FCP.
- Nonterminals: $C \subseteq \Sigma^* \times \Sigma^*$ with $1 \leq |C| \leq \mathbf{k}$.
- A production $C_0 \to w_0\, C_1\, w_1 \dots C_n\, w_n$ ($n \leq \mathbf{r}$, $w_j \in \Sigma^*$) is **valid** if

$$C_0^{\triangleleft} \supseteq w_0\, C_1^{\triangleleft}\, w_1 \dots C_n^{\triangleleft}\, w_n.$$

- This is approximated by

$$C_0^{\triangleleft} \supseteq w_0\, (E \cap C_1^{\triangleleft})\, w_1 \dots (E \cap C_n^{\triangleleft})\, w_n$$

where E is the set of substrings in the available positive data.

There are two learning algorithms, dual and primal.

Dual Learner

Parameters: positive integers **k**, **r**;
Data: positive presentation $t_1, t_2, t_3, \dots$ of L; membership oracle for L;

$D_0 := \varnothing$; $E_0 := \varnothing$; $J_0 := \varnothing$;
for i := 1, 2, 3,… **do**
  $D_i := D_{i-1} \cup \{t_i\}$; $E_i := \mathrm{Sub}(D_i)$;
  if $D_i \not\subseteq L(G_{i-1})$ **then**
    $J_i := \mathrm{Con}(D_i)$;
  else
    $J_i := J_{i-1}$;
  $N_i := \{ C \subseteq J_i \mid 1 \leq |C| \leq \mathbf{k} \}$;
  $P_i := \{ C_0 \to w_0\, C_1\, w_1 \dots C_n\, w_n \mid n \leq \mathbf{r}, C_j \in N_i, w_j \in E_i\ (j = 0,\dots,n),$
    $C_0^{\triangleleft} \supseteq w_0\, (E_i \cap C_1^{\triangleleft})\, w_1 \dots (E_i \cap C_n^{\triangleleft})\, w_n \}$;
  $I_i := \{ C \in N_i \mid E_i \cap C^{\triangleleft} \subseteq L \}$;
  output $G_i := (N_i, \Sigma, P_i, I_i)$;

$\mathrm{Sub}(D) = \{ x \mid uxv \in D \}$

$\mathrm{Con}(D) = \{ (u,v) \mid uxv \in D \}$

Why the Dual Learner Works

- If $A_0 \to w_0\, A_1\, w_1 \dots A_n\, w_n$ is a production of G and $L(G,A_j) = C_j^{\triangleleft}$, then the production $C_0 \to w_0\, C_1\, w_1 \dots C_n\, w_n$ is valid:

$$C_0^{\triangleleft} \supseteq w_0\, C_1^{\triangleleft}\, w_1 \dots C_n^{\triangleleft}\, w_n$$

- If A is an initial nonterminal of G and $L(G,A) = C^{\triangleleft}$, then $C^{\triangleleft} \subseteq L$.
- If every production of G_i is valid and every initial nonterminal C of G_i satisfies $C^{\triangleleft} \subseteq L$, then $L(G_i) \subseteq L$.
- If L = L(G) for some G with the strong FCP, the dual learner converges to some G′ such that L = L(G′).

Primal Learner

- Assume the target language has a CFG with the strong FKP.
- Nonterminals: $K \subseteq \Sigma^*$ with $1 \leq |K| \leq$ **k**.
- A production $K_0 \to w_0\, K_1\, w_1 \ldots K_n\, w_n$ ($n \leq$ **r**, $w_j \in \Sigma^*$) is **valid** if

$$K_0^{\rhd\lhd} \supseteq w_0\, K_1^{\rhd\lhd}\, w_1 \ldots K_n^{\rhd\lhd}\, w_n,$$

or, equivalently, if

$$K_0^{\rhd} \subseteq (w_0\, K_1\, w_1 \ldots K_n\, w_n)^{\rhd}.$$

- This is approximated by

$$J \cap K_0^{\rhd} \subseteq (w_0\, K_1\, w_1 \ldots K_n\, w_n)^{\rhd}$$

where J is the set of contexts contained in the available positive data.

Primal Learner

Parameters: positive integers **k**, **r**;
Data: positive presentation $t_1, t_2, t_3, \ldots$ of L; membership oracle for L;

$D_0 := \varnothing$; $E_0 := \varnothing$; $J_0 := \varnothing$;
for i := 1, 2, 3,… **do**
  $D_i := D_{i-1} \cup \{t_i\}$; $J_i := \mathrm{Con}(D_i)$;
  if $D_i \not\subseteq L(G_{i-1})$ **then**
    $E_i := \mathrm{Sub}(D_i)$;
  else
    $E_i := E_{i-1}$;
  $N_i := \{ K \subseteq E_i \mid 1 \leq |K| \leq \mathbf{k} \}$;
  $P_i := \{ K_0 \to w_0\, K_1\, w_1 \ldots K_n\, w_n \mid n \leq \mathbf{r},\ K_j \in N_i,\ w_j \in E_i\ (j = 0,\ldots,n),$
      $J_i \cap K_0^{\rhd} \subseteq (w_0\, K_1\, w_1 \ldots K_n\, w_n)^{\rhd} \}$;
  $I_i := \{ K \in N_i \mid K \subseteq L \}$;
  output $G_i := (N_i, \Sigma, P_i, I_i)$;

$\mathrm{Sub}(D) = \{ x \mid uxv \in D \}$

$\mathrm{Con}(D) = \{ (u,v) \mid uxv \in D \}$

Why the Primal Learner Works

- If $A_0 \to w_0\, A_1\, w_1 \ldots A_n\, w_n$ is a production of G and $L(G,A_j) = K_j^{\rhd\lhd}$, then the production $K_0 \to w_0\, K_1\, w_1 \ldots K_n\, w_n$ is valid:

$$K_0^{\rhd\lhd} \supseteq w_0\, K_1^{\rhd\lhd}\, w_1 \ldots K_n^{\rhd\lhd}\, w_n$$

- If A is an initial nonterminal of G and $L(G,A) = K^{\rhd\lhd}$, then $K \subseteq L$.
- If every production of G_i is valid and every initial nonterminal K of G_i satisfies $K \subseteq L$, then $L(G_i) \subseteq L$.
- If $L = L(G)$ for some G with the strong FKP, the primal learner converges to some G′ such that $L = L(G')$.

Weak FCP/FKP

A CFG $G = (N, \Sigma, P, I)$ has

- the **weak finite context property** if for each $A \in N$, there is some finite $C_A \subseteq \Sigma^* \times \Sigma^*$ such that

$$L(G,A)^{\triangleright\triangleleft} = C_A^{\triangleleft},$$

- the **weak finite kernel property** if for each $A \in N$, there is some finite $K_A \subseteq \Sigma^*$ such that

$$L(G,A)^{\triangleright\triangleleft} = K_A^{\triangleright\triangleleft}.$$

$\triangleright$, $\triangleleft$: relative to $L(G)$

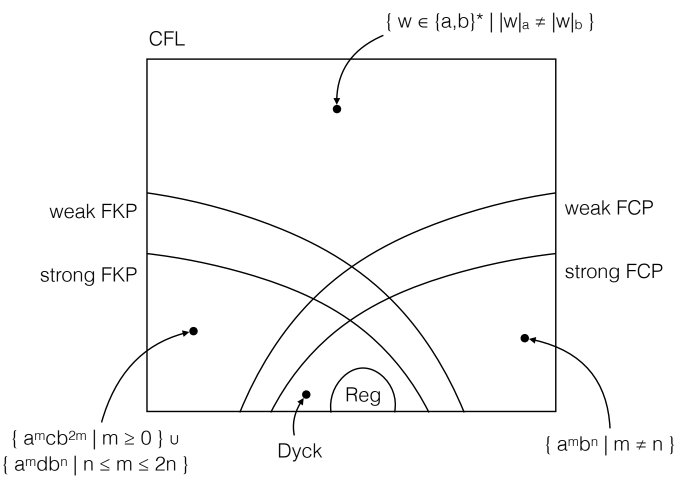
The strong FCP/FKP were soon found to be stronger than necessary.

Weak FCP/FKP Is Sufficient

- $X_0 \supseteq w_0 X_1 w_1 \dots X_n w_n$ implies

$$X_0^{\triangleright\triangleleft} \supseteq w_0 X_1^{\triangleright\triangleleft} w_1 \dots X_n^{\triangleright\triangleleft} w_n.$$

- $X \subseteq L$ implies $X^{\triangleright\triangleleft} \subseteq L$.
- The dual learner converges to a correct grammar if L has a grammar with the weak FCP.
- The primal learner converges to a correct grammar if L has a grammar with the weak FKP.



Strong and Weak FCP/FKP, Dual and Primal Learners

Ryo Yoshinaka Alex Clark

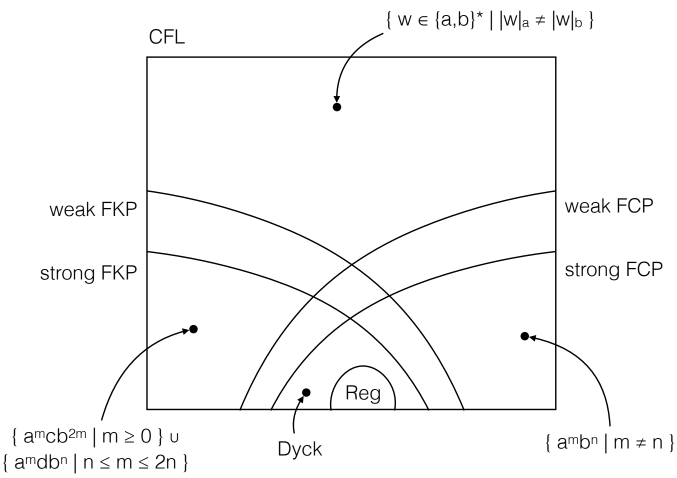
25

I’ve covered the background to this work, due to these people.

Yoshinaka’s Question

- Is there a context-free language that has a grammar satisfying the weak FCP/FKP but has no grammar satisfying the strong FCP/FKP?

26



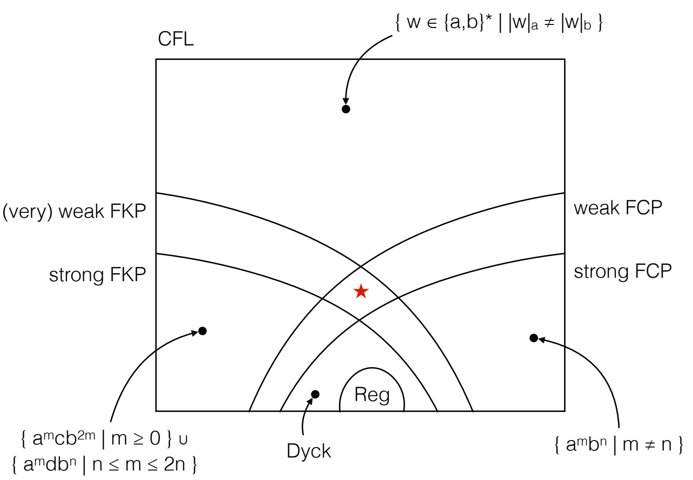
27

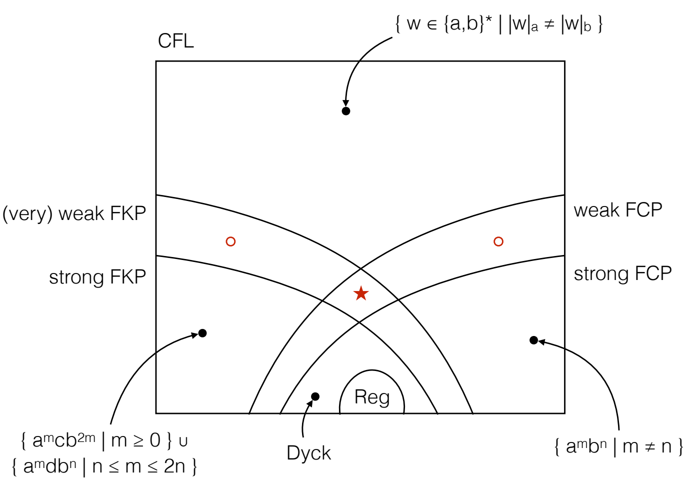
$$L = L_1 \cup L_2 \cup L_3,$$
$$L_1 = \{\, w_1 \# w_2 \# \ldots \# w_n \$ w_n^R \ldots w_2^R w_1^R \mid n \geq 1, w_1, \ldots, w_n \in \{a,b\}^* \,\},$$
$$L_2 = \{\, w y c^i d^i e^j z \mid w, z \in \{a,b\}^*, y \in (\#\{a,b\}^*)^*, i, j \geq 0, |w|_a \geq |w|_b \,\},$$
$$L_3 = \{\, w y c^i d^j e^j z \mid w, z \in \{a,b\}^*, y \in (\#\{a,b\}^*)^*, i, j \geq 0, |w|_a \leq |w|_b \,\}.$$

- L has a grammar satisfying both the weak FCP and the weak FKP.
- L has no grammar satisfying the strong FCP or the strong FKP.
- Every CFG G for L has a nonterminal A such that L(G,A) is not a closed set.

$$L(G,A)^{\triangleright\triangleleft} \cap c^*d^*e^* = \{\, c^i d^i e^i \mid i \geq 0 \,\}$$

If $c^i d^j e^k$ is accepted in all contexts that accept v\$z, then i = j = k.





The nonemptiness of two regions sometimes implies the nonemptiness of another region.
The observation that the weak FCP/ weak FKP is sufficient was meaningful.

More Fundamental Questions

- Does the weak FCP/FKP exactly capture the languages that the dual/primal learner successfully learns?
- Does the grammar that the dual/primal learner converge to always satisfy the weak FCP/FKP?

Are the weak FCP and the weak FKP weak enough?
Turns out it's easy to define properties that exactly capture the dual/primal learners.

Pre-fixed Points

A sequence $(X_A)_{A\in N}$ of subsets of Σ^* is a **pre-fixed point** of $G = (N, \Sigma, P, I)$ if for each production $A_0 \to w_0 A_1 w_1 \dots A_n w_n$ in P,

$$X_{A_0} \supseteq w_0 X_{A_1} w_1 \dots X_{A_n} w_n.$$

- $(L(G,A))_{A\in N}$ is the least pre-fixed point of G.
- $(L(G,A)^{\triangleright\triangleleft})_{A\in N}$ is the least pre-fixed point of G consisting of closed sets.

Very Weak FCP

A CFG $G = (N, \Sigma, P, I)$ has the **very weak finite context property** if there is a sequence $(C_A)_{A\in N}$ of finite subsets of $\Sigma^* \times \Sigma^*$ such that

- $(C_A^{\triangleleft})_{A\in N}$ is a pre-fixed point of G,
- $\bigcup_{A\in I} C_A^{\triangleleft} \subseteq L(G)$.

strong FCP	$L(G,A) = C_A^{\lhd}$	$\bigcup_{A\in I} L(G,A) \subseteq L(G)$
weak FCP	$L(G,A)^{\rhd\lhd} = C_A^{\lhd}$	$\bigcup_{A\in I} L(G,A)^{\rhd\lhd} \subseteq L(G)$
very weak FCP	$(C_A^{\lhd})_{A\in N}$ is a pre-fixed point of G	$\bigcup_{A\in I} C_A^{\lhd} \subseteq L(G)$

Very Weak FCP Captures Dual Learner

- The dual learner converges to a correct grammar if L has a grammar with the **very** weak FCP.
- The grammar that the dual learner converges to always satisfies the **very** weak FCP.

Very Weak FKP

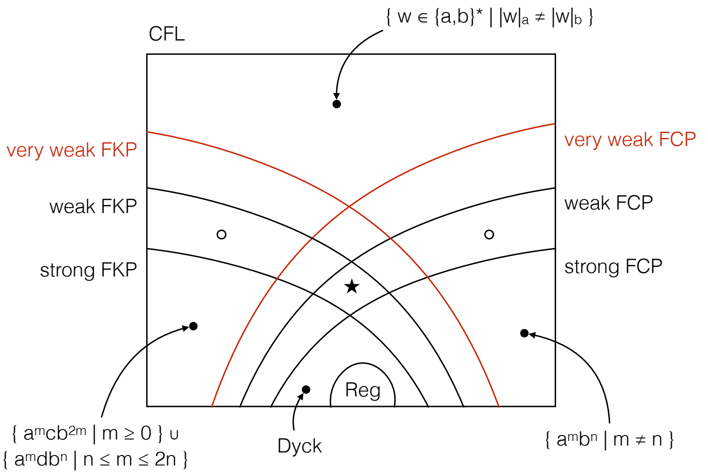
A CFG $G = (N, \Sigma, P, I)$ has the **very weak finite kernel property** if there is a sequence $(K_A)_{A\in N}$ of finite subsets of Σ^* such that

- $(K_A^{\rhd\lhd})_{A\in N}$ is a pre-fixed point of G,
- $\bigcup_{A\in I} K_A^{\rhd\lhd} \subseteq L(G)$ (or, equivalently, $\bigcup_{A\in I} K_A \subseteq L(G)$).

strong FKP	$L(G,A) = K_A^{\triangleright\triangleleft}$	$\bigcup_{A\in I} L(G,A) \subseteq L(G)$
weak FCP	$L(G,A)^{\triangleright\triangleleft} = K_A^{\triangleright\triangleleft}$	$\bigcup_{A\in I} L(G,A)^{\triangleright\triangleleft} \subseteq L(G)$
very weak FCP	$(K_A^{\triangleright\triangleleft})_{A\in N}$ is a pre-fixed point of G	$\bigcup_{A\in I} K_A^{\triangleright\triangleleft} \subseteq L(G)$

Very Weak FKP Captures Primal Learner

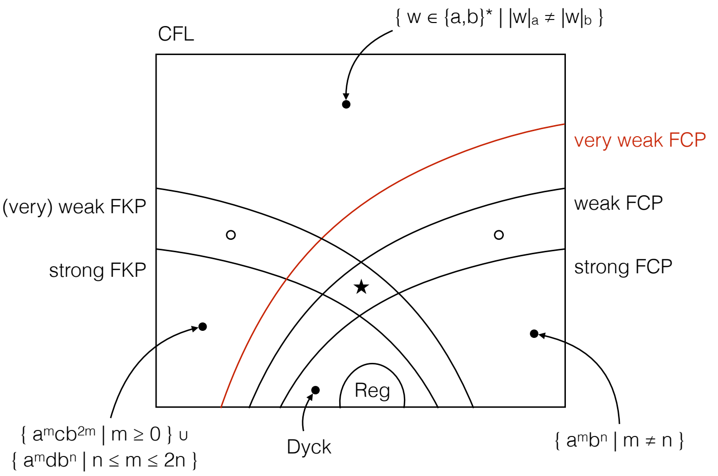
- The primal learner converges to a correct grammar if L has a grammar with the **very** weak FKP.
- The grammar that the primal learner converges to always satisfies the **very** weak FKP.

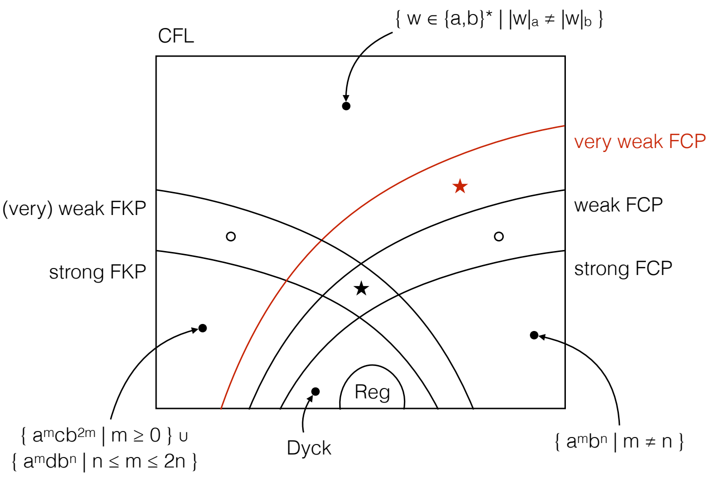


Everything inside one of the two red boundaries is polynomial-time learnable from positive data and membership queries.
Is this a real improvement?

Very Weak FKP ≡ Weak FKP

- L has a grammar with the very weak FKP if and only if L has a grammar with the weak FKP.
- The language that the primal learner converges to always satisfies the weak FKP.





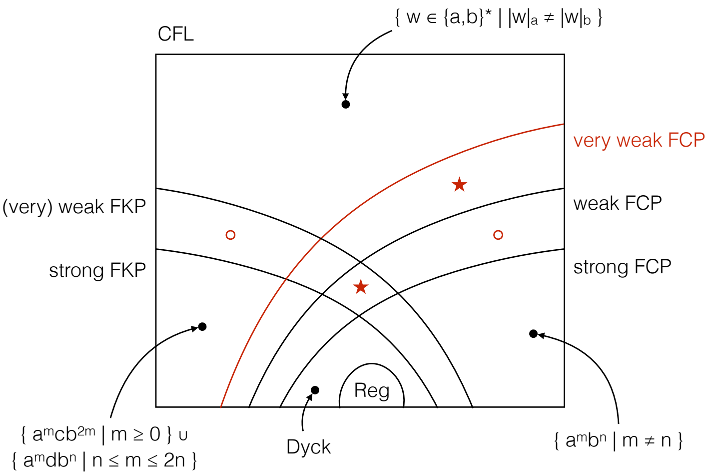
Main result of the paper.

$$L' = L_1 \cup L_2 \cup L_3 \cup L_4 \cup L_5,$$

$$L_1 = \{\, w_1 \# w_2 \# \dots \# w_n \$ w_n^R \dots w_2^R w_1^R \mid n \geq 1, w_1, \dots, w_n \in \{a,b\}^* \,\},$$
$$L_2 = \{\, w y c^i d^i e^j z \mid w, z \in \{a,b\}^*, y \in (\#\{a,b\}^*)^*, i, j \geq 0, |w|_a \geq |w|_b \,\},$$
$$L_3 = \{\, w y c^i d^j e^j z \mid w, z \in \{a,b\}^*, y \in (\#\{a,b\}^*)^*, i, j \geq 0, |w|_a \leq |w|_b \,\}.$$
$$L_4 = \{\, v \$ z c^k \% f^l \mid v \in \{a,b,\#\}^*, z \in \{a,b\}^*, k, l \geq 0 \,\},$$
$$L_5 = \{\, v c^i d^i e^j z c^k \% f^l \mid v \in \{a,b,\#\}^*, z \in \{a,b\}^*, i, j, k, l \geq 0, j \neq k \,\},$$

- L′ has a CFG satisfying the very weak FCP, but has no CFG satisfying the weak FCP.

This example involves a phenomenon similar to what we saw with { w ∈ {a,b}* | $|w|_a \neq |w|_b$ }.



New results in red.

Conclusion

- The primal learner is exactly captured by the weak FKP.
- The dual learner is exactly captured by the very weak FCP, which is genuinely weaker than the weak FCP.