Theorem (Ginsburg & Spanier '64, '66)
A set $S \subseteq \mathbb{N}^n$ is Presburger-definable iff it is semi-linear.

Proof:

"$\Leftarrow$" Consider a linear set $L(c, \{v_1, \dots, v_n\})$.
Then
$$v \in L(c, \{v_1, \dots, v_n\})$$
iff
$$\exists x_1, \dots, x_n : \; v = c + x_1 v_1 + \dots + x_n v_n.$$

"$\Rightarrow$" Consider a Presburger formula φ.
By quantifier elimination, there is an equivalent formula ψ
- that only contains atomic formulas
$$t_1 = t_2 \; , \; t_1 < t_2 \; , \; t_1 \equiv_m t_2$$
- and combines them with $\vee$ and $\wedge$.

Show that the resulting solution spaces are semi-linear:

IA: $Sol(\varphi)$ with $\varphi = t_1 = t_2 \mid t_1 < t_2 \mid t_1 \equiv_m t_2$ can be shown to be semi-linear.

IS: Assume that $Sol(\varphi)$ and $Sol(\psi)$ are semi-linear.
Then
- $Sol(\varphi \wedge \psi) = \underbrace{\underbrace{Sol(\varphi)}_{\text{semi-linear by induction hypothesis}} \cap \underbrace{Sol(\psi)}_{}}_{\text{semi-linear by closure of semi-linear sets under } \cap}$ is semi-linear.
- $Sol(\varphi \vee \psi) = Sol(\varphi) \cup Sol(\psi)$ is semi-linear by induction hypothesis and closure under $\cup$. $\square$

Corollary (Semi-linear sets are closed under complement)
If $S \subseteq \mathbb{N}^n$ is semi-linear, so is $\overline{S}$.

Picture:

Presburger-definable

$\shortparallel$ (Ginsburg & Spanier)

semi-linear

(homework) $\shortparallel$ $\supseteq$ (Parikh (now))

$\Psi(REG) \subseteq \Psi(CF)$

(by definition)

Parikh's Theorem

Show that $\Psi(\mathcal{L}(G))$ is semi-linear for every CFG G.

Recall:

- A context-free grammar (CFG) is a tuple $G = (N, \Sigma, P, S)$ where
  - N is a finite set of non-terminals
  - Σ is a finite set of terminals (with $N \cap \Sigma = \emptyset$)
  - $P \subseteq N \times (N \cup \Sigma)^*$ is a finite set of production rules
  - $S \in N$ is a start symbol.

Consider rules of the form $A \to B.C$ $A \to a$
(Chomsky normal form).

- Derivation relation:

$$\alpha_1 . A . \alpha_2 \to \alpha_1 . \gamma . \alpha_2 \text{ if } A \to \gamma \text{ in } P.$$

$$(\alpha_1, \alpha_2, \gamma \in (N \cup \Sigma)^*).$$

Language of a grammar

$$\mathcal{L}(G) := \{ w \in \Sigma^* \mid S \to^* w \}.$$

- Example:

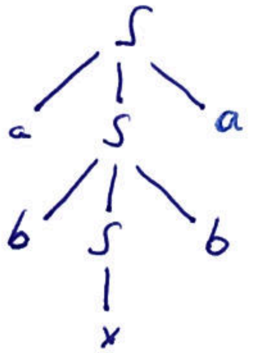
Let $G = (\{S\}, \{a, b, x\}, P, S)$
with P:

$$S \to aSa \mid bSb \mid x.$$

Grammar G generates palindromes:

$$S \to aSa \to abSba \to abxba.$$

- Derivations represented by parse trees:

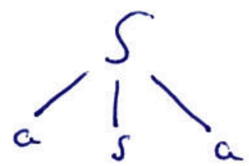
$s =$ 

$root(s) = S$
$yield(s) = abxba$
$depth(s) = 3$
$N(s) = \{S\}.$

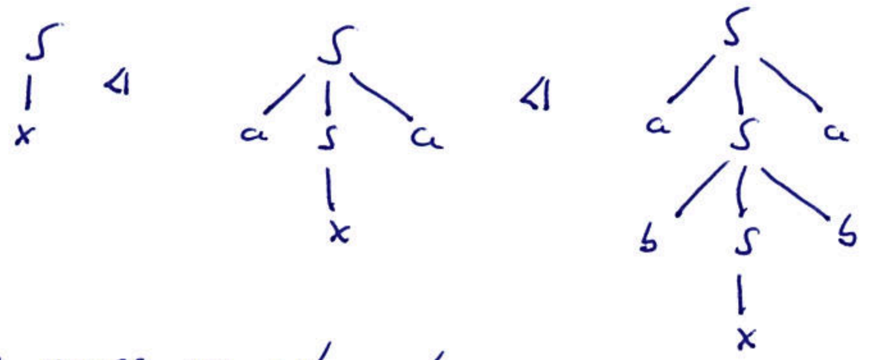
Let s be a parse tree.
Define
$root(s) :=$ non-terminal at the root.
$yield(s) :=$ string of terminals and non-terminals at the leaves (read from left to right)
$depth(s) :=$ length of longest path from leaf to root
$N(s) :=$ set of non-terminals in s.

Definition (Pump, $\triangleleft$, basic pump)

- A pump is a parse-tree with at least two nodes so that
$yield(s) = x.\,root(s).\,y$ with $x, y \in \Sigma^*$



- Define $s \triangleleft t$ on parse trees if
t can be obtained from s by inserting a pump.



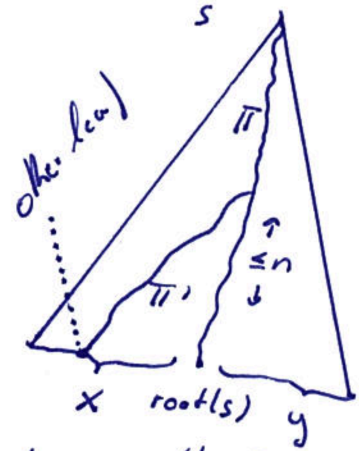
Since pumps are not empty,
there is no infinite descending chain $s_0 \triangleright s_1 \triangleright s_2 \triangleright \cdots$
- A pump is basic if it is $\triangleright$-minimal
(alternatively: t is basic if $s \triangleleft t$ implies $s = root(t)$)

Lemma 1:

If s is a basic pump, then $depth(s) \leq 2n$ with $n = |N|$.

Proof:

Consider unique path from leaf with label root(s) to root of s.



↳ Cannot be longer than n, otherwise a non-terminal repeats.

↳ If such a non-terminal repeats, have a pump that can be cut-out.

↯ minimality of s wrt. $\triangleright$.

Similarly from another leaf to first node on path π

↳ path π' cannot be longer than $n+1$.

Hence, the length from any leaf to root is at most $2n$. □

So the number of basic pumps is finite, say $p \in \mathbb{N}$.

Lemma 2:

Every parse tree t is either $\triangleright$-minimal or contains a basic pump.

Proof:

If t is not minimal, it contains pumps.

Let s be a $\triangleright$-minimal pump in t.

If s is not basic, it contains another pump u.

But this smaller pump would also be in t.

↯ contradiction to minimality of s.

So s is basic. □

Define $s \leq t$ if t is obtained from s by

- finitely many insertions of
- basic pumps u with $N(u) \subseteq N(s)$.

Let $\alpha \in (N \cup \Sigma)^*$.

$\Psi(\alpha) := \Psi(x)$ where $x = \alpha$ without non-terminals.

$\Psi(t) := \Psi(\text{yield}(t))$.

Lemma 3:

The set $\{\Psi(t) \mid s \leq t\}$ is linear.

Proof:

$$\{\Psi(t) \mid s \leq t\} = \mathcal{L}(\Psi(s), \{\Psi(u) \mid u \text{ is a basic pump with } N(u) \subseteq N(s)\}).$$

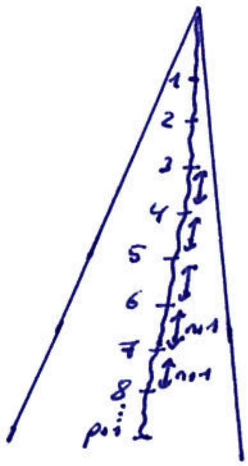
□

Lemma 4:

number of basic pumps — number of non-terminals

If s is $\leq$-minimal and $\text{yield}(s) \in \Sigma^*$, then $\text{depth}(s) \leq (p+1)(n+1)$

Proof:



↳ Assume s has a path longer than $(p+1)(n+1)$.
Then the path can be decomposed into $(p+1)$ paths of length $(n+1)$.

↳ This means each segment repeats a non-terminal.
Hence, there are $(p+1)$ pumps.
They are disjoint (share at most leaf/root).

↳ With Lemma 2, each pump is either basic or contains a basic pump.
But there are only p different basic pumps.
So one repeats.

↯ minimality of s.
One of the repeating pumps can be deleted without changing the set of non-terminals.

□

Theorem (Parikh '66)

Consider a CFG G. Then $\Psi(\mathcal{L}(G))$ is semi-linear.

Proof:

Define $M := \{s \mid s \text{ is } \leq\text{-minimal}, \text{root}(s) = S, \text{yield}(s) \in \Sigma^*\}$

Show that

$$\Psi(\mathcal{L}(G)) = \bigcup_{s \in M} \{\Psi(t) \mid s \leq t\}.$$

By Lemma 4, this union is finite.
By Lemma 3, the sets $\{\Psi(t) \mid s \le t\}$ are linear.
So the union is a semi-linear set.

<u>Proof of =</u>:

„$\supseteq$" Any t such that $s \le t$ for some $s \in M$ has

$$\mathrm{root}(t) = S \text{ and } \mathrm{yield}(t) \in \Sigma^*.$$

Thus $\mathrm{yield}(t) \in L(G)$ and hence

$$\underset{\substack{\| \\ \Psi(\mathrm{yield}(t))}}{\Psi(t)} \in \Psi(L(G)).$$

„$\subseteq$" Any word $x \in L(G)$ has a parse tree t with

$$\mathrm{root}(t) = S \text{ and } \mathrm{yield}(t) = x.$$

Hence, there is a $\le$-minimal s with $s \le t$.
Then

$$s \in M \text{ and } \Psi(x) \in \{\Psi(t) \mid s \le t\}.$$

$\square$