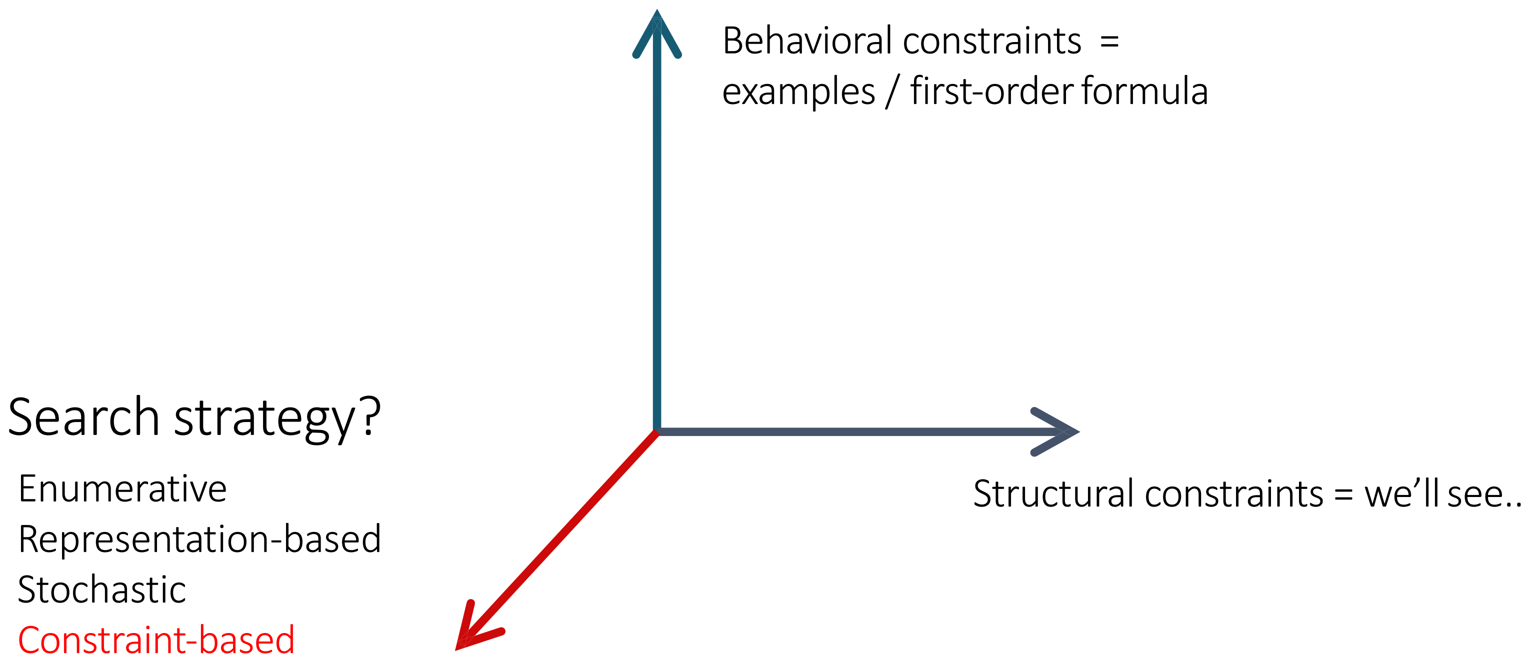
Lecture 8
Constraint-based search

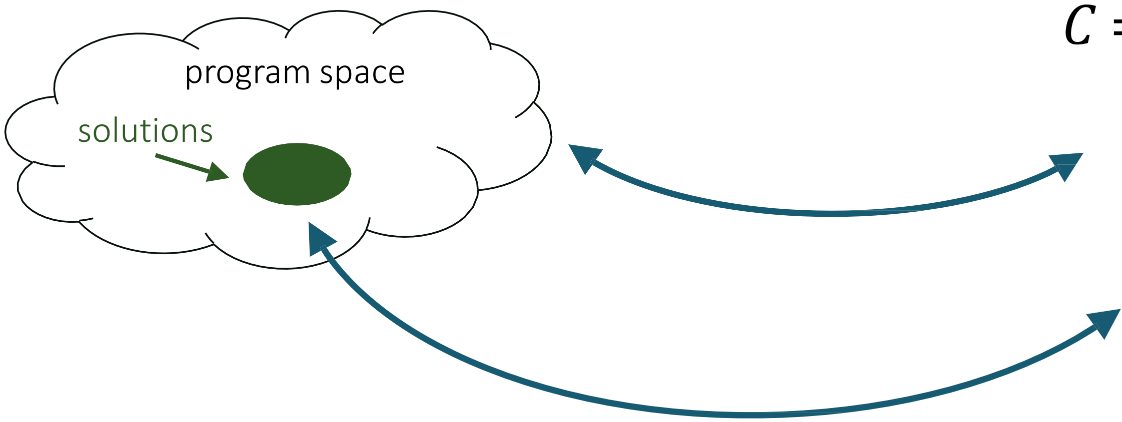
The problem statement



Constraint-based search

Idea: encode the synthesis problem as a SAT/SMT problem and let a solver deal with it

What is an encoding?



parameter space

$$C = \{c_1, \ldots, c_n\}$$

$$\{C \mid \mathrm{wf}(C)\}$$

well-formedness constraint

$$\{C \mid \mathrm{wf}(C) \wedge \mathrm{sem}(C)\}$$

semantic correctness constraint

$$\mathrm{sem}(C) = \bigwedge_{\langle i,o\rangle \in E} \varphi(C, i, o)$$

C denotes a program that gives output *o* on input *i*

SAT encoding: example

program space

`{x, x & 1}`

`x` is a two-bit word
(`x = x`$_h$`x`$_l$)

`E = [11 → 01]`

parameter space

$$C = \{c : \mathrm{Bool}\}$$

`decode[0] → x`
`decode[1] → x & 1`

$$\mathrm{wf}(c) \equiv \top$$

$$\varphi(c, i_h, i_l, o_h, o_l) \equiv (\neg c \Rightarrow o_h = i_h \wedge o_l = i_l) \wedge (c \Rightarrow o_h = 0 \wedge o_l = i_l)$$

$$\mathrm{SAT}(\varphi(c, 1, 1, 0, 1))$$

$$\mathrm{SAT}((\neg c \Rightarrow 0 = 1 \wedge 1 = 1) \wedge (c \Rightarrow 0 = 0 \wedge 1 = 1)) \xrightarrow{\text{SAT solver}} \texttt{Model } \{c \to 1\}$$

`return` `decode[1]` i.e. `x & 1`

How to define an encoding

Define the parameter space $C = \{c_1, \ldots, c_n\}$

- `encode : Prog → C`
- `decode : C → Prog` (might not be defined for all C)

Define a formula $\mathrm{wf}(c_1, \ldots, c_n)$

- that holds iff `decode[C]` is a "well-formed" program

Define a formula $\varphi(c_1, \ldots, c_n, i, o)$

- that holds iff `(decode[C])(i) = o`

Constraint-based search

```
constraint-based (wf, φ, E = [i → o]) {
  match SAT(wf(C) ∧ ⋀⟨i,o⟩∈E φ(C, i, o)) with
    Unsat -> return "No solution"
    Model C* -> return decode[C*]
}
```

Find a satisfying assignment for $c_1, \ldots, c_n$
(**i** and **o** are fixed)

SMT encoding: example

program space

`x + N | x * N`

$\text{wf}(c_{op}, c_N) \equiv \top$

`N` is an in integer literal
`x` is an integer input

`E = [2 → 9]`

parameter space

$C = \{c_{op}\text{: Bool}, c_N\text{: Int}\}$

```
decode[0,N] → x + N
decode[1,N] → x * N
```

$$\varphi(c_{op}, c_N, i, o) \equiv (\neg c_{op} \Rightarrow o = i + c_N) \wedge (c_{op} \Rightarrow o = i * c_N)$$

$$\text{SAT}(\varphi(c_{op}, c_N, 2, 9))$$

$$\text{SAT}((\neg c_{op} \Rightarrow 9 = 2 + c_N) \wedge (c_{op} \Rightarrow 9 = 2 * c_N))$$

SMT solver

`Model {`c_{op}`→0,` c_N`→ 7}`

`return` `decode[0,7]` i.e. `x + 7`

What is a good encoding?

Sound

- if $\mathrm{wf}(C) \land \mathrm{sem}(C)$ then `decode[C]` is a solution

Complete

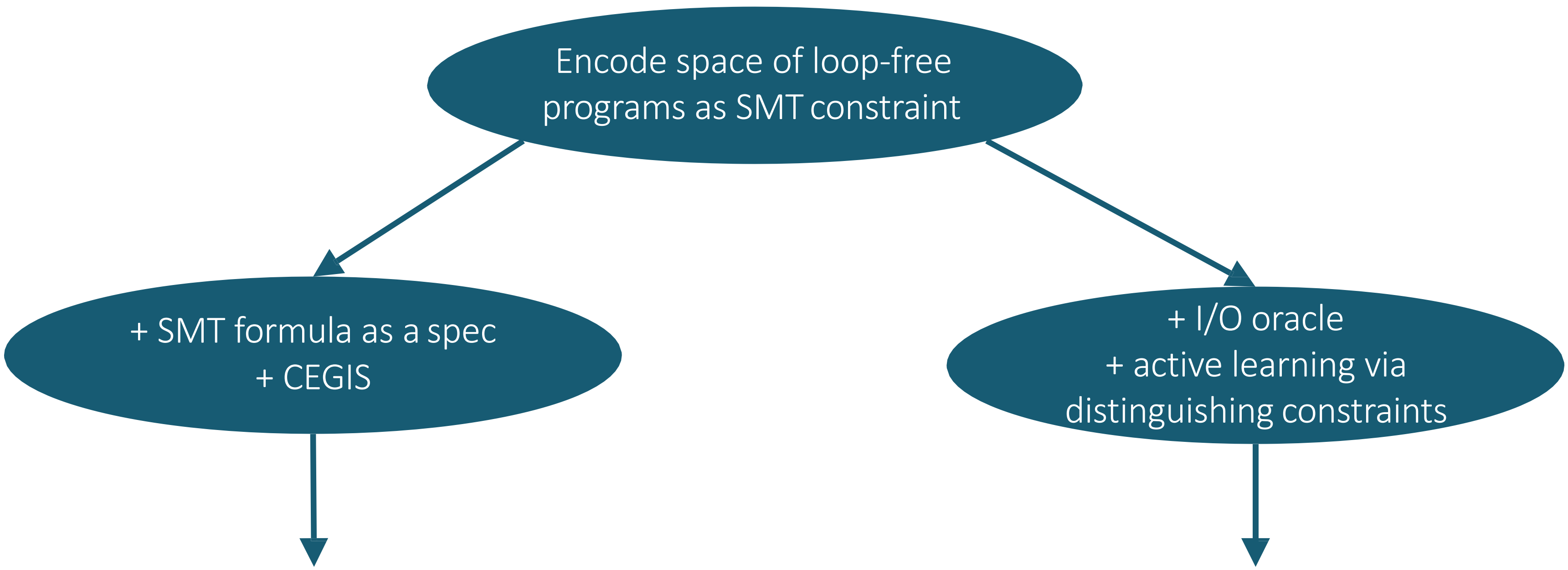
- if `decode[C]` is a solution then $\mathrm{wf}(C) \land \mathrm{sem}(C)$

Small parameter space

Solver-friendly

- decidable logic, compact constraint

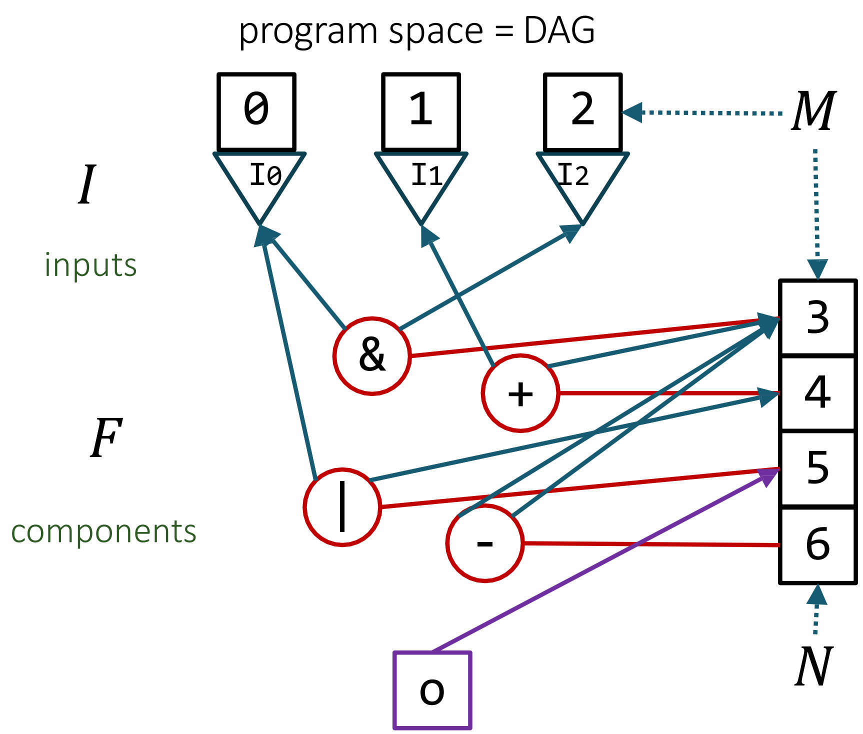
Brahma



[Gulwani, Jha, Tiwari, Venkatesan. *Synthesis of loop-free programs* PLDI'11]

[Jha, Gulwani, Seshia, Tiwari. *Oracle-guided component-based program synthesis*. ICSE'10]

Brahma encoding:



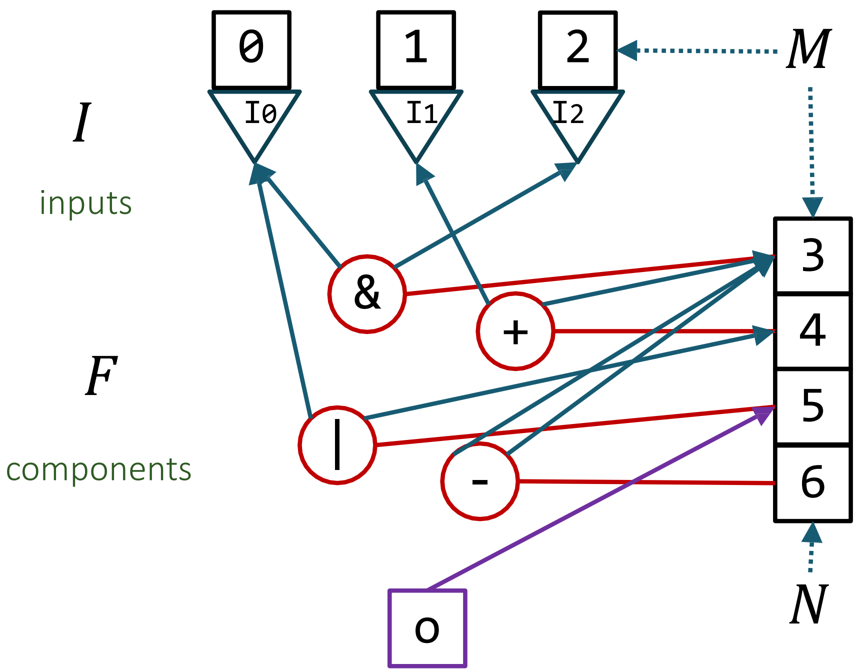
parameter space

$$C = \{c_O, \bigcup_{i\in I} c_{I_i} : Int\} \cup \bigcup_{f\in F} \{c_{O_f}, c_{I_f}, c_{J_f} : Int\}$$

$$wf(C) \equiv c_o \in M \wedge \bigwedge_{I_i \in I} c_{I_i} = I_i \wedge \bigwedge_{f\in F} c_{O_f} \in N \wedge c_{I_f, J_f} \in M$$

Brahma encoding:

program space = DAG



parameter space

$$C = \{c_O, \bigcup_{i \in I} c_{I_i} : Int\} \cup \bigcup_{f \in F} \{c_{O_f}, c_{I_f}, c_{J_f} : Int\}$$

$$P = \bigcup_{f \in F} \{I_f, J_f\} \qquad R = \bigcup_{f \in F} \{O_f\}$$

$$\varphi(C, I, O) \equiv \exists P, R. \bigwedge_{f \in F} O_f = F(I_f, J_f)$$

$$\wedge \bigwedge_{x \in P \cup R \cup I \cup \{O\}} c_x = c_y \Rightarrow x = y$$

Brahma: contributions

SMT encoding of program space

- sound?
- complete?
- solver-friendly?

SMT solver can guess constants

- e.g. 0x55555555 in P23

Brahma: limitations

Requires component multiplicities

- What happens if user provides too many? too few?
- How would you extend this approach to work without multiplicities?

Requires *precise* SMT specs for components

- What happens if we give an over-approximate spec?

Brahma: limitations

No ranking

Cannot handle:

- loops
- types
- noise
    - Can we add these things?

Brahma: questions

Behavioral Constraints? Structural Constraints? Search Strategy?

- First-order formula
- A multiset of components + straight-line program
- Constraint based + CEGIS

Can we represent these structural constraints as a grammar?

- Yes and no
- No because grammars cannot encode multiplicities
- Yes because the set is finite, so we can simply enumerate all possible programs
  - but this is not useful for synthesis

Comparison of search strategies