Outline

Regular algebraic program analysis

Semantic foundations of algebraic program analysis

Interprocedural analysis

ω-regular program analysis

Motivation

1. What it would mean to apply algebraic program analyis *beyond* the framework of algebraic path properties?
   - Set of paths of interest may not be regular (recursive procedures)
   - Paths of interest may not be finite (termination)
2. What does it mean for an algebraic program analysis to be correct?
   - How do we prove it?
3. How can we reason about the impact of program transformation on analysis?

General picture for algebraic program analysis

- Suppose we have a system of recursive $E = \{X_i = R_i\}_{i=1}^n$ defining the semantics of a program
  - Some *concrete* interpretation $\mathscr{I}^\natural = \langle A^\natural, f^\natural \rangle$
  - Interested in *least solution* $\sigma^\natural : X \to A^\natural$ to E over $\mathscr{I}^\natural$: $\sigma(X_i) = \mathscr{I}[\![R_i[\sigma^\natural]]\!]$ for all i

General picture for algebraic program analysis

- Suppose we have a system of recursive $E = \{X_i = R_i\}_{i=1}^n$ defining the semantics of a program
    - Some *concrete* interpretation $\mathscr{I}^\natural = \langle A^\natural, f^\natural \rangle$
    - Interested in *least solution* $\sigma^\natural : X \to A^\natural$ to E over $\mathscr{I}^\natural$: $\sigma(X_i) = \mathscr{I}[\![R_i[\sigma^\natural]]\!]$ for all i
- Want to *approximate* this semantics [Cousot & Cousot '77]
    - Some *abstract* interpretation $\mathscr{I}^\sharp = \langle A^\sharp, f^\sharp \rangle$
    - Some *approximation relation* $\Vdash \subseteq A^\natural \times A^\sharp$
        - $p^\natural \Vdash p^\sharp$: "$p^\natural$ is approximated by $p^\sharp$"
    - Want: $\sigma^\sharp : \{X_i\}_{i=1}^n \to A^\sharp$ s.t. $\sigma^\natural(X_i) \Vdash \sigma^\sharp(X_i)$ for all i

General picture for algebraic program analysis

- Suppose we have a system of recursive $E = \{X_i = R_i\}_{i=1}^n$ defining the semantics of a program
  - Some *concrete* interpretation $\mathscr{I}^\natural = \langle A^\natural, f^\natural \rangle$
  - Interested in *least solution* $\sigma^\natural : X \to A^\natural$ to E over $\mathscr{I}^\natural$: $\sigma(X_i) = \mathscr{I}[\![R_i[\sigma^\natural]]\!]$ for all i
- Want to *approximate* this semantics [Cousot & Cousot '77]
  - Some *abstract* interpretation $\mathscr{I}^\sharp = \langle A^\sharp, f^\sharp \rangle$
  - Some *approximation relation* $\Vdash \subseteq A^\natural \times A^\sharp$
    - $p^\natural \Vdash p^\sharp$: "$p^\natural$ is approximated by $p^\sharp$"
  - Want: $\sigma^\sharp : \{X_i\}_{i=1}^n \to A^\sharp$ s.t. $\sigma^\natural(X_i) \Vdash \sigma^\sharp(X_i)$ for all i
- The algebraic method:
  1. Symbolically compute a *closed-form* solution to the system, $E' = \{X_i = R'_i\}_{i=1}^n$
     - Right-hand-sides R'_i do not contain variables
     - E and E' have same least solution over $\mathscr{I}^\natural$
  2. Interpret the closed forms over $\mathscr{I}^\sharp$
     - $\sigma^\sharp(X_i) \triangleq \mathscr{I}^\sharp[\![R'_i]\!]$

Relational semantics

- Let P be program, given by a control flow graph $G = (V, E)$ with entry r
  - Program configurations: $V \times$ State (where, say, State $\triangleq \mathbb{Z}^X$)
  - Program transition relation: $\rightarrow_P \subseteq (V \times \text{State}) \times (V \times \text{State})$
- Relational semantics: For each vertex v,

$$R_v \triangleq \{\langle s, s' \rangle \in \text{State} \times \text{State} : \langle r, s \rangle \rightarrow_P^* \langle v, s' \rangle\}$$

Program can reach $\langle v, s' \rangle$ from initial state $\langle r, s \rangle$

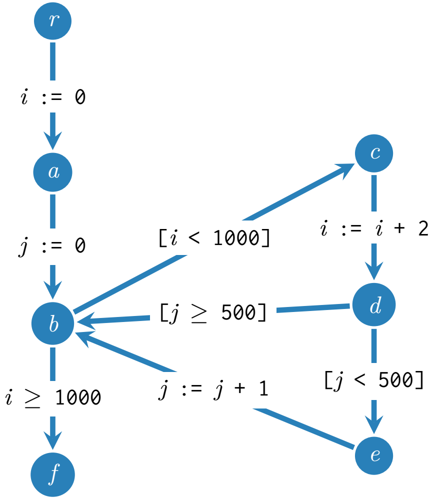
Relational interpretation

Universe: binary relations over states

$0 \triangleq \emptyset$	Empty relation
$1 \triangleq \{\langle s, s\rangle : s \in \mathbb{Z}^X\}$	Identity relation
$R \cdot S \triangleq \{(s, s'') : \exists s'.(s, s') \in R \wedge (s', s'') \in S\}$	Relational composition
$R + S \triangleq R \cup S$	Union
$R^* \triangleq \bigcup_{i=0}^{\infty} \underbrace{R \circ \cdots \circ R}_{i \text{ times}}$	Reflexive transitive closure

Equational formulation of relational semantics

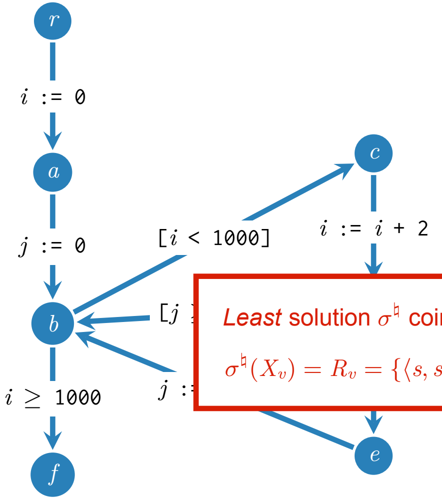
Control flow graph corresponds to **left-linear** system of equations equations



$$
\begin{aligned}
X_r &= 1 \\
X_a &= X_r \cdot \langle r, a \rangle \\
X_b &= X_a \cdot \langle a, b \rangle + X_d \cdot \langle d, b \rangle + X_e \cdot \langle e, b \rangle \\
X_c &= X_b \cdot \langle b, c \rangle \\
X_d &= X_c \cdot \langle c, d \rangle \\
X_e &= X_d \cdot \langle d, e \rangle \\
X_f &= X_b \cdot \langle b, f \rangle
\end{aligned}
$$

Equational formulation of relational semantics

Control flow graph corresponds to **left-linear** system of equations equations



$$X_r = 1$$
$$X_a = X_r \cdot \langle r, a\rangle$$
$$X_b = X_a \cdot \langle a, b\rangle + X_d \cdot \langle d, b\rangle + X_e \cdot \langle e, b\rangle$$
$$X_c = X_b \cdot \langle b, c\rangle$$
$$X_d = X_c \cdot \langle c, d\rangle$$

Least solution $\sigma^{\natural}$ coincides with relational semantics.

$$\sigma^{\natural}(X_v) = R_v = \{\langle s, s'\rangle \in \mathsf{State} \times \mathsf{State} : \langle r, s\rangle \to_P^* \langle v, s'\rangle\}$$

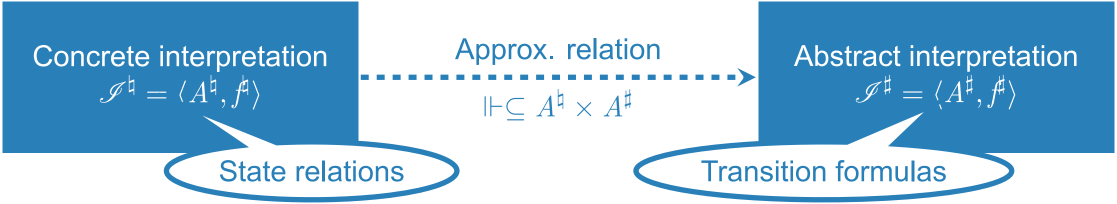
Abstract interpretation

Concrete interpretation
$\mathscr{I}^\natural = \langle A^\natural, f^\natural \rangle$

Approx. relation
$\Vdash \subseteq A^\natural \times A^\sharp$

Abstract interpretation
$\mathscr{I}^\sharp = \langle A^\sharp, f^\sharp \rangle$

Abstract interpretation



Abstract interpretation

Concrete interpretation
$\mathscr{I}^\natural = \langle A^\natural, f^\natural \rangle$

State relations

Approx. relation
$\Vdash \subseteq A^\natural \times A^\sharp$

Abstract interpretation
$\mathscr{I}^\sharp = \langle A^\sharp, f^\sharp \rangle$

Transition formulas

($R \Vdash F \iff$ every $\langle s, s' \rangle \in R$ is a model of F

Computing closed form solutions

Variable elimination $\sim$ Gauss-Jordan

$$
\begin{aligned}
X_r &= 1 \\
X_a &= X_r \cdot \langle r, a \rangle \\
X_b &= X_a \cdot \langle a, b \rangle + X_d \cdot \langle d, b \rangle + X_e \cdot \langle e, b \rangle \\
X_c &= X_b \cdot \langle b, c \rangle \\
X_d &= X_c \cdot \langle c, d \rangle \\
X_e &= X_d \cdot \langle d, e \rangle \\
X_f &= X_b \cdot \langle b, f \rangle
\end{aligned}
$$

Computing closed form solutions

Variable elimination $\sim$ Gauss-Jordan

$$
\begin{aligned}
X_r &= 1 \\
X_a &= \langle r, a \rangle \\
X_b &= X_a \cdot \langle a, b \rangle + X_d \cdot \langle d, b \rangle + X_e \cdot \langle e, b \rangle \\
X_c &= X_b \cdot \langle b, c \rangle \\
X_d &= X_c \cdot \langle c, d \rangle \\
X_e &= X_d \cdot \langle d, e \rangle \\
X_f &= X_b \cdot \langle b, f \rangle
\end{aligned}
$$

Computing closed form solutions

Variable elimination $\sim$ Gauss-Jordan

$$
\begin{aligned}
X_r &= 1 \\
X_a &= \langle r, a \rangle \\
X_b &= \langle r, a \rangle \cdot \langle a, b \rangle + X_d \cdot \langle d, b \rangle + X_e \cdot \langle e, b \rangle \\
X_c &= X_b \cdot \langle b, c \rangle \\
X_d &= X_c \cdot \langle c, d \rangle \\
X_e &= X_d \cdot \langle d, e \rangle \\
X_f &= X_b \cdot \langle b, f \rangle
\end{aligned}
$$

Computing closed form solutions

Variable elimination $\sim$ Gauss-Jordan

$$
\begin{aligned}
X_r &= 1 \\
X_a &= \langle r, a \rangle \\
X_b &= \langle r, a \rangle \cdot \langle a, b \rangle + X_d \cdot \langle d, b \rangle + X_e \cdot \langle e, b \rangle \\
X_c &= X_b \cdot \langle b, c \rangle \\
X_d &= X_b \cdot \langle b, c \rangle \cdot \langle c, d \rangle \\
X_e &= X_d \cdot \langle d, e \rangle \\
X_f &= X_b \cdot \langle b, f \rangle
\end{aligned}
$$

Computing closed form solutions

Variable elimination $\sim$ Gauss-Jordan

$$
\begin{aligned}
X_r &= 1 \\
X_a &= \langle r, a \rangle \\
X_b &= \langle r, a \rangle \cdot \langle a, b \rangle + X_b \cdot \langle b, c \rangle \cdot \langle c, d \rangle \cdot \langle d, b \rangle + X_e \cdot \langle e, b \rangle \\
X_c &= X_b \cdot \langle b, c \rangle \\
X_d &= X_b \cdot \langle b, c \rangle \cdot \langle c, d \rangle \\
X_e &= X_b \cdot \langle b, c \rangle \cdot \langle c, d \rangle \cdot \langle d, e \rangle \\
X_f &= X_b \cdot \langle b, f \rangle
\end{aligned}
$$

Computing closed form solutions

Variable elimination $\sim$ Gauss-Jordan

$$
\begin{aligned}
X_r &= 1 \\
X_a &= \langle r, a\rangle \\
X_b &= \langle r, a\rangle \cdot \langle a, b\rangle + X_b \cdot \langle b, c\rangle \cdot \langle c, d\rangle \cdot \langle d, b\rangle + X_b \cdot \langle b, c\rangle \cdot \langle c, d\rangle \cdot \langle d, e\rangle \cdot \langle e, b\rangle \\
X_c &= X_b \cdot \langle b, c\rangle \\
X_d &= X_b \cdot \langle b, c\rangle \cdot \langle c, d\rangle \\
X_e &= X_b \cdot \langle b, c\rangle \cdot \langle c, d\rangle \cdot \langle d, e\rangle \\
X_f &= X_b \cdot \langle b, f\rangle
\end{aligned}
$$

Computing closed form solutions

Variable elimination $\sim$ Gauss-Jordan

$$
\begin{aligned}
X_r &= 1 \\
X_a &= \langle r, a\rangle \\
X_b &= \langle r, a\rangle \cdot \langle a, b\rangle + X_b \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle)) \\
X_c &= X_b \cdot \langle b, c\rangle \\
X_d &= X_b \cdot \langle b, c\rangle \cdot \langle c, d\rangle \\
X_e &= X_b \cdot \langle b, c\rangle \cdot \langle c, d\rangle \cdot \langle d, e\rangle \\
X_f &= X_b \cdot \langle b, f\rangle
\end{aligned}
$$

Computing closed form solutions

Variable elimination $\sim$ Gauss-Jordan

$$
\begin{aligned}
X_r &= 1 \\
X_a &= \langle r, a\rangle \\
X_b &= \langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^* \\
X_c &= X_b \cdot \langle b, c\rangle \\
X_d &= X_b \cdot \langle b, \\
X_e &= X_b \cdot \langle b, \\
X_f &= X_b \cdot \langle b, f\rangle
\end{aligned}
$$

AB^* is least solution to $X = A + XB$

Computing closed form solutions

Variable elimination $\sim$ Gauss-Jordan

$$
\begin{aligned}
X_r &= 1 \\
X_a &= \langle r, a\rangle \\
X_b &= \langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^* \\
X_c &= \langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^* \cdot \langle b, c\rangle \\
X_d &= \langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^* \cdot \langle b, c\rangle \cdot \langle c, d\rangle \\
X_e &= \langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^* \cdot \langle b, c\rangle \cdot \langle c, d\rangle \cdot \langle d, e\rangle \\
X_f &= \langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^* \cdot \langle b, f\rangle
\end{aligned}
$$

Computing closed form solutions

Variable elimination $\sim$ Gauss-Jordan

$$
\begin{aligned}
X_r &= 1 \\
X_a &= \langle r, a \rangle \\
X_b &= \langle \\
X_c &= \langle \\
X_d &= \langle \\
X_e &= \langle r, a \rangle \cdot \langle a, b \rangle \cdot (\langle b, c \rangle \cdot \langle c, d \rangle \cdot (\langle d, b \rangle + \langle d, e \rangle \cdot \langle e, b \rangle))^* \cdot \langle b, c \rangle \cdot \langle c, d \rangle \cdot \langle d, e \rangle \\
X_f &= \langle r, a \rangle \cdot \langle a, b \rangle \cdot (\langle b, c \rangle \cdot \langle c, d \rangle \cdot (\langle d, b \rangle + \langle d, e \rangle \cdot \langle e, b \rangle))^* \cdot \langle b, f \rangle
\end{aligned}
$$

Solving single-source path expression problem
$\sim$ computing closed-form solution to left-linear equations

Abstract interpretation of closed forms

$$
\begin{aligned}
\sigma^\sharp(X_r) &\triangleq \mathscr{I}^\sharp[\![1]\!] \\
\sigma^\sharp(X_a) &\triangleq \mathscr{I}^\sharp[\![\langle r, a\rangle]\!] \\
\sigma^\sharp(X_b) &\triangleq \mathscr{I}^\sharp[\![\langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^*]\!] \\
\sigma^\sharp(X_c) &\triangleq \mathscr{I}^\sharp[\![\langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^* \cdot \langle b, c\rangle]\!] \\
\sigma^\sharp(X_d) &\triangleq \mathscr{I}^\sharp[\![\langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^* \cdot \langle b, c\rangle \cdot \langle c, d\rangle]\!] \\
\sigma^\sharp(X_e) &\triangleq \mathscr{I}^\sharp[\![\langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^* \cdot \langle b, c\rangle \cdot \langle c, d\rangle \cdot \langle d, e\rangle]\!] \\
\sigma^\sharp(X_f) &\triangleq \mathscr{I}^\sharp[\![\langle r, a\rangle \cdot \langle a, b\rangle \cdot (\langle b, c\rangle \cdot \langle c, d\rangle \cdot (\langle d, b\rangle + \langle d, e\rangle \cdot \langle e, b\rangle))^* \cdot \langle b, f\rangle]\!]
\end{aligned}
$$

Abstract semantics $\sigma^\sharp$ over-approximates concrete semantics $\sigma^\natural$

Soundness relations

- Say that a approximation relation $\Vdash$ is **soundness relation** if
  1. $f^\natural(a) \Vdash f^\sharp(a)$ for each constant a
  2. $\Vdash$ is compatible with all operations ($\Vdash$ a subalgebra of $A^\natural \times A^\sharp$)

Soundness relations

- Say that a approximation relation $\Vdash$ is **soundness relation** if
  1. $f^{\natural}(a) \Vdash f^{\sharp}(a)$ for each constant a
  2. $\Vdash$ is compatible with all operations ($\Vdash$ a subalgebra of $A^{\natural} \times A^{\sharp}$)
- **Key lemma**: $\Vdash$ is a soundness relation $\Rightarrow \mathscr{I}^{\natural}[\![e]\!] \Vdash \mathscr{I}^{\sharp}[\![e]\!]$ for any e

Soundness relations

- Say that a approximation relation $\Vdash$ is **soundness relation** if
  1. $f^\natural(a) \Vdash f^\sharp(a)$ for each constant a
  2. $\Vdash$ is compatible with all operations ($\Vdash$ a subalgebra of $A^\natural \times A^\sharp$)
- **Key lemma**: $\Vdash$ is a soundness relation $\Rightarrow \mathscr{I}^\natural[\![e]\!] \Vdash \mathscr{I}^\sharp[\![e]\!]$ for any e
- For instance:

$$R \Vdash F(X, X') \iff \text{every } \langle s, s' \rangle \in R \text{ is a model of } F$$

For all

R, S transition relations
F, G transition formulas

such that $R \Vdash F \quad S \Vdash G$
We have:

- $\cdot$: $\{(s, s'') : \exists s'.(s, s') \in R \wedge (s', s'') \in S\} \Vdash \exists X''.F(X, X'') \wedge G(X'', X')$
- $+$: $R \cup S \Vdash F \vee G$
- $*$: overapproximate transitive closure

The algebraic recipe

1. (Modeling) formulate problem of interest as extremal solution to system of equations
2. (Closed forms) design language of “closed forms” & algorithm for computing them
3. (Interpretation) design abstract interpretation & formulate soundness relation

Algebraic reasoning

- Transition formula algebras form **idempotent semirings**
  - $+$ is associative, commutative, and idempotent, and has identity 0
  - $\cdot$ is associative, has identity 1, distributes over $+$, 0 is annihilator

Algebraic reasoning

- Transition formula algebras form **idempotent semirings**
  - $+$ is associative, commutative, and idempotent, and has identity 0
  - $\cdot$ is associative, has identity 1, distributes over $+$, 0 is annihilator
- The $*$ operators from last section satisfy **pre-Kleene algebra** iteration laws.
  - Monotonicity $F \leq G \Rightarrow F^* \leq G^*$, where $x \leq y \iff x + y = y$
    - "more information in $\to$ more information out"
  - Unrolling $(F^n)^* \leq F^*$ for any n
  - ... and more

Algebraic reasoning

- Transition formula algebras form **idempotent semirings**
  - $+$ is associative, commutative, and idempotent, and has identity 0
  - $\cdot$ is associative, has identity 1, distributes over $+$, 0 is annihilator
- The $*$ operators from last section satisfy **pre-Kleene algebra** iteration laws.
  - Monotonicity $F \leq G \Rightarrow F^* \leq G^*$, where $x \leq y \iff x + y = y$
    - "more information in $\rightarrow$ more information out"
  - Unrolling $(F^n)^* \leq F^*$ for any n
  - ... and more
- Laws give users guarantees they mey rely upon
  - Every operation is monotone: user can make progress by supplying "hints"

Algebraic reasoning

- Transition formula algebras form **idempotent semirings**
  - $+$ is associative, commutative, and idempotent, and has identity 0
  - $\cdot$ is associative, has identity 1, distributes over $+$, 0 is annihilator
- The $*$ operators from last section satisfy **pre-Kleene algebra** iteration laws.
  - Monotonicity $F \leq G \Rightarrow F^* \leq G^*$, where $x \leq y \iff x + y = y$
    - "more information in $\rightarrow$ more information out"
  - Unrolling $(F^n)^* \leq F^*$ for any n
  - ... and more
- Laws give users guarantees they mey rely upon
  - Every operation is monotone: user can make progress by supplying "hints"
- Laws give analysis designers guarantees they may exploit
  - Design program transformations that are *guaranteed* to improve precision [Cyphert et al. '19]