Combining Algebraic Domains and Logical Theories by the Reduced Product

Patrick Cousot
di.ens.fr/~cousot  cs.nyu.edu/~pcousot

joint work with

Radhia Cousot  Laurent Mauborgne
di.ens.fr/~rcousot software.imdea.org/people/laurent.mauborgne/
Abstract

• In static analysis by abstract interpretation, algebraic abstract domains can be combined by the reduced product, or over-approximated by the iterated pairwise reductions.

In Satisfiability Modulo Theories (SMT) solvers, the Nelson-Oppen (NO) theory combination schema provides, under various restrictions, a sound and complete decision procedure for the combining of disjoint theories by exchanges of disjunctions of equalities and disequalities. Understood as abstract domains, we show that the NO procedure is an iterated pairwise reduction.

In the context of static analysis, theories can be understood as abstract domains. Completeness is useless in the abstract since the static analysis problem is undecidable anyway. This point of view introduces generalizations which, when combined with bounded widenings (such as interpolation), yields new combinations of algebraic and logical abstractions.

• Joint work with Radhia Cousot, ENS & CNRS, Paris and Laurent Mauborgne, IMDEA, Madrid.
References


Combined, revised and extended into:

- Patrick Cousot, Radhia Cousot, and Laurent Mauborgne. *Theories, Solvers and Static Analysis by Abstract Interpretation*. Submitted to a journal.
  Available from the authors.
Objective
Algebraic abstractions

• Used in abstract interpretation for analysis/verification of finite/infinite systems

• System properties and specifications are abstracted as an algebraic lattice (abstraction-specific encoding of properties)

• Fully automatic: system properties are computed as fixpoints of algebraic transformers

• Abstractions can be combined using the reduced product
Proof theoretic/logical abstractions

- Used in deductive methods
- System properties and specifications are expressed with formulæ of first-order theories (universal encoding of properties)
  - Partly automatic: system properties are provided manually by end-users and automatically checked to satisfy verification conditions (with implication defined by the theories)
- Theories can be combined using Nelson-Oppen procedure
Objective

• Show that proof-theoretic/logical abstractions are particular cases of algebraic abstractions

• Show that the Nelson-Oppen procedure is a particular case of the reduced product

• Use this unifying point of view to introduce a new combination of logical and algebraic abstractions

Convergence of proof theoretic/logical and algebraic property-inference and verification methods
Concrete semantics


**Programs (syntax)**

- **Expressions** *(on a signature \( \langle f, p \rangle \))*

  \[
  \begin{align*}
  x, y, z, \ldots & \in x & \text{variables} \\
  a, b, c, \ldots & \in f^0 & \text{constants} \\
  f, g, h, \ldots & \in f^n, \quad f \triangleq \bigcup_{n \geq 0} f^n & \text{function symbols of arity } n \geq 1 \\
  t & \in T(x, f) & \text{terms} \\
  p, q, r, \ldots & \in p^n, \quad p^0 \triangleq \{ \mathsf{ff}, \mathsf{tt} \}, \quad p \triangleq \bigcup_{n \geq 0} p^n & \text{predicate symbols of arity } n \geq 0, \\
  a & \in A(x, f, p) & \text{atomic formulæ} \\
  e & \in E(x, f, p) \triangleq T(x, f) \cup A(x, f, p) & \text{program expressions} \\
  \varphi & \in C(x, f, p) & \text{clauses in simple conjunctive normal form}
  \end{align*}
  \]

- **Programs** *(including assignment, guards, loops, ...)*

  \[
  \begin{align*}
  P, \ldots & \in P(x, f, p) & \text{programs} \\
  P & \triangleq x := e \mid \varphi \mid \ldots
  \end{align*}
  \]
Programs (mono-interpretation)

- **Interpretation** \( I \in \mathcal{I} \) for a signature \( \langle f, p \rangle \) is such that
  - \( I_V \) is a non-empty set of values,
  - \( \forall c \in f^0 : I_V(c) \in I_V, \quad \forall n \geq 1 : \forall f \in f^n : I_V(f) \in I_V^n \rightarrow I_V, \)
  - \( \forall n \geq 0 : \forall p \in p^n : I_V(p) \in I_V^n \rightarrow \mathcal{B}. \)
  \( \mathcal{B} \triangleq \{ \text{false, true} \} \)

- **Environments**
  \( \eta \in \mathcal{R}_I \triangleq x \rightarrow I_V \) environments

- **Expression evaluation**
  \( \llbracket a \rrbracket_{\eta} \in \mathcal{B} \) of an atomic formula \( a \in \mathcal{A}(x, f, p) \)
  \( \llbracket t \rrbracket_{\eta} \in I_V \) of the term \( t \in \mathcal{T}(x, f) \)
Programs (concrete semantics)

- The program semantics is usually specified relative to a standard interpretation $I \in \mathcal{I}$.

- The concrete semantics is given in post-fixpoint form (in case the least fixpoint which is also the least post-fixpoint does not exist, e.g. inexpressibility in Hoare logic)

\[
\begin{align*}
\mathcal{R}_\mathcal{I} & \quad \text{concrete observables}^5 \\
\mathcal{P}_\mathcal{I} & \triangleq \phi(\mathcal{R}_\mathcal{I}) \quad \text{concrete properties}^6 \\
F_{\mathcal{I}}[P] & \in \mathcal{P}_\mathcal{I} \rightarrow \mathcal{P}_\mathcal{I} \quad \text{concrete transformer of program } P \\
C_{\mathcal{I}}[P] & \triangleq \text{postfp}^\subseteq F_{\mathcal{I}}[P] \in \phi(\mathcal{P}_\mathcal{I}) \quad \text{concrete semantics of program } P
\end{align*}
\]

where $\text{postfp}^\subseteq f \triangleq \{ x \mid f(x) \leq x \}$

---

5 Examples of observables are set of states, set of partial or complete execution traces, infinite/transfinite execution trees, etc.

6 A property is understood as the set of elements satisfying this property.
Example of program concrete semantics

- **Program**
  \[ P \triangleq x=1; \text{ while true } \{ x=\text{incr}(x) \} \]

- **Arithmetic interpretation**
  \[ \mathcal{I} \text{ on integers } \mathcal{I}_\mathcal{V} = \mathbb{Z} \]

- **Loop invariant**
  \[ \text{lfp} \subseteq F_{\mathcal{I}}[P] = \{ \eta \in \mathcal{R}_\mathcal{I} | 0 < \eta(x) \} \]
  where
  \[ \mathcal{R}_\mathcal{I} \triangleq x \rightarrow \mathcal{I}_\mathcal{V} \text{ concrete environments} \]
  \[ F_{\mathcal{I}}[P](X) \triangleq \{ \eta \in \mathcal{R}_\mathcal{I} | \eta(x) = 1 \} \cup \{ \eta[x \leftarrow \eta(x) + 1] | \eta \in X \} \]

- **The strongest invariant** is
  \[ \text{lfp} \subseteq F_{\mathcal{I}}[P] = \bigcap \text{postfp} \subseteq F_{\mathcal{I}}[P] \]

- **Expressivity:** the lfp may not be expressible in the abstract in which case we use the set of possible invariants
  \[ C_{\mathcal{I}}[P] \triangleq \text{postfp} \subseteq F_{\mathcal{I}}[P] \]
Concrete domains

- The **standard semantics** describes computations of a system formalized by elements of a domain of observables $R\mathcal{J}$ (e.g. set of traces, states, etc)

  The properties $P\mathcal{J} \triangleq \varphi(R\mathcal{J})$ (a property is the set of elements with that property) form a complete lattice $\langle P\mathcal{J}, \subseteq, \emptyset, R\mathcal{J}, \cup, \cap \rangle$

- The concrete semantics $C\mathcal{J}[P] \triangleq \text{postfp}^C F\mathcal{J}[P]$ defines the system properties of interest for the verification

- The transformer $F\mathcal{J}[P]$ is defined in terms of primitives, e.g.

  $f_{\mathcal{J}}[x := e]P \triangleq \{ \eta[x \leftarrow [e]_{\mathcal{J}} \eta] \mid \eta \in P \}$ Floyd’s assignment post-condition

  $p_{\mathcal{J}}[\varphi]P \triangleq \{ \eta \in P \mid [\varphi]_{\mathcal{J}} \eta = \text{true} \}$ test
Concrete property satisfaction

- A program \( P \) satisfies a property \( P \) if and only if

\[
\exists C \in C_{\bar{y}}[P] : C \subseteq P
\]
Why using post-fixpoints is more general than using the least fixpoint?

- The least fixpoint may not exist (inexpressive logic) while post-fixpoints do exist (invariants)

- When the least fixpoint does exist, there is a bijection with the post-fixpoints (Tarski [1955])

\[
\text{lfp} \subseteq F_{\mathcal{J}}[P] = \bigcap \text{postfp} \subseteq F_{\mathcal{J}}[P] \in \text{postfp} \subseteq F_{\mathcal{J}}[P]
\]
Multiple concrete interpretations/semantics
About mathematical verification

• A verification relative to a purely mathematical semantics is of poor practical interest.

• Example (Muller’s scheme as analyzed by Kahan)

\[
\begin{align*}
x_0 & = 11/2.0; \\
x_1 & = 61/11.0; \\
\text{for} & \ (i=1 \ ; \ i<=100 \ ; \ i++) \ \{ \\
& \quad x_2 = 111 - (1130 - 3000/x_0) / x_1; \\
& \quad x_0 = x_1; \quad x_1 = x_2; \ \} \\
\end{align*}
\]

• With exact reals, converges to 6 (repulsive fixpoint)

• With any finite precision, converges to 100 (attractive fixpoint)

\[ \implies \text{Programs have many interpretations} \quad I \in \mathcal{I}(\mathcal{S}). \]
Multi-interpreted semantics

- A generalization consists in considering multiple interpretations of logics and programs

- Multi-interpreted properties:

  \[ \mathcal{P}_I \triangleq I \in \mathcal{I} \Rightarrow \varphi(\mathcal{R}_I) \]
  \[ \approx \varphi(\{<I, \eta> | I \in \mathcal{I} \land \eta \in \mathcal{R}_I\}) \]

- Multi-interpreted transformer:

  \[ F_I[P] \in \mathcal{P}_I \rightarrow \mathcal{P}_I \]
  \[ \triangleq \lambda P \in \mathcal{P}_I \cdot \lambda I \in \mathcal{I} \cdot F_I[P](P(I)) \]

- Multi-interpreted semantics:

  \[ C_I[P] \in \varphi(\mathcal{P}_I) \]
  \[ \triangleq \text{postfp} \sqsubseteq F_I[P] \]
  \[ \quad \text{where } \sqsubseteq \text{ is the pointwise subset ordering.} \]
Example I of abstraction of a multi-interpreted semantics

- The float operations have 4 possible interpretations depending on the rounding mode (towards -∞, +∞, 0, closest)
- ASTRÉÉ over-approximates all four semantics

Example II of abstraction of a multi-interpreted semantics

- Ignore some interpretations

\[ \langle \mathcal{P}_I, \subseteq \rangle \leftrightarrow_{\gamma_{I^\# \rightarrow I}} \langle \mathcal{P}_{I^\#}, \subseteq \rangle \text{ is a Galois connection where} \]

\[ \alpha_{I \rightarrow I^\#}(P) \overset{\triangle}{=} P \cap \mathcal{P}_{I^\#} \]

\[ \gamma_{I^\# \rightarrow I}(Q) \overset{\triangle}{=} \left\{ \langle I, \eta \rangle \mid I \in \mathcal{I} \land \eta \in \mathcal{R}_I \land (I \in I^\# \Rightarrow \langle I, \eta \rangle \in Q) \right\} \]
Background on abstract interpretation
Abstract domains

\[ \langle A, \sqsubseteq, \bot, \top, \sqcup, \sqcap, \sqcap, \Delta, \bar{f}, \bar{b}, \bar{p}, \ldots \rangle \]

where

\[ \bar{P}, \bar{Q}, \ldots \in A \]

\[ \sqsubseteq \in A \times A \to \mathcal{B} \]

\[ \bot, \top \in A \]

\[ \sqcup, \sqcap, \sqcap, \Delta \in A \times A \to A \]

\[ \ldots \]

\[ \bar{f} \in (x \times \mathcal{E}(x, f, p)) \to A \to A \]

\[ \bar{b} \in (x \times \mathcal{E}(x, f, p)) \to A \to A \]

\[ \bar{p} \in \mathcal{C}(x, f, p) \to A \to A \]

abstract properties

abstract partial order\(^9\)

infimum, supremum (\(\forall \bar{P} \in A : \bot \subseteq \bar{P} \subseteq \top\))

abstract join, meet, widening, narrowing

abstract forward assignment transformer

abstract backward assignment transformer

abstract condition transformer.
Abstract semantics

- **A** abstract domain
- **⊆** abstract logical implication
- \( \overline{F}[P] \in A \to A \) abstract transformer defined in term of abstract primitives
  - \( \bar{f} \in (x \times E(x, f, p)) \to A \to A \) abstract forward assignment transformer
  - \( \bar{b} \in (x \times E(x, f, p)) \to A \to A \) abstract backward assignment transformer
  - \( \bar{p} \in C(x, f, p) \to A \to A \) abstract condition transformer
- \( \overline{C}[P] \triangleq \{ \text{lfp} \subseteq \overline{F}[P] \} \) least fixpoint semantics, if any
- \( \overline{C}[P] \triangleq \{ \overline{P} \mid \overline{F}[P](\overline{P}) \subseteq \overline{P} \} \) or else, post-fixpoint abstract semantics
Concretization

\[ \gamma \in A \rightarrow P_{\mathcal{J}} \]

- **Soundness** of abstract domains:

  \[(P \subseteq Q) \Rightarrow (\gamma(P) \subseteq \gamma(Q))\]
  order \[\gamma(\bot) = \emptyset\]
  infimum

  \[\gamma(P \cup Q) \supseteq (\gamma(P) \cup \gamma(Q))\]
  join \[\gamma(\top) = \top_{\mathcal{J}}\]
  supremum

- Up to an encoding, the abstraction consists in reasoning on a subset of the concrete properties

\[ \gamma[A] \]

where

\[ \gamma[X] \triangleq \{ \gamma(x) \mid x \in X \} \]
Soundness and completeness of abstract semantics

• The abstract semantics is **sound** iff

\[ \forall P \in A : (\exists \overline{C} \in \overline{C}[P] : \overline{C} \subseteq \overline{P}) \Rightarrow (\exists C \in C[P] : C \leq \gamma(\overline{P})) \]

(any abstract proof of an abstract property can be done in the concrete)

• The abstract semantics is **complete** iff

\[ \forall P \in A : (\exists C \in C[P] : C \leq \gamma(\overline{P})) \Rightarrow (\exists \overline{C} \in \overline{C}[P] : \overline{C} \subseteq \overline{P}) \]

(any concrete proof of an abstract property can be done in the abstract)
Sufficient soundness condition

- **Theorem 4.4** (Soundness of an abstract post-fixpoint semantics). If $C[P] \triangleq postfp \sqsubseteq F[P]$, $\overline{C}[P] \triangleq postfp \sqsubseteq \overline{F}[P]$ and $\gamma : A \rightarrow C$ increasing, then

\[ \forall \overline{P} \in A : F[P] \circ \gamma(\overline{P}) \leq \gamma \circ \overline{F}[P](\overline{P}) \]

implies that the abstract semantics is sound.

- This is usually implied by local conditions to be checked on the abstract domain

\[
\begin{align*}
\gamma(f[x := e]\overline{P}) & \supseteq f_{\overline{x}}[x := e] \gamma(\overline{P}) \\
\gamma(b[x := e]\overline{P}) & \supseteq b_{\overline{x}}[x := e] \gamma(\overline{P}) \\
\gamma(p[x \overline{\varphi} \overline{P}) & \supseteq p_{\overline{x}}[\varphi] \gamma(\overline{P})
\end{align*}
\]

- **Compositionality:**
  
  **Theorem 4.3** (Compositionality of abstractions). The composition of sound (resp. complete) abstractions is sound (resp. complete).
Best abstraction

• If the concretization preserves existing meets then we have a Galois connection

\[ \langle \mathcal{P}_J, \subseteq \rangle \xrightarrow{\gamma} \langle A, \sqsubseteq \rangle \xleftarrow{\alpha} \]

• If no two abstract properties have the same concretization, the abstraction is surjective

\[ \langle \mathcal{P}_J, \subseteq \rangle \xrightarrow{\gamma} \langle A, \sqsubseteq \rangle \xleftarrow{\alpha} \]
Beyond bounded verification: Widening

**Definition 4.6 (Widening).** Let \( \langle A, \sqsubseteq \rangle \) be a poset. Then an over-approximating widening \( \sqcap \in A \times A \mapsto A \) is such that

(a) \( \forall x, y \in A : x \sqsubseteq x \sqcap y \land y \sqsubseteq x \sqcap y \).

A terminating widening \( \sqcap \in A \times A \mapsto A \) is such that

(a) Given any sequence \( \langle x^n, n \geq 0 \rangle \), the sequence \( y^0 = x^0, \ldots, y^{n+1} = y^n \sqcap x^n \), \( \ldots \) converges (i.e. \( \exists \ell \in \mathbb{N} : \forall n \geq \ell : y^n = y^\ell \) in which case \( y^\ell \) is called the limit of the widened sequence \( \langle y^n, n \geq 0 \rangle \)).

Traditionally a widening is considered to be both over-approximating and terminating. \( \square \)
Iteration with widening

Definition 4.7 (Iterates with widening). The iterates of a transformer $\overline{F}[P] \in A \mapsto A$ from the infimum $\bot \in A$ with widening $\nabla \in A \times A \mapsto A$ in a poset $(A, \sqsubseteq)$ are defined by recurrence as $\overline{F}^0 = \bot$, $\overline{F}^{n+1} = \overline{F}^n$ when $\overline{F}[P](\overline{F}^n) \sqsubseteq \overline{F}^n$ and $\overline{F}^{n+1} = \overline{F}^n \nabla \overline{F}[P](\overline{F}^n)$ otherwise.

Theorem 4.8 (Limit of the iterates with widening). The iterates in a poset $(A, \sqsubseteq, \bot)$ of a transformer $\overline{F}[P]$ from the infimum $\bot$ with widening $\nabla$ converge and their limit is a post-fixpoint of the transformer.

Invited talk, SAS 2011, Ca' Foscari, Venezia, Wednesday, September 14th, 2011, 14:00-15:00.
Implementation notes

• Each **abstract domain** \( \langle A, \subseteq, \bot, T, \cup, \cap, \nabla, \Delta, \bar{f}, \bar{b}, \bar{p}, \ldots \rangle \) is **implemented separately**, by providing a specific computer **representation** of properties in \( A \), and **algorithms** for the logical operations \( \subseteq, \bot, T, \cup, \cap, \nabla, \Delta \), and transformers \( \bar{f}, \bar{b}, \bar{p}, \ldots \).

• Different **abstract domains** are combined into a **reduced product**

• Very **efficient** but requires skilled experts
Multi-interpreted first-order logic
First-order logical formulæ & satisfaction

- **Syntax**
  
  \[ \Psi \in \mathcal{F}(x, f, p) \quad \Psi ::= a \mid \neg \Psi \mid \Psi \land \Psi \mid \exists x : \Psi \]  
  
  quantified first-order formulæ

  a distinguished predicate \( = (t_1, t_2) \) which we write \( t_1 = t_2 \).

- **Free variables** \( \vec{x}_\Psi \)

- **Satisfaction**
  
  \( I \models_\eta \Psi \),

  interpretation \( I \) and an environment \( \eta \) satisfy a formula \( \Psi \)

- **Equality**

  \[ I \models_\eta t_1 = t_2 \triangleq \langle t_1 \rangle, \eta =_I \langle t_2 \rangle, \eta \]

  where \( =_I \) is the unique reflexive, symmetric, antisymmetric, and transitive relation on \( I \).
Extension to multi-interpretations

- A property is described by a formula for multiple interpretations

\[ \mathcal{I} \in \varphi(\mathcal{I}) \]

- Semantics of first-order formulae

\[ \gamma_{\mathcal{I}}^a \in F(x, f, p) \rightarrow \mathcal{P}_\mathcal{I} \]
\[ \gamma_{\mathcal{I}}^a(\Psi) \triangleq \{ \langle I, \eta \rangle \mid I \in \mathcal{I} \land I \models_\eta \Psi \} \]

- But how are we going to describe sets of interpretations \( \mathcal{I} \in \varphi(\mathcal{I}) \)?
Defining multiple interpretations as models of theories

- **Theory**: set of theorems (closed sentences without any free variable)

- **Models** of a theory (interpretations making true all theorems of the theory)

\[
\mathcal{M}(\mathcal{T}) \triangleq \{I \in \mathcal{I} \mid \forall \Psi \in \mathcal{T} : \exists \eta : I \models_{\eta} \Psi\} \\
= \{I \in \mathcal{I} \mid \forall \Psi \in \mathcal{T} : \forall \eta : I \models_{\eta} \Psi\}
\]
Classical properties of theories

- **Decidable theories:** \( \forall \Psi \in F(x, f, p) : \text{decide}_T(\Psi) \equiv (\Psi \in \mathcal{T}) \) is computable

- **Deductive theories:** closed by deduction
  \( \forall \Psi \in \mathcal{T} : \forall \Psi' \in F(x, f, p), \text{ if } \Psi \Rightarrow \Psi' \text{ implies } \Psi' \in \mathcal{T} \)

- **Satisfiable theory:**
  \( \mathcal{M}(\mathcal{T}) \neq \emptyset \)

- **Complete theory:**
  for all sentences \( \Psi \) in the language of the theory, either \( \Psi \) is in the theory or \( \neg \Psi \) is in the theory.
Checking satisfiability modulo theory

- **Validity modulo theory**
  \[
  \text{valid}_T(\Psi) \triangleq \forall I \in \mathcal{M}(T) : \forall \eta : I \models_\eta \Psi
  \]

- **Satisfiability modulo theory (SMT)**
  \[
  \text{satisfiable}_T(\Psi) \triangleq \exists I \in \mathcal{M}(T) : \exists \eta : I \models_\eta \Psi
  \]

- **Checking satisfiability** for decidable theories
  \[
  \text{satisfiable}_T(\Psi) \iff \neg (\text{decide}_T(\forall \overline{x}_\Psi : \neg \Psi)) \quad (\text{when } T \text{ is decidable and deductive})
  \]
  \[
  \text{satisfiable}_T(\Psi) \iff (\text{decide}_T(\exists \overline{x}_\Psi : \Psi)) \quad (\text{when } T \text{ is decidable and complete})
  \]

- **Most SMT solvers support only limited forms of quantified formulae**
Example of abstraction: Logical abstractions
Logical abstract domains

- \( \langle A, \mathcal{T} \rangle : A \in \wp(F(x, f, p)) \) abstract properties
  \( \mathcal{T} \) theory of \( F(x, f, p) \)

- Abstract domain \( \langle A, \subseteq, \&, \#, \vee, \wedge, \Delta, \overline{f}_a, \overline{b}_a, \overline{p}_a, \ldots \rangle \)

- Logical implication \( (\Psi \subseteq \Psi') \triangleq ((\forall \overline{x}_\Psi \cup \overline{x}_{\Psi'} : \Psi \Rightarrow \Psi') \in \mathcal{T}) \)

- A lattice but in general not complete

- The concretization is

  \[ \gamma^a_{\mathcal{T}}(\Psi) \triangleq \left\{ \langle I, \eta \rangle \middle| I \in \mathcal{M}(\mathcal{T}) \land I \models_{\eta} \Psi \right\} \]
Logical abstract semantics

- Logical abstract semantics
  \[ \overline{C^a[P]} \triangleq \{ \Psi \mid \overline{F^a[P]}(\Psi) \subseteq \Psi \} \]

- The logical abstract transformer defined in terms of primitives
  - \( \overline{f^a} \in (x \times T(x, f)) \rightarrow A \rightarrow A \)
  - \( \overline{b^a} \in (x \times T(x, f)) \rightarrow A \rightarrow A \)
  - \( \overline{\rho^a} \in \mathbb{L} \rightarrow A \rightarrow A \)

abstract forward assignment transformer
abstract backward assignment transformer
condition abstract transformer
Implementation notes ...

- Universal representation of abstract properties by logical formulæ
- Trivial implementations of logical operations \( \mathbb{S}, \mathbb{T}, \lor, \land \)
- Provers or SMT solvers can be used for the abstract implication \( \subseteq \)
- Concrete transformers are purely syntactic

\[
\begin{align*}
\text{f}_a & \in (x \times T(x, f)) \to F(x, f, p) \to F(x, f, p) \\
\text{f}_a[x := t][\Psi] & \triangleq \exists x' : \Psi[x \leftarrow x'] \land x = t[x \leftarrow x'] \\
\text{b}_a & \in (x \times T(x, f)) \to F(x, f, p) \to F(x, f, p) \\
\text{b}_a[x := t][\Psi] & \triangleq \Psi[x \leftarrow t] \\
\text{p}_a & \in C(x, f, p) \to F(x, f, p) \to F(x, f, p) \\
\text{p}_a[\varphi][\Psi] & \triangleq \Psi \land \varphi
\end{align*}
\]

axiomatic forward assignment transformer

axiomatic backward assignment transformer

axiomatic transformer for program test of condition \( \varphi \).

.../...
but ...

.../... so the **abstract transformers** follow by abstraction

\[
\begin{align*}
\tilde{f}_a[x := t]\Psi & \triangleq \alpha^T_A(f_a[x := t]\Psi) \\
\tilde{b}_a[x := t]\Psi & \triangleq \alpha^T_A(b_a[x := t]\Psi) \\
\tilde{p}_a[\varphi]\Psi & \triangleq \alpha^T_A(p_a[\varphi]\Psi)
\end{align*}
\]

abstract forward assignment transformer
abstract backward assignment transformer
abstract transformer for program test of condition

- The **abstraction algorithm** to abstract properties in \( A \) may be **non-trivial** (e.g. quantifiers elimination)

- A **widening** is needed to ensure convergence of the fixpoint iterates (or else ask the end-user)
Example I of widening: **thresholds**

- Choose a subset $W$ of $A$ satisfying the ascending chain condition for $\sqsubseteq$,
- Define $X \sqcup Y$ to be (one of) the strongest $\Psi \in W$ such that $Y \Rightarrow \Psi$

Example II of bounded widening: **Craig interpolation**

- Use Craig interpolation (knowing a bound e.g. the specification)
- Move to thresholds to enforced convergence after $k$ widenings with Craig interpolation
Reduced Product

Cartesian product

• Definition of the Cartesian product:

Let $\langle A_i, \sqsubseteq_i \rangle$, $i \in \Delta$, $\Delta$ finite, be abstract domains with increasing concretization $\gamma_i \in A_i \rightarrow \Psi_{\text{I}}^{\Sigma_0}$. Their Cartesian product is $\langle \vec{A}, \sqsubseteq \rangle$ where $\vec{A} \triangleq \bigtimes_{i \in \Delta} A_i$, $(\vec{P} \sqsubseteq \vec{Q}) \triangleq \land_{i \in \Delta}(\vec{P}_i \sqsubseteq_i \vec{Q}_i)$ and $\vec{\gamma} \in \vec{A} \rightarrow \Psi_{\text{I}}^{\Sigma_0}$ is $\vec{\gamma}(\vec{P}) \triangleq \bigcap_{i \in \Delta} \gamma_i(\vec{P}_i)$. 
Reduced product

• Definition of the Reduced product:

Let \( \langle A_i, \subseteq_i \rangle, i \in \Delta, \Delta \text{ finite}, \) be abstract domains with increasing concretization \( \gamma_i \in A_i \to \Psi_i^{\Sigma_0} \) where \( \hat{A} \triangleq \times_{i \in \Delta} A_i \) is their Cartesian product. Their reduced product is \( \langle \hat{A}, \hat{\subseteq} \rangle \) where \( \langle \hat{P}, \subseteq \rangle \triangleq (\hat{\gamma}(\hat{P}) = \hat{\gamma}(\hat{Q})) \) and \( \hat{\gamma} \) as well as \( \hat{\subseteq} \) are naturally extended to the equivalence classes \( [\hat{P}]/\subseteq, \) \( \hat{P} \in \hat{A}, \) of \( \subseteq \) by \( \hat{\gamma}([\hat{P}]/\subseteq) = \hat{\gamma}(\hat{P}) \) and \( [\hat{P}]/\subseteq \subseteq [\hat{Q}]/\subseteq \triangleq \exists \hat{P}' \in [\hat{P}]/\subseteq : \exists \hat{Q}' \in [\hat{Q}]/\subseteq : \hat{P}' \subseteq \hat{Q}' . \)

• In practice, the reduced product may be complex to compute but we can use approximations such as the iterated pairwise reduction of the Cartesian product.
Reduction

• **Example:** intervals x congruences

\[ \rho( x \in [-1,5] \land x = 2 \mod 4) \equiv x \in [2,2] \land x = 2 \mod 0 \]

are equivalent

• **Meaning-preserving reduction:**

Let \( \langle A, \sqsubseteq \rangle \) be a poset which is an abstract domain with concretization \( \gamma \in A \rightarrow C \) where \( \langle C, \leq \rangle \) is the concrete domain. A meaning-preserving map is \( \rho \in A \rightarrow A \) such that \( \forall \overline{P} \in A : \gamma(\rho(\overline{P})) = \gamma(\overline{P}) \). The map is a reduction if and only if it is reductive that is \( \forall \overline{P} \in A : \rho(\overline{P}) \sqsubseteq \overline{P} \). \( \square \)
Implementing the reduced product

- **Mathematically**, we can choose any representant of the equivalence class (and normalize to this representant)

- In practice, **normalization is hard** to do

- It is better to choose a **minimal representant**
Why is reduction to a minimal representant important?

- **Without reduction** (signs and parity):

  ```plaintext
  \[\begin{align*}
  & x \geq 0 \rightarrow \text{odd}(x) \\
  & \text{if } (x \leq 0) \text{ then} \\
  & \quad \text{// } x == 0 \rightarrow \text{odd}(x)
  \end{align*}\]

- **With reduction**:

  ```plaintext
  \[\begin{align*}
  & x \geq 0 \rightarrow \text{odd}(x) \\
  & x > 0 \rightarrow \text{odd}(x) \quad \text{minimal representant} \\
  & \text{if } (x \leq 0) \text{ then} \\
  & \quad \text{// false } \rightarrow \text{odd}(x) \quad \text{minimal representant} \\
  & \quad \text{// false } \rightarrow \text{false}
  \end{align*}\]
  ```
Iterated reduction

- Definition of iterated reduction:

Let \( \langle A, \sqsubseteq \rangle \) be a poset which is an abstract domain with concretization \( \gamma \in A \rightarrow C \) where \( \langle C, \sqsubseteq \rangle \) is the concrete domain and \( \rho \in A \rightarrow A \) be a meaning-preserving reduction.

The iterates of the reduction are \( \rho^0 \triangleq \lambda \overline{P} \cdot \overline{P}, \rho^{\lambda+1} = \rho(\rho^\lambda) \) for successor ordinals and \( \rho^\lambda = \bigcap_{\beta < \lambda} \rho^\beta \) for limit ordinals.

The iterates are well-defined when the greatest lower bounds \( \bigcap (\text{glb}) \) do exist in the poset \( \langle A, \sqsubseteq \rangle \).
Finite versus infinite iterated reduction

- **Finite iterations** of a meaning preserving reduction are meaning preserving (and more precise)

- **Infinite iterations**, limits of meaning-preserving reduction, may not be meaning-preserving (although more precise). It is when \( \gamma \) preserves glbs.
Pairwise reduction

- **Definition of pairwise reduction**

  Let \(\langle A_i, \sqsubseteq_i \rangle\) be abstract domains with increasing concretization \(\gamma_i \in A_i \rightarrow L\) into the concrete domain \(\langle L, \leq \rangle\).

  For \(i, j \in \Delta, i \neq j\), let \(\rho_{ij} \in \langle A_i \times A_j, \sqsubseteq_{ij} \rangle \mapsto \langle A_i \times A_j, \sqsubseteq_{ij} \rangle\) be pairwise meaning-preserving reductions (so that \(\forall \langle x, y \rangle \in A_i \times A_j : \rho_{ij}(\langle x, y \rangle) \sqsubseteq_{ij} \langle x, y \rangle\) and \((\gamma_i \times \gamma_j) \circ \rho_{ij} = (\gamma_i \times \gamma_j)^{24}\)).

  Define the pairwise reductions \(\tilde{\rho}_{ij} \in \langle \tilde{A}, \sqsubseteq \rangle \mapsto \langle \tilde{A}, \sqsubseteq \rangle\) of the Cartesian product as

  \[
  \tilde{\rho}_{ij}(\tilde{P}) \triangleq \text{let } \langle \tilde{P}', \tilde{P}' \rangle \triangleq \rho_{ij}(\langle \tilde{P}_i, \tilde{P}_j \rangle) \text{ in } \tilde{P}[i \leftarrow \tilde{P}'_i][j \leftarrow \tilde{P}'_j]
  \]

  where \(\tilde{P}[i \leftarrow x]_i = x\) and \(\tilde{P}[i \leftarrow x]_j = \tilde{P}_j\) when \(i \neq j\).

---

24 We define \((f \times g)(\langle x, y \rangle) \triangleq \langle f(x), g(y) \rangle\).
Define the iterated pairwise reductions $\bar{\rho}^n, \bar{\rho}^\lambda, \bar{\rho}^*$ in $\langle \hat{A}, \vec{\subseteq} \rangle \mapsto \langle \hat{A}, \vec{\subseteq} \rangle$, $n \geq 0$ of the Cartesian product for

$$\bar{\rho} \triangleq \bigcirc_{i,j \in \Delta, i \neq j} \bar{\rho}_{ij}$$

where $\bigcirc_{i=1}^n f_i \triangleq f_{\pi_1} \circ \ldots \circ f_{\pi_n}$ is the function composition for some arbitrary permutation $\pi$ of $[1, n]$.  

□
Iterated pairwise reduction

- The iterated pairwise reduction of the Cartesian product is meaning preserving.

If the limit \( \hat{\rho}^* \) of the iterated reductions is well defined then the reductions are such that \( \forall \vec{P} \in \vec{A} : \forall n \in \mathbb{N}_+ : \hat{\rho}^*(\vec{P}) \subseteq \hat{\rho}^n(\vec{P}) \subseteq \hat{\rho}_{ij}(\vec{P}) \subseteq \vec{P} \), \( i, j \in \Delta, i \neq j \) and meaning-preserving since \( \hat{\rho}^*(\vec{P}), \hat{\rho}_{ij}(\vec{P}), \vec{P} \in [\vec{P}]/\equiv \).

If, moreover, \( \gamma \) preserves greatest lower bounds then \( \hat{\rho}^*(\vec{P}) \in [\vec{P}]/\equiv \). \( \square \)
Iterated pairwise reduction

• In general, the iterated pairwise reduction of the Cartesian product is not as precise as the reduced product

• Sufficient conditions do exist for their equivalence
Counter-example

- $L = \emptyset(\{a, b, c\})$
- $A_1 = \{\emptyset, \{a\}, \top\}$
- $A_2 = \{\emptyset, \{a, b\}, \top\}$
- $A_3 = \{\emptyset, \{a, c\}, \top\}$
- $\langle \top, \{a, b\}, \{a, c\}\rangle/\equiv = \langle \{a\}, \{a, b\}, \{a, c\}\rangle$
- $\vec{\rho}_{ij}(\langle \top, \{a, b\}, \{a, c\}\rangle) = \langle \top, \{a, b\}, \{a, c\}\rangle$
- $\vec{\rho}^*(\langle \top, \{a, b\}, \{a, c\}\rangle) = \langle \top, \{a, b\}, \{a, c\}\rangle$ is not

for $\Delta = \{1, 2, 3\}, i, j \in \Delta, i \neq j$

where $\top = \{a, b, c\}$

a minimal element of $[\langle \top, \{a, b\}, \{a, c\}\rangle]/\equiv$
Nelson-Oppen combination procedure
The Nelson-Oppen combination procedure

- Prove **satisfiability in a combination of theories** by exchanging equalities and disequalities

- **Example:** \( \varphi \overset{\triangle}{=} (x = a \lor x = b) \land f(x) \neq f(a) \land f(x) \neq f(b) \)

- **Purify:** introduce auxiliary variables to separate alien terms and put in conjunctive form

\[
\varphi \overset{\triangle}{=} \varphi_1 \land \varphi_2 \quad \text{where} \\
\varphi_1 \overset{\triangle}{=} (x = a \land x = b) \land y = a \land z = b \\
\varphi_2 \overset{\triangle}{=} f(x) \neq f(y) \land f(x) \neq f(z)
\]

---

\(^{22}\) where \(a, b\) and \(f\) are in different theories
The Nelson-Oppen combination procedure

\[ \varphi \triangleq \varphi_1 \land \varphi_2 \text{ where} \]
\[ \varphi_1 \triangleq (x = a \lor x = b) \land y = a \land z = b \]
\[ \varphi_2 \triangleq f(x) \neq f(y) \land f(x) \neq f(z) \]

- Reduce \( \tilde{\rho}(\varphi) \): each theory \( T_i \) determines \( E_{ij} \), a (disjunction) of conjunctions of variable (dis)equalities implied by \( \varphi_j \) and propagates it in all other components \( \varphi_i \)

\[ E_{12} \triangleq (x = y) \lor (x = z) \]
\[ E_{21} \triangleq (x \neq y) \land (x \neq z) \]

- Iterate \( \tilde{\rho}^*(\varphi) \) : until satisfiability is proved in each theory or stabilization of the iterates
The Nelson-Oppen combination procedure

Under **appropriate hypotheses** (disjointness of the theory signatures, stably-infiniteness/shininess, convexity to avoid disjunctions, etc), the Nelson-Oppen procedure:

- **Terminates** (finitely many possible (dis)equalitys)
- **Is sound** (meaning-preserving)
- **Is complete** (always succeeds if formula is satisfiable)
- Similar techniques are used in theorem provers
Is completeness of the Nelson-Oppen procedure needed?

- **Yes**, if you want to win the SMT-COMP competition (*)
- **No**, for program static analysis/verification
  - Verification is *undecidable* anyway so requiring completeness is useless.
  - Therefore these hypotheses (disjointness of the theory signatures, stably-infiniteness/shininess, convexity, etc) can be lifted, the procedure is then sound and incomplete.
  - **No change to SMT solvers is needed.**

(*) congratulations to Z3 for SMT-COMP 2011, http://www.smtexec.org/exec/?jobs=856
The Nelson-Oppen procedure is an iterated pairwise reduced product
Observables in Abstract Interpretation

• (Relational) abstractions of values \((v_1, \ldots, v_n)\) of program variables \((x_1, \ldots, x_n)\) is often too imprecise.

Example: when analyzing quaternions \((a, b, c, d)\) we need to observe the evolution of \(\sqrt{a^2+b^2+c^2+d^2}\) during execution to get a precise analysis of the normalization.

• An observable is specified as the value of a function \(f\) of the values \((v_1, \ldots, v_n)\) of the program variables \((x_1, \ldots, x_n)\) assigned to a fresh auxiliary variable \(x_0\)

\[ x_0 == f(v_1, \ldots, v_n) \]

(with a precise abstraction of \(f\))
Purification = Observables in A.I.

• The purification phase consists in introducing new observables

• The program can be purified by introducing auxiliary assignments of pure sub-expressions so that forward/backward transformers of purified formulæ always yield purified formulæ

• Example (f and a,b are in different theories):

\[ y = f(x) == f(a+1) \land f(x) == f(2*b) \]

becomes

\[ z=a+1; t=2*b; y = f(x) == f(z) \land f(x) = f(t) \]
Reduction

• The transfer of a (disjunction of) conjunctions of variable (dis-)equalities is a pairwise iterated reduction

• This can be incomplete when the signatures are not disjoint
Static analysis combining logical and algebraic abstractions
When checking satisfiability of $\varphi_1 \land \varphi_2 \land \ldots \land \varphi_n$, the Nelson-Oppen procedure generates (dis)-equalities that can be propagated by $\rho_{la}$ to reduce the $P_i$, $i = 1, \ldots, m$, or

- $\alpha_i(\varphi_1 \land \varphi_2 \land \ldots \land \varphi_n)$ can be propagated by $\rho_{la}$ to reduce the $P_i$, $i = 1, \ldots, m$

- The purification to theory $\mathcal{T}_i$ of $\gamma_i(P_i)$ can be propagated to $\varphi_i$ by $\rho_{al}$ in order to reduce it to $\varphi_i \land \gamma_i(P_i)$ (in $\mathcal{T}_i$)
Advantages

• **No need for completeness hypotheses** on theories

• **Bidirectional reduction** between logical and algebraic abstractions

• **No need for end-users to provide** inductive invariants (discovered by static analysis) (*)

• Easy interaction with end-user (through logical formulæ)

• Easy introduction of **new abstractions** on either side

⇒ Extensible expressive static analyzers / verifiers

(*) may need occasionally to be strengthened by the end-user
Future work

• Still at a conceptual stage

• More experimental work on a prototype is needed to validate the concept
Conclusion

• Future convergence between logic-based proof-theoretic deductive methods using SMT solvers/theorem provers and algebraic methods using abstract interpretation for infinite-state systems?

• Expressiveness is important

• Efficiency is decisive

• Reproducibility is crucial
The End

Thank You