Prime Factorizations of Abstract Domains
Using First-Order Logic

Elena Marchiori

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
and
University of Leiden, P.O. Box 9512, 2300 RA Leiden, The Netherlands
e-mail: elena@cwi.nl

Abstract. A methodology is introduced based on first-order logic, for the design and decomposition of abstract domains for abstract interpretation. First, an assertion language is chosen that describes the properties of interest. Next, abstract domains are defined to be suitably chosen sets of assertions. Finally, computer representations of abstract domains are defined in the expected way, as isomorphic copies of their specification in the assertion language. In order to decompose abstract domains, the notion of prime (conjunctive) factorization of sets of assertions is introduced. We illustrate this approach by considering typical abstract domains for ground-dependency and aliasing analysis in logic programming.

1 Introduction

In the theory of abstract interpretation [3], abstract domains are (computer) representations of properties. The semantics of an abstract domain is given by a function called concretization, that maps elements of the abstract domain into elements of a 'concrete domain'. Two fundamental aspects of the study of abstract domains are the investigation of representations supporting efficient implementations, and the comparative analysis of the properties represented by abstract domains. This paper is concerned with the latter aspect.

Previous work on this subject is mainly based on two equivalent techniques (cf. [3]): Galois connections and closure operators. In [3] comparison of abstract domains is defined by means of the notion of abstraction, where an abstract domain is more abstract than another one if there is a Galois insertion from the first into the latter. This notion is weakened in [7], where the comparison is defined w.r.t. a given property, by means of the notion of quotient of one abstract domain w.r.t. another one, describing the part of the former abstract domain that is useful for computing the information described by the latter one. In [4], the approach based on closure operators is used for investigating domain complementation in abstract interpretation. The authors formalize the concept of decomposition of an abstract domain, as a set of abstract domains whose reduced product yields the initial abstract domain and use the notion of pseudo-complement for decomposing abstract domains.
In this paper we propose a method based on first-order logic for the design and decomposition of abstract domains. First, an assertion language is chosen whose syntax specifies the properties of interest, and whose semantics is fixed by means of a structure characterizing the meaning of the predicates in accordance with the properties they are supposed to describe. Next, an abstract domain is defined to be a suitably chosen set of assertions. Finally, computer representations of abstract domains are defined in the expected way, i.e., they have to respect (i.e., be isomorphic to) their specification in the assertion language. In order to decompose abstract domains, the notion of prime (conjunctive) factorization of sets of assertions is introduced. This is a standard algebraic notion of factorization, where an abstract domain is factorized in pairwise ‘disjoint’ parts. This method has various benefits. First, it allows one to focus only on the abstract domains that describe the properties of interest, that are those expressible in the chosen assertion language. This is not the case for the standard methods above mentioned, where all possible abstract domains (on the concrete domain) are taken into account. Moreover, using our method abstract domains can be decomposed in ‘disjoint’ factors. This desirable property is not guaranteed in the decompositions obtained using the approach of [4]. Finally, the two phases of design and computer representation of an abstract domain are neatly separated, where the design phase is performed at the logical level.

We illustrate this approach by considering typical abstract domains for ground-dependency and aliasing analysis in logic programming. The fragment $\mathcal{L}$ of a first-order assertion language introduced in [13] (actually, a slight extension of this) is used. Logical descriptions of various abstract domains are given: $\text{Def}$ [8] and $\text{Pos}$ [14, 15] for ground-dependency analysis; $\text{Sharing}$ [10] and $\text{ASub}$ [18] for aliasing analysis. Maximal factorizations for these domains are obtained by inspecting the structure of the assertions in the abstract domains, and they are used for analyzing and comparing the abstract domains.

The paper is organized as follows. The next section introduces our methodology. Section 3 presents an assertion language for the design of typical abstract domains for logic programming, and Section 4 contains a comparative study of various abstract domains for logic programming. Finally, Section 5 contains a discussion on related work and some conclusive remarks.

2 Abstract Domains in Assertion Form

First-order logic is a familiar formalism, used for specifying as well as for reasoning about properties. We show in this section how first-order logic can be used for the design and decomposition of abstract domains for abstract interpretation.

Here and in the sequel $\mathcal{L}$ denotes a generic assertion language. We assume that the semantics of the predicates in $\mathcal{L}$ is fixed according to their intended meaning, by a given structure denoted by $\mathcal{M}$. Assertions are indicated by $\phi, \psi$. As already mentioned, abstract domains represent properties of some syntactic objects, usually a subset of the variables of the considered program. Thus, the definition of abstract domain we give is parametric with respect to a set $V$
of syntactic objects. We adopt the following convenient assumptions: 1. \( V \) is (identified with) a set of distinct variables of \( \mathcal{L} \); 2. in the definition of abstract domain, only the set of assertions of \( \mathcal{L} \) whose free variables are contained in \( V \) is considered, denoted by \( A(\mathcal{L}, V) \); 3. assertions with the same meaning are identified.

This last assumption amounts to consider equivalence classes of assertions of \( A(\mathcal{L}, V) \), where \([\phi]\) denotes all the assertions that are logically equivalent to \( \phi \). For simplicity, in the sequel the squares in \([\phi]\) are often omitted.

**Definition 1. (Abstract Domain on \( \mathcal{L} \))** An abstract domain (on \( \mathcal{L} \)), denoted by \( \mathcal{A} \) (possibly subscripted), is a set of assertions of \( A(\mathcal{L}, V) \) containing \( \text{false} \), and closed under conjunction and variance

Observe that this definition of abstract domain is consistent with the original one given by the Cousots (cf. [3]): the 'concrete domain' is the set of sets of valuations, and the 'concretization function' maps an assertion \( \mathcal{A} \) into the set of valuations that satisfy \( \mathcal{A} \).

In the sequel, for simplicity, we shall often avoid to mention the element \( \text{false} \) when specifying the set of assertions of an abstract domain.

**Example 1.** A simple abstract domain for the study of the sign of program variables assuming integer values is given in [3]. For a considered set \( V \) of program variables, this domain can be specified in our formalism as follows: \( \mathcal{L} \) contains the constants and function symbols of the program, and the unary predicates \( \geq, \leq \); \( \mathcal{M} \) maps terms into integers according with their intended interpretation, and specify the semantics of \( \geq, \leq \) in the expected way. Then the abstract domain for the study of the sign of the variables in \( V \) can be described by the set \( \text{Sign}_V \) of assertions that are conjunctions of atoms of the form \( x \geq 0 \), or \( x \leq 0 \), with \( x \) in \( V \).

Viewing abstract domains as sets of assertions allows us to use the following notion of (conjunctive) factorization for decomposing (in \( \mathcal{L} \)) an abstract domain in 'disjoint' parts.

The notation \( \mathcal{A}_1 = \mathcal{A}_2 \) is used, meaning that \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \) contain the same equivalence classes. Moreover, for two sets \( \mathcal{A}_1, \mathcal{A}_2 \) of assertions, \( \mathcal{A}_1 \cap \mathcal{A}_2 \) denotes the set \( \{ [\phi_1 \land \phi_2] \mid \phi_1 \in \mathcal{A}_1, \phi_2 \in \mathcal{A}_2 \} \).

**Definition 2. (Prime Factorization on \( \mathcal{L} \))** The set \( \{ \mathcal{A}_1, \ldots, \mathcal{A}_n \} \) is a (conjunctive) prime factorization of \( \mathcal{A} \) if the following conditions hold:

(a) If \( n > 1 \) then \( \mathcal{A}_i \neq \{ \text{true, false} \} \), for \( i \in [1, n] \);
(b) for every \( i \neq j \) \( \mathcal{A}_i \cap \mathcal{A}_j = \{ \text{true, false} \} \);
(c) \( \mathcal{A}_1 \land \ldots \land \mathcal{A}_n = \mathcal{A} \).

We call \( \mathcal{A} \) reduced if it has only one factorization. Moreover, a factorization of \( \mathcal{A} \) is maximal if \( \mathcal{A}_i \) is reduced, for \( i \in [1, n] \).

---

1 Recall that a variant of an assertion \( \phi \) is any assertion \( \phi \sigma \) obtained by applying to \( \phi \) a substitution \( \sigma \) that renames the variables of \( \phi \).
The name ‘prime’ in the above definition is used to underline the similarity of our definition with the standard algebraic notion of factorization of integers in relatively prime factors. For simplicity, in the sequel we write ‘factorization’ instead of ‘prime factorization’. Clearly, if $A$ is reduced then $\{A\}$ is its only factorization, and it is maximal.

Example 2. It is easy to check that $\{\text{Sign}_{\leq 0}, \text{Sign}_{> 0}\}$ is a maximal factorization of $\text{Sign}_V$, where $\text{Sign}_{\leq 0}$ is the set of assertions that are conjunctions of atoms of the form $x \leq 0$, with $x$ in $V$, and where $\text{Sign}_{> 0}$ is defined analogously.

In order to improve the precision of the static analysis of logic programs, abstract domains can be composed by means of the notion of reduced-product ([3]). Intuitively, the reduced product of two domains is obtained from the cardinal product of the domains by identifying pairs of elements whose conjunction represent the same information. A factorization yields a reduced-product in the expected way.

Proposition 3. If $\{A_1, \ldots, A_n\}$ is a factorization of $A$ then $A$ is the reduced-product of $A_1, \ldots, A_n$.

Proof. Let Val denote the set of valuations. $M$ provides a Galois insertion of an abstract domain $A$ into the concrete domain $2^{\text{Val}}$ consisting of sets of valuations. This Galois insertion is determined by the concretization function $\gamma_A$ that maps an assertion $\phi$ of $A$ into the set of valuations that satisfy $\phi$. Observe that $\gamma_A$ is injective because equivalent assertions are identified. Then the operator $\wedge$ on abstract domains (on $L$) is a reduced-product operator. □

The benefit of using this first-order framework is that the definition, decomposition and comparison of abstract domains can be performed in a uniform and familiar setting. However, (computer) representations of abstract domains for their efficient manipulation ([9]) often need different lattice structures (see, e.g., [2] for ground-dependency analysis). Therefore the notion of representation of an abstract domain is defined as follows. First, we need some preliminary terminology. The following notion of embedding of an abstract domain into $L$ is used. Here and in the the sequel $D$ denotes an abstract domain (on any complete lattice) and $\gamma_D$ denotes its concretization function (cf. [3]).

Definition 4. (Embedding) An embedding of $D$ in $L$ is an injective mapping $\varepsilon_D : D \to L$ s.t. for every $D$ in $D$, $\alpha$ is in $\gamma_D(D)$ if and only if $\varepsilon(D)$ is true under $\alpha$. □

Thus an embedding of a domain into $L$ consists of the (equivalence classes of the) assertions $\phi_D$ characterizing the sets $\gamma_D(D)$ of valuations, with $D$ in $D$. The following result is an easy consequence of the definition of concretization function ([3]).
Proposition 5. The image $\varepsilon_D(D)$ of an embedding is an abstract domain on $L$.

We can now formalize the concept of representation domain.

Definition 6. (Representation Domain) $D$ is a representation of $A$ (or equivalently $A$ and $D$ are isomorphic, denoted by $A \simeq D$) if there exists an embedding $\varepsilon_D$ s.t. $A = \varepsilon_D(D)$.

The definition of representation domain clarifies the role of domains in assertion form, as those used in the design phase, in contrast to the representation domains used in the (efficient) implementation.

Example 3. Suppose $V = \{x\}$. Then a representation of $\text{Sign}_V$ is the familiar lattice pictured below

\[ \begin{array}{c}
\top \\
\geq 0 \\
\leq 0 \\
\bot
\end{array} \]

We conclude this section with a discussion on the relationship of our approach with the one based on closure operators. Following [3], one can associate with each abstract domain an (upper) closure operator (on sets of valuations) by means of the concretization function $\gamma$ mapping an assertion into the set of valuations that satisfy that assertion. The closure operator associated with an abstract domain is the set of sets of valuations obtained by applying $\gamma$ to each of its assertions. In the standard approach, also the vice versa holds, i.e., the lattice of abstract domains is isomorphic to the lattice of upper closure operators. This result does not hold when Definition 1 is considered, for two reasons. One is the hypothesis of closure under variance w.r.t. $V$: a set of valuations that is not closed under variance (w.r.t. $V$)\(^2\) is an (upper) closure operator, but it is not an abstract domain (on $L$). The other reason is related to the expressivity of the chosen assertion language $L$: (the image of) a closure operator is an abstract domain (on $L$) only if it can be described by means of a set of assertions (of $L$). However, if one assumes that the assertion language allows to describe all sets of valuations closed under variance (w.r.t. $V$), then the lattice of abstract domains (according with Definition 1) is isomorphic to the lattice of upper closure operators on sets of valuations closed under variance (w.r.t. $V$).

\(^2\) The notion of variant w.r.t. $V$ of a set $d$ of valuations is defined in the expected way: let $\rho$ be a substitution that renames the variables of $V$ with other variables of $V$. Then a variance of $d$ is obtained by applying $\rho$ to the domain of every valuation.
Abstract Domains for Logic Programming

In this section, we show how a slight extension of the first-order assertion language $L$ introduced in [13] can be used for the design and decomposition of typical abstract domains for the static analysis of logic programs.

Term properties, like groundness and sharing, have been identified as crucial when analyzing the run-time behaviour of logic programs. For instance, ground-dependency analysis can be used for compile optimization, by using matching instead of unification when it is known that at a given program point a variable is bound to a ground term every time the execution reaches that point. Information on the sharing among variables in a logic program is useful for important optimizations, like and-parallelism. The assertion language here considered allows to express properties of terms, like groundness, freeness, linearity, sharing, covering and independency. Informally, a term is ground if it does not contain variables, it is free if it is a variable, and it is linear if every variable occurs in it at most once. Moreover, a set of terms share if they have at least one common variable, while they are independent if they do not share. Finally, a term is covered by a set of terms if the set of its variables is contained in the union of the sets of variables of the terms in that set. For instance, the term $f(x, y)$ is covered by the set $\{g(x), g(y)\}$.

In order to define $L$, a countable set $\text{Var}$ of (logical) variables is used, denoted by $v, x, y, z$, possibly subscripted. Here and in the sequel, $S$ represents a finite set of logical variables, and $|S|$ its cardinality. Moreover, the notation $S \subset S'$ indicates that $S$ is a proper subset of $S'$.

**Definition 7. (The Assertion Language)** Let $L'$ be the smallest set $F$ of formulas containing atoms of the form $\text{var}(x)$, $\text{ground}(x)$, $\text{linear}(x)$, $\text{share}(S)$, and s.t. if $\phi_1$ and $\phi_2$ are in $F$ then $\lnot \phi_1$ and $\phi_1 \land \phi_2$ are also in $F$. The assertion language $L$ consists of all the formulas of the form $\forall x_1, \ldots, x_n(\phi)$, with $\phi \in L'$, and $n \geq 0$.

The formula $\phi \lor \psi$ is used as a shorthand for $\lnot (\lnot \phi \land \lnot \psi)$, $\phi \Rightarrow \psi$ denotes $\lnot \phi \lor \psi$, and $\phi \Leftrightarrow \psi$ stands for $(\phi \Rightarrow \psi) \land (\psi \Rightarrow \phi)$. Moreover, the propositional constants true and false are assumed to be in $L$, where true is identified with the conjunction over the empty set of assertions $\land \emptyset$ and false with $\lor \emptyset$. In the sequel, the notation $\text{share}(x, y)$ is used as shorthand of $\text{share}(\{x, y\})$, with $x, y$ distinct.

Observe that only a weak form of universal quantification is allowed, where $\forall$ does not occur in the scope of any $\lnot$. For instance, $\forall z (\text{var}(z) \land \lnot \text{share}(\{z, x\}))$ is in $L$, but $\lnot \forall z (\text{var}(z) \land \lnot \text{share}(\{z, x\}))$ is not in $L$.

The meaning of assertions in $L$ is specified by means of the following structure $M$. Let $\text{OVar}$ be the set of (object) variables, here identified for simplicity with $\text{Var}$, and let $\text{Fun}$ be a set of functors with rank (constants are identified with functors of rank 0). In the following, $\text{occ}(x, \tau)$ denotes the number of occurrences of the variable $x$ in the term $\tau$, and $\text{OVar}(\tau)$ the set of (object) variables occurring in $\tau$. 

Definition 8. The structure $\mathcal{M}$ contains the universe $\mathcal{U}$ consisting of the (object) terms built on $OVar$ and $Fun$. Moreover, for each predicate symbol $p$ of $\mathcal{L}$, $\mathcal{M}$ contains a predicate in $\mathcal{U}$, also denoted by $p$, with the following semantics:

- $\mathcal{M} \models \text{var}(\tau)$ if $\tau \in OVar$
- $\mathcal{M} \models \text{ground}(\tau)$ if $OVar(\tau) = \emptyset$
- $\mathcal{M} \models \text{linear}(\tau)$ if $\text{occ}(x, \tau) = 1$ for every $x$ in $OVar(\tau)$
- $\mathcal{M} \models \text{share}(\{\tau_1, \ldots, \tau_n\})$ if $\bigcap_{i=1}^n OVar(\tau_i) \neq \emptyset$

Example 4. The assertion $\neg \text{share}(\{x, y, z\}) \lor \text{share}(\{x, y\})$ is valid in $\mathcal{M}$. In fact, for every valuation $\alpha$, if $OVar(x\alpha) \cap OVar(y\alpha) \neq \emptyset$ then $\mathcal{M} \models \text{share}(\{x, y\})\alpha$, otherwise $\mathcal{M} \models \neg \text{share}(\{x, y\})\alpha$.

Note that even if $\text{share}$ is not first-order (its argument is a set), it can be expressed in first-order logic by means of a family of first-order predicates $\text{share}_n$ of rank $n$, with $n \geq 0$. The set of valid (in $\mathcal{M}$) assertions of $\mathcal{L}$ has been characterized by means of a complete and decidable theory $T$, by means of a simple axiomatization (see [13]).

The completeness and decidability of $T$ provides an automatic tool for proving properties of some elements of an abstract domain, in the following way. In order to prove that an element $\phi$ of a domain satisfies a property $P$, specified in $\mathcal{L}$ by means of the assertion $\psi$, it is sufficient to check the validity of the implication $\phi \Rightarrow \psi$.

In order to use $\mathcal{L}$ for the static analysis of logic programs, it is necessary to assume that $\mathcal{U}$ contains the constants and function symbols of the considered class of programs. Moreover, we adopt the notation of the previous section: $V$ denotes the set of (logical) variables representing the considered (program) variables, and $A(\mathcal{L}, V)$ the set of assertions of $\mathcal{L}$ whose free variables are contained in $V$. Therefore substitutions are identified with valuations.

An abstract domain (on $\mathcal{L}$) is specified according with Definition 1. Observe that we obtain a more specific notion of abstract domain than the original one (cf. [3]), because of the choice of the assertion language, and because of the condition of closure under variance. For instance, $\{\text{ground}(x), \text{true}, \text{false}\}$ would represent an abstract domain in the original definition, but it is not a legal one in our definition (unless $V = \{x\}$). The condition of closure under variance w.r.t. $V$ has been implicitly assumed in the literature on abstract interpretation of logic programs, but it has not been taken into account when reasoning about these domains using the standard techniques based on Galois insertions or closure operator (cf. [4]).

We conclude this section with a simple example.

Example 5. Consider the abstract domain $\text{Con}$ introduced by Mellish [17] and used in early mode and groundness analyzers [12]. $\text{Con}$ consists of the bottom element $\bot$, and of the sets $S = \{x_1, \ldots, x_n\}$ of variables of $V$, with concretization function mapping $\bot$ into $\emptyset$ and $\gamma_{\text{Con}}(S) = \{\sigma \mid OVar(x\sigma) = \emptyset \text{ for all } x \in S\}$. 

\[ \text{Example 5. Consider the abstract domain } \text{Con} \text{ introduced by Mellish [17] and used in early mode and groundness analyzers [12]. } \text{Con} \text{ consists of the bottom element } \bot, \text{ and of the sets } S = \{x_1, \ldots, x_n\} \text{ of variables of } V, \text{ with concretization function mapping } \bot \text{ into } \emptyset \text{ and } \gamma_{\text{Con}}(S) = \{\sigma \mid OVar(x\sigma) = \emptyset \text{ for all } x \in S\}. \]
Let $A_{Con}$ be the set of assertions that are conjunctions of atoms of the form $\text{ground}(x)$, with $x$ in $V$. It is easy to show that $A_{Con}$ satisfies Definition 1, and that $Con$ is a representation of $A_{Con}$, by considering the embedding $\epsilon_{Con}$ that maps $\perp$ into $false$ and a set $\{x_1, \ldots, x_n\}$ into the assertion $\text{ground}(x_1) \land \ldots \land \text{ground}(x_n)$.

4 Ground-Dependency and Aliasing Analysis

This section contains a comparative analysis of various abstract domains for the static analysis of logic programs, namely $Def$, $Pos$, $Sharing$ and $ASub$. Each of these domains is shown to be the representation of an abstract domain on $L$. These logical characterizations in $L$ of the domains are used for deriving their maximal factorizations, for studying and comparing the original domains, as well as for defining new ones.

4.1 $Def$ in Logical Form

The abstract domain $Def$ was introduced by Marriott and Søndergaard for ground-dependency analysis in [15], based on previous work by Dart ([8]) on groundness analysis in deductive databases. We show that $Def$ can be factorized into two reduced domains, describing groundness and covering, respectively.

First, we recall the definition of $Def$. $Def$ is the largest class of positive boolean functions whose models are closed under intersection, augmented with the bottom element $false$. Recall that a boolean function $F$ is positive if $F(\text{true}, \ldots, \text{true}) = \text{true}$. Here boolean functions are represented by (equivalence classes of) propositional formulas, as e.g. in [15]. In order to define the concretization function $\gamma_{Def}$, substitutions are viewed as truth assignments as follows. For a substitution $\sigma$, the truth assignment $\text{groundss}\sigma$ maps a propositional variable $x$ to $true$ if $x\sigma$ is ground, and to $false$ otherwise. Moreover, the notion of instance $\sigma'$ of a substitution $\sigma$ is used, meaning that $\sigma'$ is obtained by composing $\sigma$ with some substitution. The concretization function $\gamma_{Def}$ maps an element $F$ of $Def$ into the set $\gamma_{Def}(F)$ of those substitutions $\sigma$ s.t. for every instance $\sigma'$ of $\sigma$, $F$ under the truth assignment $\text{groundss}\sigma'$ is $true$. Intuitively, $\gamma_{Def}(F)$ extracts the 'monotonic' (in the sense that its truth is preserved under instantiation) information described by the propositional formula $F$.

Consider the following abstract domain $A_{Def}$ on $L$.

**Definition 9.** $A_{Def}$ is the set of assertions that are conjunctions of formulas of the form $\forall z (\text{var}(z) \land \text{share}(z, x) \Rightarrow \text{share}(z, y_1) \lor \ldots \lor \text{share}(z, y_n))$, with $n \geq 0$, where $x, y_1, \ldots, y_n$ are in $V$, and $z$ is a fresh variable.

We show that $Def$ is a representation of $A_{Def}$, and provide a maximal factorization of $A_{Def}$.

First, $Def$ is characterized in logical form by means of the following transformation. We use the representation of an element $F$ in $Def$ as a conjunction of formulas, called definite clauses, of the form $y_1 \land \ldots \land y_n \rightarrow x$ with $n \geq 0$ (see [8, 2]).
Definition 10. The transformation $\varepsilon_{Def} : Def \rightarrow \mathcal{L}$ maps $F$ into $\phi_F$, defined as follows:

- $\phi_F = \forall z (\text{var}(z) \land \text{share}(z, x) \Rightarrow \text{share}(z, y_1) \lor \ldots \lor \text{share}(z, y_n))$ if $F = y_1 \land \ldots \land y_n \rightarrow x$.
- $\phi_F = \phi_{F_1} \land \ldots \land \phi_{F_k}$ if $F = F_1 \land \ldots \land F_k$, $k \geq 0$, and all the $F_i$'s are definite clauses.

Observe that, for $n = 0$ we obtain the assertion $\forall z (\text{var}(z) \land \text{share}(z, x) \Rightarrow \text{false})$, that is equivalent to ground($x$).

**Example 6.** The element $x \land (y \leftrightarrow w)$ is mapped by $\varepsilon_{Def}$ into the assertion $\text{ground}(x) \land \forall z (\text{var}(z) \land \text{share}(z, w) \Rightarrow \text{share}(z, y)) \land \forall z (\text{var}(z) \land \text{share}(z, y) \Rightarrow \text{share}(z, w))$.

Next, the transformation of Definition 10 is shown to be correct.

**Lemma 11.** $\varepsilon_{Def}$ is an embedding of Def into $\mathcal{L}$.

Finally, using the above Lemma we can prove that Def is a representation of $A_{Def}$.

**Theorem 12.** Def $\simeq A_{Def}$.

In order to analyze Def and to compare it with other abstract domains, a maximal factorization of $A_{Def}$ is given. To this end, we use the following domains. For every $|V| \geq n \geq 0$, consider the domain $A_{Def^n}$ consisting of the conjunctions of formulas of the form $\forall z (\text{var}(z) \land \text{share}(z, x) \Rightarrow \text{share}(z, y_1) \lor \ldots \lor \text{share}(z, y_n))$, with $y_1, \ldots, y_n$ distinct variables of $V$. The following result holds.

**Lemma 13.** $\{A_{Def^n} \mid n \in [0, |V|]\}$ is a maximal factorization of $A_{Def}$.

Let $A_{Def^+} = \land_{n \in [1, |V|]} A_{Def^n}$. A representation of $A_{Def^+}$ is provided by the set Def$^+$ of positive boolean functions that can be represented as conjunctions of clauses $y_1 \land \ldots \land y_n \rightarrow x$, with $n \geq 1$, plus the bottom element false, with concretization function the one of Def. Then by Lemma 13 it follows that Def is (isomorphic to) the reduced-product of the domain $Con$ and Def$^+$. It has been recently shown in [4] that Def characterizes the ground-dependency information on $V$ described by the domain Sharing. We shall see that this result is easily derived from the logical descriptions of these domains.

### 4.2 Pos in Logical Form

In order to study ground-dependency analysis, the abstract domain Pos was introduced by Marriott and Søndergaard [14, 15], consisting of the positive boolean functions, plus the bottom element false, with concretization function equal to $\gamma_{Def}$.

Consider the following abstract domain $A_{Pos}$. 

Definition 14. $A_{Pos}$ is the set of assertions that are conjunctions of formulas of the form \( \forall z (\text{var}(z) \land \text{share}(z, x_1) \Rightarrow Q(z, y_1, \ldots, y_n)) \lor \ldots \lor \forall z (\text{var}(z) \land \text{share}(z, x_m) \Rightarrow Q(z, y_1, \ldots, y_n)) \), with \( m \geq 1 \), and \( n \geq 0 \), where \( x_1, \ldots, x_m, y_1, \ldots, y_n \) are in \( V \), and \( z \) is a fresh variable.

We show that $Pos$ is a representation of $A_{Pos}$, and provide a maximal factorization (on $\mathcal{L}$) of $A_{Pos}$.

First, $Pos$ is characterized in logical form by means of the following transformation. We use the representation of an element $F$ of $Pos$ as a conjunction of clauses, of the form \( y_1 \land \ldots \land y_n \rightarrow x_1 \lor \ldots \lor x_m \), \( m \geq 1 \), \( n \geq 0 \) (cf. [2]). In the sequel $Q(z, y_1, \ldots, y_n)$ denotes the assertion $\text{share}(z, y_1) \lor \ldots \lor \text{share}(z, y_n)$.

Definition 15. The transformation $\epsilon_{Pos} : Pos \rightarrow \mathcal{L}$ maps $F$ into $\phi_F$, defined as follows:

1. $\phi_F = \forall z (\text{var}(z) \land \text{share}(z, x_1) \Rightarrow Q(z, y_1, \ldots, y_n)) \lor \ldots \lor \forall z (\text{var}(z) \land \text{share}(z, x_m) \Rightarrow Q(z, y_1, \ldots, y_n))$ if $F = y_1 \land \ldots \land y_n \rightarrow x_1 \lor \ldots \lor x_m$.
2. $\phi_F = \phi_{F_1} \land \ldots \land \phi_{F_k}$ if $F = F_1 \land \ldots \land F_k$, $k \geq 0$, and all the $F_i$'s are clauses.

It is easy to check that the above transformation restricted to the elements of $Def$ coincides with $\epsilon_{Def}$.

Example 7. The element $x \lor y$ is mapped by $\epsilon_{Pos}$ into the assertion $\forall z (\text{var}(z) \land \text{share}(z, x) \Rightarrow \text{false}) \lor \forall z (\text{var}(z) \land \text{share}(z, y) \Rightarrow \text{false})$, equivalent to $\text{ground}(x) \lor \text{ground}(y)$.

Next, the transformation of Definition 15 is shown to be correct.

Lemma 16. $\epsilon_{Pos}$ is an embedding of $Pos$ into $\mathcal{L}$.

Finally, using Lemma 16, we can prove that $Pos$ is a representation of $A_{Pos}$.

Theorem 17. $Pos \simeq A_{Pos}$.

In order to give a maximal factorization of $A_{Pos}$, we use the decomposition of $A_{Def}$, and the following domains. For every $|V| \geq n \geq 0$ and $|V| \geq m \geq 2$, consider the domain $A_{Pos}'$ consisting of the conjunctions of formulas of the form $\forall z (\text{var}(z) \land \text{share}(z, x_1) \Rightarrow Q(z, y_1, \ldots, y_n)) \lor \ldots \lor \forall z (\text{var}(z) \land \text{share}(z, x_m) \Rightarrow Q(z, y_1, \ldots, y_n))$ with $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ distinct variables of $V$. The following result holds.

Lemma 18. $\{A_{Def}, A_{Pos}' \mid n \in [0, |V|], m \in [2, |V|]\} \text{ is a maximal factorization of } A_{Pos}$.

Let $A_{Pos}' = \bigwedge_{n \in [0, |V|], m \in [2, |V|]} A_{Pos}'$. A representation of $A_{Pos}'$ is provided by the set $Pos'$ of positive boolean functions that can be represented as conjunctions of clauses $y_1 \land \ldots \land y_n \rightarrow x_1 \lor \ldots \lor x_m$, with $n \geq 0$, $m \geq 2$, plus the bottom element $\text{false}$, with concretization function the one of $Pos$. Then by Lemma 18 it follows that $Pos$ is (isomorphic to) the reduced-product of the domains $\text{Con}$, $\text{Def}^+$ and $Pos'$. It has been shown in [6] that $Def$ is properly contained in $Pos$. Lemma 18 characterizes logically the other part of $Pos$. 
4.3 Sharing in Logical Form

In order to study information on the possible sharing among abstract variables, an abstract domain extensively used in abstract interpretation is the domain Sharing by Jacobs and Langen [10]. Sharing is the set of sets \( \mathcal{D} \in 2^{2^{2^V}} \) s.t. if \( \mathcal{D} \neq \emptyset \) then \( \emptyset \in \mathcal{D} \). Its concretization function \( \gamma_{\text{Sharing}} \) maps an element \( \Delta \) of Sharing into the set \( \gamma_{\text{Sharing}}(\Delta) \) of those substitutions \( \sigma \) whose approximation set \( A(\sigma) \) is an element of \( \Delta \). The approximation set \( A(\sigma) \) consists of all the sets \( \text{occ}(\sigma, x) = \{ v \mid v \text{ in the domain of } \sigma \text{ s.t. } x \text{ occurs in } v \sigma \} \), for all the variables \( x \) occurring in the range of \( \sigma \).

Consider the following abstract domain \( A_{\text{Sharing}} \).

**Definition 19.** \( A_{\text{Sharing}} \) is the set of assertions of \( \mathcal{L} \) that are conjunctions of formulas of the form \( \forall z (\text{var}(z) \land \text{share}(z, x_1) \land \ldots \land \text{share}(z, x_m) \Rightarrow \text{share}(z, y_1) \lor \ldots \lor \text{share}(z, y_n)) \) with \( m \geq 1, n \geq 0 \), where \( x_1, \ldots, x_m, y_1, \ldots, y_k \) are in \( V \), and \( z \) is a fresh variable.

We show that Sharing is a representation of \( A_{\text{Sharing}} \), and provide a maximal factorization (on \( \mathcal{L} \)) of \( A_{\text{Sharing}} \).

First, Sharing is characterized in logical form by means of the following transformation. In the sequel, for the sake of simplicity, we write \( \text{share}(x, S) \) instead of \( \text{share}(\{x\} \cup S) \).

**Definition 20.** The transformation \( \varepsilon_{\text{Sharing}} \) maps \( \Delta \) into the assertion

\[
\phi_\Delta = \bigwedge_{S \subseteq V} \forall z (\text{var}(z) \land \text{share}(z, S) \Rightarrow \text{share}(z, S_1) \lor \ldots \lor \text{share}(z, S_k)),
\]

with \( \{S_1, \ldots, S_k\} = \{S' \mid S' \in \Delta \text{ s.t. } S \subseteq S'\} \).

Let \( \phi_S \) denote the conjunct of \( \phi_\Delta \) corresponding to the subset \( S \) of \( V \).

Observe that if \( S \) is not contained in any set of \( \Delta \), then \( \phi_S \) is the assertion \( \forall z (\text{var}(z) \land \text{share}(z, S) \Rightarrow \text{false}) \), which says that the variables of \( S \) can only be bound to terms sharing no variables. If \( S \) is a singleton, say \( S = \{x\} \), then \( \phi_S \) describes information on ground-dependency for \( x \). Indeed, it is not difficult to see that in this case \( \phi_S \) can be rewritten into an assertion of \( A_{\text{Def}} \). The other assertions \( \phi_S \), for \( S \) not singleton and \( k > 0 \), describe information about sharing of sets containing at least three variables.

**Example 8.** Consider \( \Delta = \{\emptyset, \{x\}, \{x, y\}, \{y, z\}\} \), and \( V = \{x, y, z\} \). Then \( \phi_\Delta \) is (equivalent to) \( \neg\text{share}(x, z) \land \neg\text{share}(\{x, y, z\}) \land \forall v (\text{var}(v) \land \text{share}(v, y) \Rightarrow \text{share}(v, z) \lor \text{share}(v, x)) \land \forall v (\text{var}(v) \land \text{share}(v, z) \Rightarrow \text{share}(v, y)) \).

Next, the correctness of this transformation is shown.

**Lemma 21.** \( \varepsilon_{\text{Sharing}} \) is an embedding of Sharing into \( \mathcal{L} \).
Finally, Lemma 21 is used to prove that Sharing is a representation of $A_{\text{Sharing}}$. In the proof, we use the fact that the assertion $\forall z (\text{var}(z) \land \text{share}(z, S)) \Rightarrow \text{share}(z, S_1) \lor \ldots \lor \text{share}(z, S_k)$ is equivalent to the assertion consisting of the conjunction of the formulas $\forall z (\text{var}(z) \land \text{share}(z, x_1) \land \ldots \land \text{share}(z, x_m) \Rightarrow \text{share}(z, y_1) \lor \ldots \lor \text{share}(z, y_k))$, for all $(y_1, \ldots, y_k)$ occurring in $S_1 \times \ldots \times S_k$.

**Theorem 22.** Sharing $\simeq A_{\text{Sharing}}$.

In order to give a maximal factorization of $A_{\text{Sharing}}$, we use the following domains. For every $|V| \geq n \geq 0$ and $|V| \geq m \geq 1$, consider the domain $A_{\text{Sharing}}^{m,n}$ consisting of the conjunctions of formulas of the form $\forall z (\text{var}(z) \land \text{share}(z, x_1) \land \ldots \land \text{share}(z, x_m) \Rightarrow \text{share}(z, y_1) \lor \ldots \lor \text{share}(z, y_n))$, with $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ distinct variables of $V$.

**Lemma 23.** $\{A_{\text{Sharing}}^{m,n} \mid n \in [0,|V|], m \in [1,|V|]\}$ is a maximal factorization of $A_{\text{Sharing}}$.

Consider the abstract domain $\text{Sharing}^+$ introduced in [4], containing as elements the empty set, and the sets $\Delta^+$ of the form $\Delta \cup T$, with $\Delta$ in $\text{Sharing}$ and $T = \{\{x\} \mid x \in V\} \cup \{\emptyset\}$. One can prove that $\text{Sharing}^+$ is a representation of $\bigwedge_{m \geq 2, n \geq 0} A_{\text{Sharing}}^{m,n}$. Moreover, $\text{Def}^+$ is a representation of $\bigwedge_{n \geq 0} A_{\text{Sharing}}^{1,n}$. Therefore, by Lemma 23 it follows that $\text{Sharing}$ is (isomorphic to) the reduced product of $\text{Sharing}^+, \text{Def}^+$ and $\text{Con}$.

### 4.4 $A_{\text{Sub}}$ in Logical Form

The pair-sharing domain $A_{\text{Sub}}$ was introduced by Søndergaard [18] for sharing and linearity analysis. Its elements are pairs $(G, R)$ where the first component is a subset of $V$, and the second one is a symmetric binary relation on $V$, s.t. $(G \times V) \cap R = \emptyset$. Moreover, the element $\bot$, representing the empty set of substitutions, is in $A_{\text{Sub}}$. Its concretization function $\gamma_{A_{\text{Sub}}}$ maps an element $(G, R)$ of $A_{\text{Sub}}$ into the set of substitutions $\sigma$ s.t. for all $(x, y)$ in $V$: (i) $x$ in $G$ implies $x \sigma$ ground; (ii) $x, y$ distinct and $\text{OVar}(x \sigma) \cap \text{OVar}(y \sigma) \neq \emptyset$ implies $(x, y)$ in $R$; (iii) $(x, x) \notin R$ implies $x \sigma$ linear.

Consider the following abstract domain $A_{A_{\text{Sub}}}$.

**Definition 24.** $A_{A_{\text{Sub}}}$ is the set of assertions that are conjunctions of literals of the form $\text{ground}(x)$, $\neg \text{share}(x, y)$, and $\text{linear}(x)$, with $x, y$ in $V$.

We show that $A_{\text{Sub}}$ is a representation of $A_{A_{\text{Sub}}}$, and provide a maximal factorization of $A_{A_{\text{Sub}}}$.

First, $A_{\text{Sub}}$ is characterized in logical form by means of the following transformation.

**Definition 25.** The transformation $\varepsilon_{A_{\text{Sub}}}$ maps $\bot$ into $\text{false}$, and $(G, R)$ into the assertion $\phi_{(G, R)} = \phi_1 \land \phi_2 \land \phi_3$, where:
1. $\phi_1$ is the conjunction of the atoms $\text{ground}(x)$, for all $x$ in $G$.
2. $\phi_2$ is the conjunction of the literals $\neg \text{share}(x, y)$, for all $(x, y)$ not in $R$ with $x, y$ distinct.
3. $\phi_3$ is the conjunction of the atoms $\text{linear}(x)$, for all $(x, x)$ not in $R$.

Assertions $\phi_1$, $\phi_2$ and $\phi_3$ characterize $A_{Sub}$ in logical form, by means of its information on groundness, independence, and linearity, respectively.

Example 9. Consider the element $(G, R)$ of $A_{Sub}$, with $G = \{x\}$ and $R = \{(y, z), (z, z), (z, w)\}$ and suppose that $V = \{x, y, z, w\}$. Then $\phi(G, R)$ is (equivalent to) $\text{ground}(x) \land \text{linear}(y) \land \text{linear}(w) \land \neg \text{share}(y, w)$.

Next, this transformation is shown to be correct.

**Lemma 26.** $\varepsilon_{A_{Sub}}$ is an embedding of $A_{Sub}$ into $L$.

Finally, Lemma 26 is used to prove that $A_{Sub}$ is a representation of $A_{ASub}$.

**Theorem 27.** $A_{Sub} \simeq A_{ASub}$.

In order to give a maximal factorization of $A_{ASub}$, the domain $A_{Linear}$ is used, consisting of the conjunctions of atoms the form $\text{linear}(x)$, with $x$ in $V$.

**Lemma 28.** $\{A_{Sharing}^{m, 0}, A_{Linear} \mid m \in [1, 2]\}$ is a maximal factorization of $A_{ASub}$.

5 Conclusion

In this paper a simple framework based on first-order logic has been proposed for reasoning about abstract domains for static analysis. The notions of domain representation and of conjunctive factorization have been introduced for analyzing and comparing abstract domains. The usefulness of this framework has been illustrated by considering a number of abstract domains used in abstract interpretation of logic programs.

We discuss now some related work.

In [7], the Galois insertion approach is used to define the notion of quotient of a domain $D$ w.r.t. another domain $P$, describing the part of $D$ that is useful for computing the information described by $P$. In this paper the logical characterization and factorization of the domains allows to perform a similar analysis, where $D$ and $P$ are first characterized logically, and next factorized. Then the reduced product of the common factors of the domains corresponds to the quotient of $D$ w.r.t. $P$.

In [4] the approach based on closure operators is used. To this end, the lattice of all the abstract domains (according with the original definition of the Cousots, cf. [3]) is considered. Abstract domains are decomposed by means of the notion of pseudo-complement, a kind of inverse of the reduced-product. Instead, in our method the set of abstract domains considered depends on the set $V$ of the
program variables, as well as on the class of properties described in the assertion language. Moreover, we use a direct approach for decomposing a domain, by inspecting the syntactic form of the relative set of assertions.

The abstract domains analysed in Section 4 have been extensively studied in previous work. In [6] it is proven that the part of Sharing describing groundness dependencies is contained in Pos. In [4] this result is strengthen by showing that this part coincides with Def, and that Sharing$^+$ is the pseudo-complement of Def in Sharing. In this paper these results are directly derived from the logical characterization of Sharing. Moreover, we have obtained the finest (in L) decomposition of Sharing. Finally, the factors of this decomposition have been used for other purposes, e.g. for comparing Sharing with ASub.

The classes of Boolean functions used to represent Def and Pos have been extensively analyzed (e.g. [5, 2]). The difference from these works is that they focus on the representation, while we focus on the design and reasoning, by considering a syntactic characterization in first-order logic of their image under the concretization function.

An interesting topic that seems worth of investigation, is the study of the relationship between abstract interpretations and proof methods. This topic has been tackled in the functional programming setting, where a domain-theoretic approach is used in [11] for proving that strictness analysis by abstract interpretation and non-standard type inference are equivalent. For logic programming, our framework could be used for defining a program logic for the comparison of data-drivenness analysis using type inference (cf. e.g. [1]) and abstract interpretation (cf. [16]).

Acknowledgements

This work was partially supported by NWO, the Dutch Organization for Scientific Research, under grant 612-32-001. I would like to thank the referees for their useful comments; Livio Colussi, Tino Cortesi, Gilberto Filé, Maurizio Gabbrielli, Massimo Marchiori, and Catuscia Palamidessi for stimulating discussions on the subject of this paper; and Jan Rutten for his encouragement.

References


