A minimalistic look at widening operators

David Monniaux

CNRS / VERIMAG[*]

December 8, 2009

Abstract

We consider the problem of formalizing in higher-order logic the familiar notion of widening from abstract interpretation. It turns out that many axioms of widening (e.g. widening sequences are ascending) are not useful for proving correctness. After keeping only useful axioms, we give an equivalent characterization of widening as a lazily constructed well-founded tree. In type systems supporting dependent products and sums, this tree can be made to reflect the condition of correct termination of the widening sequence.

1   The usual framework

We shall first recall the usual definitions of abstract interpretation and widening operators.

1.1   Abstraction and concretization maps

Abstract interpretation is a framework for formalizing approximation relationships arising in program semantics and static analysis [8, 9]. *Soundness* of the abstraction is expressed by the fact that the approximation takes place in a controlled direction. In order to prove that a given set of undesirable states is unreachable, we can compute a superset of the set of reachable states (an *over-approximation* thereof), in the hope that this set does not intersect the set of undesirable states. If order to prove that we eventually reach a given set of states, we can compute a subset of the set of states that eventually reach them (an *under-approximation* thereof), in the hope that this set includes the initial states.

Most introductory materials on abstract interpretation describe abstraction as a *Galois connection* between a concrete space S (typically, the powerset $\mathcal{P}(\Sigma)$ of the set of states Σ of the program, or the powerset of the set of finite execution traces Σ^* of the program) and an abstract space $S^\sharp$. For instance, if the program state consists in a program counter location, taken within a finite set P of program locations, and three integer variables, $\Sigma = P \times \mathbb{Z}^3$, $S = \mathcal{P}(P \times \mathbb{Z}^3)$, the abstract state can be, for instance, a member of $S^\sharp = P \to (\{\bot\} \cup I^3)$, where P is the set of program locations, $a \to b$ denotes the set of functions from a

---

[*]VERIMAG is a joint laboratory of CNRS, Université Joseph Fourier and Grenoble-INP.

to b, I is the set of well-formed pairs (a, b) defining intervals ($a \in \mathbb{Z} \cup \{-\infty\}$, $b \in \mathbb{Z} \cup \{+\infty\}$ and $a \le b$) and $\perp$ is a special element meaning "unreachable". S and $S^\sharp$ are ordered; here, S is ordered by set inclusion $\subseteq$ and $S^\sharp$ is ordered by $\sqsubseteq_P$, the pointwise application of $\sqsubseteq$ for all program locations: $\perp \sqsubseteq x^\sharp$ for all x in $S^\sharp$, and $((a_1, b_1), (a_2, b_2), (a_3, b_3)) \sqsubseteq ((a'_1, b'_1), (a'_2, b'_2), (a'_3, b'_3))$ if for all $1 \le i \le 3$, $a'_i \le a_i$ and $b_i \le b'_i$. For the sake of simplicity, we shall give examples further on where P is a singleton; the generalization to any finite P is straightforward. $P \to (\{\perp\} \cup I^3)$ is then isomorphic to $\{\perp\} \cup I^3$ and we shall thus consider, as a running example, the case where S is $\mathcal{P}(\mathbb{Z}^3)$ and $S^\sharp$ is $\{\perp\} \cup I^3$.

S and $S^\sharp$ are connected by an *abstraction map* α and a *concretization map* γ. γ maps any abstract state $x^\sharp$ to the set of concrete states that it represents. Here, $\gamma((a_1, b_1), (a_2, b_2), (a_3, b_3))$ is the set of triples (v_1, v_2, v_3) such that for all $1 \le i \le 3$, $a_i \le v_i \le b_i$. α maps a set x of concrete states to the "best" (least) abstract element $x^\sharp$ such that $x \subseteq \gamma(x^\sharp)$. Here, if $x \subseteq \mathbb{Z}^3$, then for all $1 \le i \le 3$, $a_i = \inf_{(v_1, v_2, v_3) \in x} v_i$ and $b_i = \sup_{(v_1, v_2, v_3) \in x} v_i$. γ must be monotone with respect to $\subseteq$ and $\sqsubseteq$: if $x^\sharp \sqsubseteq y^\sharp$, then $\gamma(x^\sharp) \subseteq \gamma(y^\sharp)$.

In some presentations of abstract interpretation, abstract elements $x^\sharp$ are identified with their concretization $\gamma(x^\sharp)$. For instance, one talks directly of the interval $[a, b]$, not of the pair (a, b). This can make explanations smoother by clearing up notations. It is however important for some purposes to distinguish the machine representation of an abstract element $x^\sharp$ from its concretization $\gamma(x^\sharp)$, if only because γ may not be injective. For instance, $x = y \wedge x \le 1$ and $x = y \wedge y \le 1$ define exactly the same part of the plane (as geometrical convex polyhedra) but are different in their machine representation. This is the same difference as that between the *syntax* and the *semantics* of a logic.

In this article, we ditinguish this syntactic and semantical aspects, for several reasons. First, certain abstract operations may be sensitive to the syntax of an abstract element; that is, they may yield different results for $x^\sharp$ and $y^\sharp$ even though $\gamma(x^\sharp) = \gamma(y^\sharp)$, as we shall recall in §1.3 about the polyhedra and octagons.

Also, while in many cases $\sqsubseteq$ is defined by $a \sqsubseteq b \iff \gamma(a) \subseteq \gamma(b)$, this relation may sometimes be too costly or impossible to compute, and some smaller relation may be used. For instance, if one uses a product of several abstract domains $D_1^\sharp \times \ldots \times D_m^\sharp$, each D_i fitted with a decidable ordering $\sqsubseteq_i$, and $\gamma(x_1^\sharp, \ldots, x_m^\sharp) = \gamma_1(x_1^\sharp) \cap \ldots \cap \gamma_m(x_m^\sharp)$ then it is straightforward to consider the product ordering $(x_1, \ldots, x_m) \sqsubseteq (x'_1, \ldots, x'_m) \iff x_1 \sqsubseteq_i x'_1 \wedge \ldots \wedge x_m \sqsubseteq_m x'_m$. If $x \sqsubseteq x'$ for this ordering, then $\gamma(x) \subseteq \gamma(x')$, but the two are not necessarily equivalent. Consider for instance a simplification of the domain of difference bounds [15], expressed as a product of simpler domains: the concrete states in $\mathbb{Q}^3$, the abstract domains $D_1^\sharp = D_2^\sharp = D_3^\sharp = \mathbb{Q}$, $\gamma_1(c_1) = \{(x, y, z) \in \mathbb{Q}^3 \mid x - y \le c_1\}$, $\gamma_2(c_2) = \{(x, y, z) \in \mathbb{Q}^3 \mid y - z \le c_2\}$, $\gamma_3(c_3) = \{(x, y, z) \in \mathbb{Q}^3 \mid x - z \le c_3\}$. Obviously, $\gamma(1, 1, 2) = \gamma(1, 1, 3)$, yet $(1, 1, 3) \not\sqsubseteq (1, 1, 2)$. In order to use the product ordering, one has to perform beforehands a reduction operation mapping $(1, 1, 3$ to $(1, 1, 2)$, but such an operation may be nontrivial: the one in the octagon abstract domain involves a Floyd-Warshall shortest path computation, the one in the template linear constraints [20] involves linear programming. In the case of real-life static analysis tools, e.g. the Astrée static analyzer [4], with many nontrivial abstract domains interacting, it is not obvious whether

$\gamma(a) \subseteq \gamma(b)$ is decidable, and even if it were, how to decide it within acceptable time.

Finally, since our goal is to write programs and proofs in a proof assistant based on intuitionistic type theory, we thought it best to clearly separate the computational, constructive content from the non-computational content: membership in the set of reachable states of a program is, in general, recursively enumerable but not recursive (from Turing's halting problem: one cannot in general decide whether the "end" line of the program is reachable); thus the characteristic function of that set cannot be defined by constructive logic, since this would involve describing an algorithm computing that function.

1.2 Obtaining invariants

Abstract interpretation replaces a possibly infinite number of concrete program execution, which cannot be simulated in practice, by a simpler "abstract" execution. For instance, one may replace running a program using our three integer variables over all possible initial states by a single abstract execution with interval arithmetic. The resulting final intervals are guaranteed to contain all possible outcomes of the concrete program. More formally, if one has a transition relation $\tau \subseteq \Sigma \times \Sigma$, one defines the forward concrete transfer function $f_\tau : S \to S$ as $f_\tau(x) = \{\sigma' \mid \sigma \to_\tau \sigma' \land \sigma \in x\}$. $f_\tau(W)$ is the set of states reachable in one forward step from W. We say that the *abstract transfer function* $f_\tau^\sharp(x^\sharp)$ is a correct abstraction for f_τ if for all $x^\sharp$, $f_\tau \circ \gamma(x^\sharp) \subseteq \gamma \circ f_\tau^\sharp(x^\sharp)$. This *soundness property* means that if we have a superset of the concrete set of states before the execution of τ, we get a superset of the concrete set of states after the execution of τ.

As usual in program analysis, obtaining loop invariants is the hardest part. Given a set $x_0 \subseteq \Sigma$ of initial states, we would like to obtain a superset of the set of reachable states $x_\infty = \{\sigma' \mid \sigma \to_\tau^* \sigma' \land \sigma \in x_0\}$. The set of states x_n reachable in at most n steps from x_0 is defined by induction: $x_{n+1} = \phi(x_n)$, where $\phi(x) = f_\tau(x) \cup x_0$ is monotone, because f_τ is by definition a $\cup$-morphism. The sequence (x_n) is ascending, and its limit is x_∞, which is the least fixed point of ϕ by Kleene's fixed point theorem; this sequence is thus often known as *Kleene iterations*. x_∞ is also known as the *strongest invariant* of the program. An *inductive invariant* or *post-fixpoint* is a set x such that $x_0 \subseteq x$ and $f_\tau(x) \subseteq x$, and by Tarski's theorem, the intersection of all such sets is x_∞.

Obviously, the set of all possible states (often noted $\top$) is an inductive invariant, but it is uninteresting since it cannot be used to prove any non-trivial property of the program. A major goal of program analysis is to obtain program invariants x that are strong enough to prove interesting properties, without being too costly to establish.

In some cases, interesting inductive invariants may be computed directly. Various approaches have recently been proposed for the direct computation of invariants, without Kleene iterations. Costan et al. [6] proposed a method for computing least fixed points in the lattice of real intervals by downward *policy iteration*, also known as *strategy iteration*, a technique borrowed from game theory; they later extended their framework to other domains. Gawlitza and Seidl [12] proposed a method for computing least fixed points in certain lattices by upward strategy iteration. Monniaux [17, 18] showed that least fixed point problems in some lattices expressing numerical constraints can be reduced to

quantifier elimination problems, which in turn can be solved algorithmically. Other recent proposals include expressing the least invariant problem in the abstract lattice directly as a constrained minimization problem, then solving it with operational research tools [7]. One common factor to these approaches is that they target specific classes of abstract domains and programs; in addition, they may also suffer from high complexity.

1.3 Abstract Kleene iterations and widening operators

The more traditional approach to finding inductive invariants by abstract interpretation is to perform *abstract Kleene iterations*. Let $x_0^\sharp$ be an abstraction of x_0. Define $\phi^\sharp(x^\sharp) = f_\tau^\sharp(x^\sharp) \sqcup x_0^\sharp$, where $\sqcup$ is a sound overapproximation of the concrete union $\cup$: $\gamma(x^\sharp) \cup \gamma(y^\sharp) \subseteq \gamma(x^\sharp \sqcup y^\sharp)$. From the soundness of $f_\tau^\sharp$ and $\sqcup$, $\phi^\sharp$ is a sound abstraction of ϕ: for all $x^\sharp$, $\phi \circ \gamma(x^\sharp) \subseteq \gamma \circ \phi^\sharp(x^\sharp)$. By induction, for all n, $x_n \subseteq \gamma(x_n^\sharp)$: assuming $x_n \subseteq \phi(x_n^\sharp)$, $x_{n+1} = \phi(x_n) \subseteq \phi \circ \gamma(x_n^\sharp) \subseteq \gamma \circ \phi^\sharp(x_n^\sharp) = x_{n+1}^\sharp$.

In many presentations of abstract interpretation, it is supposed that the abstract transfer function $f_\tau^\sharp$ and the abstract union $\sqcup$ are monotonic. Intuitively, this means that if the analysis has more precise information at its disposal, then its outcome is more precise. This is true for elementary transfer functions in most abstract domains, and thus of their composition into abstract transfer functions of more complex program constructions. A well-known exception is when the abstract transfer function is itself defined as the overapproximation of a least fixed-point operation using a widening operator (see below), yet there exist less well-known cases where the abstract transfer function may be non-monotonic.[1]

Let us nevertheless temporarily assume that $f_\tau^\sharp$ and $\sqcup$ and, thus, $\phi^\sharp$, are monotonic, and that $a^\sharp, b^\sharp \sqsubseteq a^\sharp \sqcup b^\sharp$ for all $a^\sharp$ and $b^\sharp$. Then $x_0^\sharp \sqsubseteq x_1^\sharp$ and by induction, for all n, $\phi^\sharp$ being monotonic, $x_n^\sharp = \phi^{\sharp n}(x_0^\sharp) \sqsubseteq \phi^{\sharp n}(x_1^\sharp) = x_{n+1}^\sharp$; the sequence $x_n^\sharp$ is therefore ascending. If this sequence is stationary, there is a N such that $x_{N+1}^\sharp = x_N^\sharp$. Then, $\gamma(x_N^\sharp) = \gamma(x_{N+1}^\sharp) = \gamma(f_\tau^\sharp(x_N^\sharp) \sqcup x_0^\sharp) \supseteq \gamma \circ f_\tau^\sharp(x_N^\sharp) \supseteq f_\tau \circ \gamma(x_N^\sharp)$, and $\gamma(x_N^\sharp) = \gamma(x_{N+1}^\sharp) = \gamma(f_\tau^\sharp(x_N^\sharp) \sqcup x_0^\sharp) \supseteq \gamma(x_0^\sharp)$, which means that $\gamma(x_N^\sharp)$ is an inductive invariant. Obviously, if the abstract domain $S^\sharp$ is finite, then any ascending sequence is stationary.[2]

More generally, the same results hold for any domain of *finite height* (there exists an integer L such that any strictly ascending sequence has at most length

---

[1]Such is for instance the case of the symbolic constant propagation domain proposed by Miné [16, §5][15, §6.3.4]. The full symbolic propagation strategy can induce non-monotonic effects: if the analysis knows more relationships, it can perform spurious rewritings and paradoxically provide a less precise result.

The same is true of Miné's linearization step, which dynamically abstracts nonlinear expressions as linear expressions. Consider the nonlinear expression $x \times y$ where $x \in [m_x, M_x]$, $y \in [m_y, M_y]$ and $m_x, m_y > 0$: a choice has to be made between several valid linearizations, here $x \times [m_y, M_y]$ and $[m_x, M_x] \times y$. While all choices between candidate linearizations lead to sound results, they do not have the same precision and the choice heuristic does not necessarily choose the one leading to the most precise results later on.

[2]This explains the popularity of Boolean abstractions: $S^\sharp$ is the set of sets of bit vectors of fixed length L, and these sets are often represented by *reduced ordered binary decision diagrams* (ROBDD) [5]. Reachability analysis in BDD-based model-checkers is thus a form of Kleene iteration in the BDD space. Very astute implementation techniques, involving generalized hashing of data structures, ensure that equality tests take constant time and that $\phi^\sharp$ is computed efficiently.

L), and, even more generally, for any domain satisfying the *ascending chain condition* (there does not exist any infinite strictly ascending sequence). Yet, even the very simple domain of products of intervals that we defined earlier does not satisfy the ascending chain condition!

In domains that do not satisfy the ascending condition, the abstract Kleene iterations may fail to converge in finite time. Such is the case, for instance, of the interval abstraction of the program with a single integer variable defined by the transition system τ: for all n, $n \to_\tau n + 1$, and the initial state is 0. The best abstract transfer function $\phi^\sharp$ maps a pair $(0, n)$ representing an integer interval $\{0, \ldots, n\}$ to the pair $(0, n+1)$, thus the abstract Kleene iterations are $x_n^\sharp = (0, n)$ and the analysis fails to converge in finite time.

The traditional solution to the convergence problem in domains that do not satisfy the ascending chain condition is to use a *widening operator*, which is a form of convergence accelerator applied to abstract Kleene iterations [8, Def. 4.1.2.0.4][9, §4]. Intuitively, the widening operation examines the first abstract Kleene iterations and conjectures some possible over-approximation of the limit, which is then checked for stability; further iterations may be necessary until an inductive invariant is reached. For each infinite height domain, one or more widening operators must be designed. Consequently, most literature on abstract interpretation domains includes descriptions of widening operators.

For instance, the interval abstract domain can be fitted with a simple widening discarding unstable bounds [8], then later with the less brutal "widening up to" [13, §3.2] or "widening with thresholds"[3, §6.4][4, §7.1.2]. The domain of convex polyhedra was first fitted with a very simple widening that discarded all unstable constraints [10], but this widening was later refined in order to make it insensitive to syntactic variations in the way semantically equivalent constraints were given [14, p. 56–57][13, §2.2]. Miné [15] fitted the octagon abstract domain with a similar construction, widening to $+\infty$ the unstable constraints. Again, this widening was sensitive to syntax, which lead to proposals of semantic widenings [1]. Widening techniques are not restricted to numerical domains; for instance there are specific techniques for widening over automata [11] (roughly speaking, they overapproximate a language defined by an automaton by the language defined by a quotient, of limited size, of that automaton; the limited size ensures termination).

Here is the most common definition:

Definition 1. A widening operator ∇ on an abstract domain $(S^\sharp, \sqsubseteq)$ is a binary operator that satisfies the three following properties:

1. $x^\sharp \sqsubseteq x^\sharp \nabla y^\sharp$

2. $y^\sharp \sqsubseteq x^\sharp \nabla y^\sharp$

3. for any sequence $v_n^\sharp$, a sequence of the form $u_{n+1}^\sharp = u_n^\sharp \nabla v_n^\sharp$ is ultimately stationary.

We can then use $u_0^\sharp = x_0^\sharp$, $u_{n+1}^\sharp = u_n^\sharp \nabla \phi^\sharp(u_n^\sharp)$. By the third property of the widening operator, there exists N such that $u_N^\sharp = u_N^\sharp \nabla \phi^\sharp(u_N^\sharp)$. Thus, $\phi^\sharp(u_N^\sharp) \sqsubseteq u_N^\sharp$, and $\gamma \circ \phi^\sharp(u_N^\sharp) \subseteq \gamma(u_N^\sharp)$. But $x_0 \cup f_\tau \circ \gamma(u_N^\sharp) = \phi \circ \gamma(u_N^\sharp) \subseteq \gamma \circ \phi^\sharp(u_N^\sharp) \subseteq \gamma(u_N^\sharp)$ thus $f_\tau \circ \gamma(u_N^\sharp) \subseteq \gamma(u_N^\sharp)$ and $\gamma(u_N^\sharp)$ is an inductive invariant.

Let us now have a second look at the hypotheses that we used to establish that result. Though it is often assumed that the abstract domain is a complete
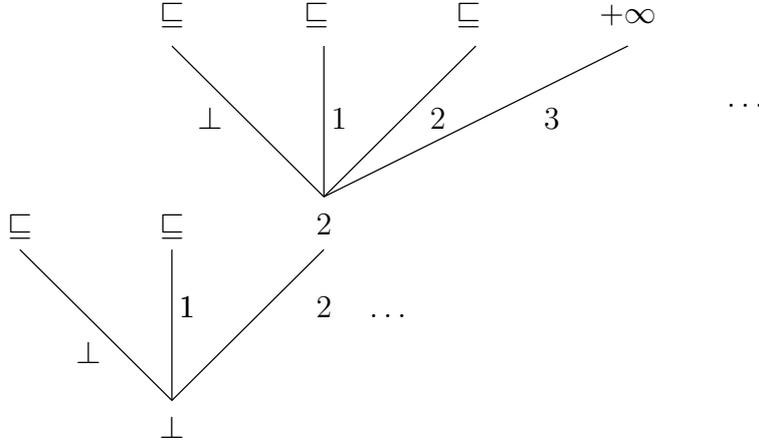
Figure 1: Interpretation of widening as a well-founded tree for the domain $1 \sqsubset 2 \sqsubset 3 \sqsubset \ldots + \infty$. This domain may be used to construct the domain of intervals: an interval $[x, y]$ is represented by the pair $(-x, y) \in \mathbb{N}^2$, pointwise ordered, and the widening operation described here is applied to each coordinate. Each node represents a proposal $u_n^\sharp$ from the widening system. Each edge is labelled with the answer $v_n^\sharp$ from the analysis system. The widening system either answers $\sqsubseteq$ when it determines that $v_n^\sharp \sqsubseteq u_n^\sharp$, or makes a new proposal. A proposal of $+\infty$ forces termination: whatever $u_n^\sharp$ the analysis system then supplies, $u_n^\sharp \sqsubseteq +\infty$ (we left out its outgoing branches, all finishing in $\sqsubseteq$). A path from the root of the tree is an abstract Kleene iteration sequence. The well-foundedness of the tree ensures the termination of such sequences.

lattice, and that the abstract transfer function is monotonic, we never used either hypotheses. In fact, the only hypotheses that we used are:

- f_τ is monotonic and the concrete domain $\mathcal{P}(S)$ is a complete lattice, thus ϕ has a least fixed point which is the least inductive invariant of the program.

- For all $a^\sharp$ and $b^\sharp$, $b^\sharp \sqsubseteq a^\sharp \triangledown b^\sharp$.

- For all sequence $v_n^\sharp$, any sequence defined by $u_{n+1}^\sharp = u_n^\sharp \triangledown v_n^\sharp$ is stationary.

2 Relaxation of conditions and interpretation in inductive types

During our work on the Astrée tool [4], and when formalizing the notion of widening in the Coq proof assistant [2],[3] we realized that the usual definitions of abstract domains and widenings are unnecessarily restrictive for practical purposes. Pichardie [19, §4.4] already proposed a relaxation of these conditions, but his definition of widenings is still fairly complex. We propose here a drastically reduced informal definition of widenings, which subsumes both the $\sqsubseteq$ ordering and the $\triangledown$ operator; this definition will be made formal as Def. 3.

---

[3]Coq is a proof assistant based on higher order logic, available from `http://coq.inria.fr`.

Definition 2. A widening system is an algorithm that proposes successive abstract elements $u_0^\sharp, u_1^\sharp, \ldots, u_n^\sharp$ to the rest of the analyzer, and receives $v_n^\sharp$ from it. It can then either terminate with some guarantee that $\gamma(v_n^\sharp) \subseteq \gamma(u_n^\sharp)$, or propose the next element $u_{n+1}^\sharp$. The system never provides infinite sequences.

In practical use, $v_n^\sharp = \phi^\sharp(u_n^\sharp)$ and $\phi^\sharp$ is an abstraction of the concrete transformer ϕ of a loop or, more generally, of a monotonic system of semantic equations.

It is obvious that any widening that verifies the conditions of Def. 1 also verifies these conditions. Note that Def. 2 is strictly laxer than Def. 1. For instance, we make no requirement that $\gamma(u_n^\sharp) \subseteq \gamma(u_{n+1}^\sharp)$; a widening system could first try some ascending sequence $u_0^\sharp, \ldots, u_n^\sharp$, regret, and restart with another sequence $u_{n+1}^\sharp, \ldots$.

A more mathematical way of seeing this definition is by interpreting the widening system as a well-founded tree:

Definition 3. Let $S^\sharp$ be an abstract domain with the associated concretization map γ. Let $\sqsubseteq$ be a preorder over $S^\sharp$ such that γ is monotonic. A widening system is a well-founded tree whose nodes are labeled by elements of $S^\sharp$ (there may be several nodes with the same label). From a node labeled with $u^\sharp$, there are branches labeled with every $v^\sharp$ such that $v^\sharp \not\sqsubseteq u^\sharp$.

Let $u_0^\sharp$ be the label for the root of the tree, and let $u_0^\sharp, v_0^\sharp, u_2^\sharp, \ldots$ be a path into the tree consisting in successive nodes and edges. Because the tree is well-founded, this path is finite, which means that it terminates with $u_N^\sharp, v_N^\sharp$ such that $v_N^\sharp \sqsubseteq u_N^\sharp$. This recalls the termination property of Def. 1.

Definition 3, combined with the $\sqsubseteq$ test can be easily recast as couple of mutually inductive types :

$$\begin{aligned} widening & \equiv S^\sharp \times (S^\sharp \to answer) \\ answer & \equiv termination \mid next \text{ of } widening \end{aligned} \qquad (1)$$

From each node labeled by $u^\sharp$, for each $v^\sharp$ there is an edge labeled by $v^\sharp$, which either leads to "termination" if $v^\sharp \sqsubseteq u^\sharp$, or to another node (see Fig. 1).

Note that, even in an eager language such as Objective Caml, the widening tree is never constructed in memory: its nodes are constructed on demand by application of the function $S^\sharp \to answer$.

In a higher-order type system with dependent sums and products such as the Calculus of inductive constructions (as in Coq), the above inductive datatype can be adorned with proof terms. A tree node *widening* is a pair $(u^\sharp, a)$ where a maps each $v^\sharp$ to an *answer*. $a(v^\sharp)$ is either "$\sqsubseteq$", carrying a proof term stating that $\gamma(v^\sharp) \subseteq \gamma(u^\sharp)$, or another *widening* tree node.

3 Implementation in Coq

We shall first show how to implement our concept of widening system in Coq, then we shall give a few concrete examples of how common abstract interpretation techniques can be implemented within this framework.[4]

---

[4]Source code may be downloaded from
`http://www-verimag.imag.fr/~monniaux/download/domains_coq.zip`.

3.1 Framework

We assume we have an abstract domain `S` with an ordering `domain_le` (representing $\sqsubseteq$). In practice, this ordering is supposed to be decidable: there exists a function `domain_le_decide` that takes x and y as inputs and decides whether $x \sqsubseteq y$.

The *answer* is the disjunctive sum `{domain_le y x} + widening`: it either provides a new `widening` object, or a proof that $y \sqsubseteq x$. By inlining this type into the definition of `widening`, we obtain:

```
Variable S : Set.
Hypothesis domain_le : S -> S -> Prop.
Hypothesis domain_le_decide :
    forall x y : S,
      { domain_le x y } + {~ (domain_le x y) }.


Inductive widening: Set :=
  widening_intro : forall x : S,
    (forall y : S, widening + {domain_le y x}) -> widening.
```

Note that all properties desired of the widening are lumped in this definition. The `Inductive` keyword introduces a type whose elements are all well-founded by construction; Coq will make it impossible to create widening trees that are not well-founded. The correct termination property (termination only if $v^\sharp \sqsubseteq u^\sharp$) is also ensured by construction: a leaf edge corresponding to $u^\sharp$ and $v^\sharp$ may be constructed only by giving a proof of $v^\sharp \sqsubseteq u^\sharp$ (a term belonging to the type `domain_le` $v^\sharp\ u^\sharp$).

In the above definition, we have added the hypothesis that $\sqsubseteq$ is decidable (`domain_le_decide`). This is not needed for this definition, but is useful in many constructions, and is a very reasonable assumption to make. Indeed, the reason why we introduced $\sqsubseteq$ as just any order such that γ is monotonic, and not the most precise one, is that the most precise one might not be decidable, or too costly to decide effectively.

Since `widening` is an inductive type, defining well-founded trees, it is possible to define functions by induction over elements of that type. One especially interesting inductively defined function takes $f^\sharp : S^\sharp \to S^\sharp$ as a parameter and computes $x^\sharp$ such that $f^\sharp(x^\sharp) \sqsubseteq x^\sharp$ by well-founded induction over the widening tree. On a widening node labeled by $u^\sharp$, it computes $v^\sharp = f^\sharp(u^\sharp)$ then requests the "answer" from the widening node on the value $v^\sharp$:

- Either it answers with another widening node and the function is called recursively.

- Or it answers with a proof that $v^\sharp \sqsubseteq u^\sharp$ and the algorithm terminates with the requested answer (both $u^\sharp$ and a proof that $f^\sharp(u^\sharp) \sqsubseteq u^\sharp$).

```
Section Recursor.
Variable f : S -> S.

Fixpoint abstract_lfp_rec
  (iteration_step : widening) :
  { lfp : S | domain_le (f lfp) lfp } :=
```

8

```
  let (x, xNext) := iteration_step in
  match xNext (f x) with
  | inleft next_widening => abstract_lfp_rec next_widening
  | inright fx_less_than_x => exist (fun x => domain_le (f x) x)
      x fx_less_than_x
  end.
```

```
End Recursor.
```

For ease of use, we pack S, domain_le, an abstraction relation domain_abstracts and other related constructs into one single domain record.

3.2 Examples

In numerical abstract domains, it is common to use "widening up to" [13, §3.2] or "widening with thresholds" [3, §6.4][4, §7.1.2]: one keeps an ascending sequence $z_1^\sharp, \ldots, z_n^\sharp$ of "magical" values, and $x^\sharp \triangledown y^\sharp$ is the least element $z_k^\sharp$ greater than $x^\sharp \sqcup y^\sharp$. For instance, instead of widening a sequence of integer intervals $[0, 1]$, $[0, 2]$ etc. to $[0, +\infty[$, we may try some "magical" values such as $[0, 255]$, $[0, 32767]$ etc. Yet, if all elements in the sequence fail to define an inductive invariant, we are forced to try $[0, +\infty[$. In other words, after trying the "magical" values, we revert to the usual brutal widening on the intervals.

This is easily modeled within our framework by a "widening transformer": taking a widening W as input and a finite "ramp" l of values, it outputs a widening W' that first applies the thresholds and, as a last resort, calls W. Variable T : domain is a parameter defining the original domain and original widening, which is used as the last resort by our transformed widening. Function ramp_widening_search searches for the next threshold in the "ramp".

```
Section Widening_ramp.
Variable T : domain.

Fixpoint ramp_widening_search (bound : (domain_set T))
  (ramp : (list (domain_set T))) { struct ramp } : (list (domain_set T)) :=
  match ramp with
  | nil => ramp
  | (cons ramp_h ramp_t) =>
    match (domain_le_decide T bound ramp_h) with
    | left _ => ramp
    | right _ => ramp_widening_search bound ramp_t
    end
  end.

Fixpoint ramp_widening (ramp : (list (domain_set T))) :
  (widening (domain_set T) (domain_le T)) :=
  match ramp with
  | nil => domain_widening T
  | (cons ramp_h ramp_t) =>
    (widening_intro (domain_set T) (domain_le T) ramp_h
      (fun (y : (domain_set T)) =>
```

```
        match domain_le_decide T y ramp_h with
        | left STOP =>
            inright
              (widening (domain_set T) (domain_le T)) STOP
        | right _ =>
            inleft
              (domain_le T y ramp_h)
              (ramp_widening (ramp_widening_search y ramp_t))
        end))
    end.
```

A trick often used in static analysis is to delay the widening [4, §7.1.3]. Instead of performing ∇ at each iteration, one performs $\sqcup$ for a finite number of steps, then one tries ∇ again. For termination purposes, it suffices that there is some "fairness property": ∇ should not be delayed infinitely. One can for instance choose to delay widening by n steps of $\sqcup$ after each widening step. This is again implemented as a "widening transformer":

```
Definition delayed_widening_each_step :
  nat -> (widening (domain_set T) (domain_le T)).
```

We can similarly build a product domain $S_1^\sharp \times S_2^\sharp$. The widening on couples $(a_1, a_2)\nabla(b_1, b_2) = (a_1 \nabla_1 b_1, a_2 \nabla_2 b_2)$ is implemented by a "widening transformer" taking one widening W_1 on $S_1^\sharp$ and a widening W_2 on $S_2^\sharp$ as inputs, and producing a widening on $S_1^\sharp \times S_2^\sharp$ by syntactic induction on W_1 and W_2: if $a_1 \sqsubseteq_1 b_1 \wedge a_2 \sqsubseteq_2 b_2$, then $(a_1, a_2) \sqsubseteq (b_1, b_2)$ for the product ordering and one terminates; if $a_1 \sqsubseteq_1 b_1$ but $a_2 \not\sqsubseteq_2 b_2$ then one stays on a_1 but moves one step into W_2 (and *mutatis mutandis* reversing the coordinates); if $a_1 \not\sqsubseteq_1 b_1$ and $a_2 \not\sqsubseteq_2 b_2$, then one moves into both W_1 and W_2. This implements the usual widening on products. This construct can be generalized to any finite products of domains.

4   Conclusion

By seeing the combination of the computational ordering $\sqsubseteq$ and the widening operator ∇ as a single inductive construct, one obtains an elegant characterization extending the usual notion of widening in abstract interpretation, suitable for implementation in higher order logic.

Acknowledgments

References

[1] Roberto Bagnara, Patricia M. Hill, Elena Mazzi, and Enea Zaffanella. Widening operators for weakly-relational numeric abstractions. In Chris

Hankin, editor, *Static Analysis (SAS)*, volume 3672 of *LNCS*, pages 3–18. Springer, 2005. ISBN 3-540-28584-9. doi: 10.1007/11547662_3.

[2] Yves Bertot and Pierre Castéran. *Interactive Theorem Proving and Program Development, Coq'Art: The Calculus of Inductive Constructions.* Texts in Theoretical Computer Science. Springer, 2004. ISBN 3-540-20854-2.

[3] Bruno Blanchet, Patrick Cousot, Radhia Cousot, Jérôme Feret, Laurent Mauborgne, Antoine Miné, David Monniaux, and Xavier Rival. Design and implementation of a special-purpose static program analyzer for safety-critical real-time embedded software. In Torben Æ. Mogensen, David A. Schmidt, and I. Hal Sudborough, editors, *The Essence of Computation: Complexity, Analysis, Transformation*, number 2566 in LNCS, pages 85–108. Springer, 2002. ISBN 3-540-00326-6. doi: 10.1007/3-540-36377-7_5.

[4] Bruno Blanchet, Patrick Cousot, Radhia Cousot, Jérôme Feret, Laurent Mauborgne, Antoine Miné, David Monniaux, and Xavier Rival. A static analyzer for large safety-critical software. In *Programming Language Design and Implementation (PLDI)*, pages 196–207. ACM, 2003. ISBN 1-58113-662-5. doi: 10.1145/781131.781153.

[5] Edmund M. Clarke, Jr, Orna Grumberg, and Doron A. Peled. *Model Checking.* MIT Press, 1999. ISBN 0-262-03270-8.

[6] Alexandru Costan, Stephane Gaubert, Éric Goubault, Matthieu Martel, and Sylvie Putot. A policy iteration algorithm for computing fixed points in static analysis of programs. In Kousha Etessami and Sriram K. Rajamani, editors, *Computer Aided Verification (CAV)*, number 4590 in LNCS, pages 462–475. Springer, 2005. ISBN 3-540-27231-3. doi: 10.1007/11513988_46.

[7] Patrick Cousot. Proving program invariance and termination by parametric abstraction, Lagrangian relaxation and semidefinite programming. In *Sixth International Conference on Verification, Model Checking and Abstract Interpretation (VMCAI'05)*, pages 1–24. Springer, January 17–19 2005. ISBN 3-540-24297-X. doi: 10.1007/b105073. URL `http://www.di.ens.fr/~cousot/COUSOTpapers/VMCAI05.shtml`.

[8] Patrick Cousot. *Méthodes itératives de construction et d'approximation de points fixes d'opérateurs monotones sur un treillis, analyse sémantique des programmes.* State doctorate thesis, Université scientifique et médicale de Grenoble & Institut national polytechnique de Grenoble, 1978. URL `http://tel.archives-ouvertes.fr/tel-00288657/en/`. In French.

[9] Patrick Cousot and Radhia Cousot. Abstract interpretation frameworks. *J. of Logic and Computation*, pages 511–547, August 1992. ISSN 0955-792X. doi: 10.1093/logcom/2.4.511.

[10] Patrick Cousot and Nicolas Halbwachs. Automatic discovery of linear restraints among variables of a program. In *Principles of Programming Languages (POPL)*, pages 84–96. ACM, 1978. doi: 10.1145/512760.512770.

[11] Vijay D'Silva. Widening for automata. Diplomarbeit, Universität Zürich, 2006.

[12] Thomas Gawlitza and Helmut Seidl. Precise fixpoint computation through strategy iteration. In Rocco de Nicola, editor, *Programming Languages and Systems (ESOP)*, volume 4421 of *LNCS*, pages 300–315. Springer, 2007. ISBN 978-3-540-71316-6. doi: 10.1007/978-3-540-71316-6_21.

[13] Nicolas Halbwachs. Delay analysis in synchronous programs. In *Computer Aided Verification (CAV)*, pages 333–346. Springer, 1993. ISBN 3-540-56922-7. doi: 10.1007/3-540-56922-7_28.

[14] Nicolas Halbwachs. *Détermination automatique de relations linéaires vérifiées par les variables d'un programme*. PhD thesis, Université scientifique et médicale de Grenoble & Institut national polytechnique de Grenoble, 1979. URL http://tel.archives-ouvertes.fr/tel-00288805/en/. In French.

[15] Antoine Miné. *Weakly Relational Numerical Abstract Domains*. PhD thesis, École polytechnique, Palaiseau, France, December 2004. In English.

[16] Antoine Miné. Symbolic methods to enhance the precision of numerical abstract domains. In *Verification, Model Checking, and Abstract Interpretation (VMCAI'06)*, volume 3855 of *LNCS*, pages 348–363. Springer, January 2006. ISBN 3-540-31139-4. doi: 10.1007/11609773.

[17] David Monniaux. Automatic modular abstractions for linear constraints. In *POPL (Principles of programming languages)*. ACM, 2009. ISBN 978-1-60558-379-2. doi: 10.1145/1480881.1480899.

[18] David Monniaux. Optimal abstraction on real-valued programs. In Gilberto Filé and Hanne Riis Nielson, editors, *Static analysis (SAS '07)*, volume 4634 of *LNCS*, pages 104–120. Springer, 2007.

[19] David Pichardie. *Interprétation abstraite en logique intuitionniste : extraction d'analyseurs Java certifiés*. PhD thesis, Université Rennes 1, 2005. In French.

[20] Sriram Sankaranarayanan, Henny B. Sipma, and Zohar Manna. Scalable analysis of linear systems using mathematical programming. In *VMCAI*, volume 3385 of *LNCS*, pages 21–47. Springer Verlag, 2005.