



Completeness in static analysis by abstract interpretation, a personal point of view

David Monniaux

► To cite this version:

David Monniaux. Completeness in static analysis by abstract interpretation, a personal point of view. Vincenzo Arceri; Agostino Cortesi; Pietro Ferrara; Martina Olliaro. Challenges of Software Verification, 238, Springer Nature Singapore, pp.93-108, 2023, Intelligent Systems Reference Library, 978-981-19-9600-9. 10.1007/978-981-19-9601-6_6 . hal-03857312v2

HAL Id: hal-03857312

<https://hal.science/hal-03857312v2>

Submitted on 25 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Completeness in Static Analysis by Abstract Interpretation, a Personal Point of View

David Monniaux

May 25, 2023

Abstract

Static analysis by abstract interpretation is generally designed to be “sound”, that is, it should not claim to establish properties that do not hold—in other words, not provide “false negatives” about possible bugs. A rarer requirement is that it should be “complete”, meaning that it should be able to infer certain properties if they hold. This paper describes a number of practical issues and questions related to completeness that I have come across over the years.

1 Introduction

The concept of *completeness*, under several definitions, permeates mathematical logic. A proof system is deemed complete with respect to a semantics if it can prove all properties that are true in this semantics. For instance, Gödel’s completeness theorem states that first-order logic (that is, any reasonable proof system for it) can prove any property true in all models. A decision procedure for a logic (that is, a procedure that answers “true” or “false” for any formula in the logic F) is expected to be *sound* (it does not declare to be true formulas that are not true) and *complete* (it declares to be true all formulas that are true).

The concepts of completeness (a) for a proof system with respect to a class of properties (b) for a procedure that searches for proofs of such properties within such a proof system are related, but distinct. Obviously, if a proof system is incomplete, then so is any procedure that searches for proofs within that system. However, it is possible to have a complete proof system and an incomplete procedure for searching within it, for instance for practical efficiency reasons.

In abstract interpretation, completeness, at least at the global level (proving properties of whole programs), is often forgone straight from the start when designing the abstraction (for instance, interval arithmetic is used even though it is obvious that it cannot prove safety in general, even if the property to prove is itself an interval), and there thus may be little reluctance to adding further incompleteness in the solving procedure by using approaches such as widening operators [5, 6]. Yet, it is interesting to control this further incompleteness, both from a theoretical and a practical points of view. Completeness in the solving method ensures some predictability in the analysis outcome, while the brittleness of some incomplete approaches (the approach succeeds or fails depending on seemingly inconsequential aspects of the program and the property) surprises

end users: for instance, providing more precise information on the precondition of a program may prevent the analysis procedure from proving a property that it could prove with a less precise precondition.

In this paper, I discuss various instances of completeness and incompleteness that I have come across, without any pretense of thorough theoretical discussion and, ironically, no pretense of completeness. For more theoretical discussion, Giacobazzi et al. [14, Sec. 7] surveyed the literature extensively. For a study of completeness with respect to the relationship between the semantics and the abstract domain, and constructions to make the domain complete, see [13].

2 Completeness of the Abstraction: the Case of LRU Caches

It is well-known that it is equivalent to (a) evaluate an integer multivariate polynomial over integer inputs, then take the final result modulo N (b) do the same computation but reducing modulo N at every step.¹ Due to a strong mathematical property (a ring morphism), it is possible to replace an expensive computation, possibly involving large numbers, by a simpler one that abstracts the expensive one by operating on small abstractions of the data.

This example is ideal, and one could doubt the existence of complete abstractions that significantly simplify computations for nontrivial program analysis problems; yet we came across one such example, where we perform an exact analysis defined by a composition of exact abstractions and an exact solving procedure.

All current processors use some form of cache memory: frequently accessed code or data is retained in fast memory close to the processor core, avoiding much slower accesses to main memory. Cache analysis consists in determining, given a program and a description of a processor cache subsystem, which memory accesses are always “cache hits” (data already in cache), which are always “cache misses” (data not in cache, thus access to external memory), or more complex properties (“the first time this statement is executed, the access is a cache miss, then subsequent accesses are hits”). Such an analysis is for instance used as a prerequisite for worst-case execution time analysis.

A possible approach to solving the cache analysis problem would be to apply a model-checker to the combination of the program under analysis (or some suitably simplified model thereof) and of the cache system. A first and obvious abstraction is, for the cache system, to retain only the addresses of the memory blocks currently stored within the system, but not their contents. This abstraction is exact: the cache system finds data, or decides to evict data to make room for new contents, according to addresses only, never according to contents. The resulting model of the cache system, for a fixed processor model, is finite: the cache consists of

¹For $N = 9$, this is the basis for the method of “casting out nines”, named so because reduction modulo N can be implemented over numbers in decimal form by summing the digits of the number, with 9 being replaced by 0 (repeat the process until the result consists on one single digit). Schoolchildren used to apply this method to check their hand computations: the final result of an error-prone integer computation, taken modulo 9, should be identical to the result obtained with reduction at every step. This is of course not a sound check: if results differ, one is sure that there has been some error somewhere (no false positives), but they can coincide by chance.

a fixed number of cache lines, buffers, etc., labeled with addresses taken within a fixed range. However, such a model is in practice intractable on all but the simplest examples.

The naive vision of cache memory is that of a *fully associative* cache: any cache line may be used to store any block from the main memory, regardless of its address. In reality, a cache is usually split into several *cache sets*, each of which able to retain blocks present only at a given class of addresses in the main memory. The classes of addresses for the different cache sets form a partition of the possible addresses in main memory. In almost all cases² these cache sets operate completely separately from each other. It is thus possible to perform cache analysis separately for each cache set; this is again an exact abstraction. However, this still results in intractable models.

Each cache set is composed of a fixed number of *cache lines*; the number N of lines is also known as the number of *ways* or the *associativity* of the cache. Each cache line may be empty or nonempty; a nonempty cache line stores a block from main memory: its address and its contents (we have already seen that the contents are not relevant for analysis). For simplicity, we shall consider here the case of single-level, read-only caches. A read from a memory location not currently in the cache triggers a load from main memory into the cache. Except in the rare case when the relevant cache set still has empty lines, this involves evicting a block from the cache set. The selection of the block to be evicted is made according to a *cache replacement policy*. The most obvious replacement policy is to evict the *least recently used* (LRU) block.³ A cache set is therefore modeled as a sequence of at most N block addresses, in increasing *age*: the *youngest* block is the one accessed most recently, the *oldest* is the least recently accessed. The positions of blocks within the set change as data is accessed.

Let us take an example with $N = 4$. Assume a, b, c, \dots denote distinct addresses in memory, and assume the cache set initially contains $abcd$, meaning that a is the youngest block and d the oldest. If the processor requires block d , then it is rejuvenated and the cache set then contains $dabc$. If then the processor requires block e , then the oldest block (c) is evicted and the cache set contains $edab$.

Let us now see our approach to analyzing programs over LRU caches [27]. Consider the case where we want to know whether a certain fixed block a is within the cache at certain program locations. Then, the blocks older than a in the cache are unimportant; so is the ordering of the blocks younger than a . The idea is that if a block is younger than a , then accessing it again will just change the ordering of blocks younger than a , but not the age of a ; and if it is older, then accessing it will increment the age of a .

A cache set may thus be described, with respect to a block a , by an abstract configuration: either “ a is absent” or “ a is here and the set of blocks younger than a is $\{\dots\}$ ”. Again, this is an exact abstraction.

²The exception is the “pseudo round robin” cache replacement policy, which involves a counter global to all cache sets. Also, dependencies between cache sets arise if one considers an integrated model of the processor pipeline, pre-fetching units, and caches, because whether or not some data is in some cache set influences indirectly whether some other data will be fetched.

³Some processors implement this LRU policy. It is however more common to implement so-called pseudo-LRU policies, meant to provide the same practical performance at a fraction of the hardware cost [19, 15]. Unfortunately, these pseudo-LRU policies are very different from the point of view of cache analysis [19, 23, 25].

Cache analysis using that abstraction specialized for accesses to a thus amounts to collecting the sets (for all program locations) of reachable abstract configurations in the cache set where a is to be stored. This is still very costly.

A collection of abstract configurations for that cache set consists of a Boolean “is it possible that a is absent” and a set of sets of size less than N of blocks distinct from a . A crucial observation is that, in this set of sets, only the minimal and maximal elements (with respect to the inclusion ordering) matter: if it is possible that a is preceded by $\{b\}$, $\{b, c\}$, $\{b, c, d\}$, then, for checking if it is possible that a (at some later point) is out of the cache, it is only necessary to retain $\{b, c, d\}$, which subsumes the other two cases; and for checking if it is possible that a (at some later point) is in the cache, it is only necessary to retain $\{b\}$, which again subsumes the other two cases. The reason for this subsumption is that if a sequence of steps leads from a state where a is preceded in the cache by a set P of blocks to the next access to a , and that access is a miss, then the same sequence of steps, starting with any cache state where the set of blocks younger than a is a superset of P , also leads to a miss (*mutatis mutandis* for subsets and hits). Retaining only the extremal elements is thus, again, an exact abstraction from the point of view of cache analysis.

To summarize, we first gradually simplified the cache state by abstracting away parts and aspects that are irrelevant to the analysis under way, then we simplified the collection of abstract cache states by using a subsumption relation. The resulting abstraction is exact with respect to the properties under analysis.

A this point, what remains is how to implement this abstraction. What is needed is an efficient way to represent sets of incomparable sets of blocks (*antichains*); we opted for zero-suppressed binary decision diagrams (ZDD). For better efficiency, this analysis, still somewhat costly, is applied to memory accesses only after some cheap approximate analyses [26] have failed to classify these accesses.

An advantage of an exact approach is that, since the result is uniquely defined, it is possible to test implementations of various exact abstractions against each other; the results should be identical. This way we discovered a subtle ZDD bug by comparison with a (less scalable) model-checking approach.

Granted, this approach is “exact” or “complete” only because a simplified model of the program (a control-flow graph, not taking into account the data being processed) is used. This model introduces approximation: for instance, in the following program

```
if (flag) access(a); else access(b);
if (flag) access(a); else access(b);
```

the only feasible sequences of accesses are a, a and b, b , (and thus in either case the second access is a hit), but an analysis based on control-flow only will not track the value of the flag and consider that sequences a, b and b, a are also feasible, and conclude that the second access may be a miss. However, tracking the value of data precisely makes the model undecidable. An intermediate solution that retains decidability is to consider control-flow with precise modeling of procedure calls [23].

On a final note, on many examples [27], worst-case execution time (WCET) analysis ended up being overall faster with the help of this apparently expensive analysis. The reason is that the WCET analysis in use

involved enumerating all reachable pipeline states. If the cache analysis is imprecise and cannot conclude about some access, where the exact analysis would conclude to a “always miss”, this does not normally change the bound on WCET, but this increases the number of pipeline states to consider. In short:

- incomplete analyses are first applied in order to answer most sub-problems exactly; only the subproblems for which the answer is definitely unknown are passed to the more expensive complete analysis;
- completeness is ensured by a composition of abstractions that do not lose any precision with respect to the properties we are interested in, but which greatly simplify the model to analyze;
- the resulting model is analyzed exactly;
- complete analyses have an exactly defined result, thus testing for bugs is easy by comparing results;
- for the same reason, it is possible to study the complexity of the problem [25, 23];
- the extra cost of the complete analysis may be offset by savings in client analyses, because a more precise result means fewer case analyses down the road.

3 Completeness or Incompleteness of the Analysis Method

A classical question of completeness in abstract interpretation is whether the abstract domain in use is sufficient to prove the properties being sought. Yet, even if the abstract domain is sufficient, it may happen that the abstract interpretation method being used cannot find the necessary invariants within the domain.

3.1 Widening Operators

Let us see an example: some simplified version of a dataflow program where `i` is an index into some circular array of length 42.

```
i = 0;
while (true) {
    if (trigger()) {
        i = i+1;
        if (i > 42) i = 0;
    }
    assert (i < 1000);
}
```

Textbook forward static analysis by intervals [5, 4] will compute, for variable `i`, a sequence of intervals $[0, 0]$, $[0, 1]$, $[0, 2]$, \dots , and widen it to $[0, \infty)$. If narrowing iterations are used, in their simplest form by starting from the $[0, \infty)$ invariant, running the loop body once more and adding the initialization states, one obtains $[0, 999]$. The assertion cannot be proved using that invariant. Yet the assertion would have been proved if the analysis had guessed the least possible interval, $[0, 42]$. Clearly, the problem is with the inference method (iterations with widening), not the abstract domain (intervals). Iterations with widening are an *incomplete*

method with respect to the abstract domain: they may fail to discover suitable inductive invariants when some exist in the domain.

Note also a surprising characteristic of this incomplete approach: if instead of the precondition $i = 0$ we had analyzed the loop with the precondition $0 \leq i \leq 42$, we would be able to prove the assertion correct. This *non monotonic* behavior (a less precise precondition, or a coarser abstraction of some program semantics leads to more precise analysis results) may surprise end users.

Another surprising characteristic of widening operators for relational domains is that they do not commute, in general, with projection, meaning that adding some extra variables may change the result of the analysis on the original variables even if the extra variables have no influence on the original ones (for instance, if these variables are mere “observers”). Consider for instance the program:

```
i=0; j=0;
while (true) {
  if (*) i=0; else {i=1; j=0; }
  if (i==0) j=j+1;
}
```

In this program, j observes the number of last iterations during which i was 0. The set of reachable states of this program for (i, j) is $(1, 0) \cup \{(0, n) | n \in \mathbb{N}\}$; in particular the set of reachable values for i is $\{0, 1\}$, and it can be computed exactly with interval analysis by postponing widening by one step, or by applying one step of narrowing. Now consider the sequence of polyhedra, that is, polygons, computed by polyhedral analysis with widening [7]: at step n , the polygon is defined by vertices $(0, 0)$, $(1, 0)$, $(0, n)$. The lines passing through vertices $(0, 0)$ and $(1, 0)$ (constraint $j \geq 0$) and the one passing through vertices $(0, 0)$ and $(0, n)$ (constraint $i \geq 0$) are stable; the line passing through vertices $(1, 0)$ and $(0, n)$ is unstable and will be discarded by widening. The resulting system of constraints is $i \geq 0 \wedge j \geq 0$; we have lost the $i \leq 1$ constraint obtained by interval analysis. In some cases, a *stratified* approach, where successive analyses take increasing subsets of variables in consideration, may be able to recoup precision [24].

3.2 Exact Solving

A safety property is established using forward abstract interpretation by inferring some inductive invariants in the abstract domain and checking that these invariants imply the property. A complete abstract interpretation method is one that would compute always such invariants if they exist.

Clearly, any method that computes the least inductive invariants in the abstract domain is complete: if there exist invariants in the domain that can prove the safety property, then, a fortiori, the least inductive invariants in the domain can prove that property.

One such method is ascending iterations *without widening* within a domain that satisfies the ascending chain condition (no infinite strictly ascending sequences), which ensures termination. This includes domains of finite height (there is a bound on the length of strictly ascending chains), and in particular finite domains, such as powersets of finite sets. If $f : L \rightarrow L$ is a monotone function, and \perp is the infimum of L , then the sequence $f^n(\perp)$ is ascending and, in all the above cases, becomes stationary. When

$f^n(\perp) = f^{n+1}(\perp)$, then $f^n(\perp)$ is the least solution of $x = f(x)$. The algorithms used in practice are algorithmic improvements over this idea.⁴

An example of an infinite domain of finite height is that of solution sets of systems of linear equations over a fixed number of variables [20]. When a solution set S is strictly included in a solution set S' , the dimension of S' is strictly greater than that of S ; this dimension cannot be more than n , which thus bounds the height of the lattice.

The domain of intervals has none of these characteristics; but it is however possible to compute the least inductive invariant within that domain of programs such as the one above, by specifying this least inductive invariant as a least fixed point, writing a system of numeric equations that this fixed point should satisfy, and solving this system for the least solution. Indeed, for the above example, let us write down the equations that h should satisfy for $[0, h]$ to be a fixed point for the loop:

$$h = \min(\max(\min(42, h + 1), h), 999) \quad (1)$$

Let us solve this equation. In any solution, the outermost “minimum” operator must be equal to one of its arguments. Assume it is the second argument, 999. One can indeed verify that 999 is a solution to this equation. Now consider the case where it evaluates to its left argument. The equation then becomes $h = \max(\min(42, h + 1), h)$. Similarly, the “maximum” operator evaluates to one of its operands. Assume it evaluates to its right argument. The system then simplifies to $h = h$; but this is under the condition that $h \leq 999$ and $h \geq \min(42, h + 1)$. When $h + 1 > 42$, that is, $h \geq 42$, this minimum is equal to 42; thus all $42 \leq h \leq 999$ yield valid solutions. When $h + 1 \leq 42$, this minimum is equal to $h + 1$; but then the condition becomes $h \geq h + 1$, which would imply $h = \infty$, which contradicts $h \leq 999$. Now assume the “maximum” operator evaluates to its left argument. The system then simplifies to $h = \min(42, h + 1)$, which yields only $h = 42$ as solution. We conclude that $h = 42$ is the least solution.

The above reasoning by case analysis over the minimum and maximum operators can be automated through exhaustive case analysis or, most subtly, by a SMT (satisfiability modulo theory) solver. Many such solvers can also optimize within the solution space, and thus directly produce the least fixpoint. Alternatively, for proving safety properties, it is sufficient to query the solver for any solution sufficient for proving the safety property.

A more principled approach to the case analysis for the maximum operators is ascending policy iteration [11], which considers an ascending sequence of fixpoints of systems obtained by picking arguments to the “maximum” operators. All the above approaches generalize to domains other than intervals, for instance to zones [12], at the expense of some algorithmic complications.

Let us go further. Assume the abstract domain consists of the sets defined by a parameterized predicate $I(\mathbf{p}, \mathbf{x})$ where \mathbf{p} is the vector of parameters, \mathbf{x} is the program state. Note that this encompasses intervals, octagons, template polyhedra domains, or even polyhedra with a fixed number of faces. Also note that I may contain disjunctions and express non-convex relations.

⁴In many cases, $L = L_b^m$ where L_b is a base lattice. An element x of L is then decomposed into x_1, \dots, x_m , and f is decomposed into its m projections f_1, \dots, f_m . The problem is then to find a solution of a system of equations $x_1 = f_1(x_1, \dots, x_m), \dots, x_m = f_m(x_1, \dots, x_m)$, typically using a system of working set of variables being updated, with some judiciously chosen ordering for choosing the next update to process.

The inductiveness condition for a transition $\tau_{i,j}$ from program location i to program location j may thus be written as the Horn clause

$$\forall \mathbf{x}, \mathbf{x}' I(\mathbf{p}_i, \mathbf{x}) \wedge \tau_{i,j}(\mathbf{x}, \mathbf{x}') \Rightarrow I(\mathbf{p}_j, \mathbf{x}) \quad (2)$$

Safety conditions may be expressed as $\forall \mathbf{x} I_i(\mathbf{p}_i, \mathbf{x}) \Rightarrow C_i(\mathbf{x})$, and program startup may be specified as $\forall \mathbf{x} S_i(\mathbf{x}) \Rightarrow I_i(\mathbf{p}_i, \mathbf{x})$. Inductive invariants within the abstract domain suitable for proving the safety properties are thus expressed as the solutions, in the parameters \mathbf{p}_i , of a system of Horn clauses.

If I , the transitions $\tau_{i,j}$, the safety conditions C_i and the start conditions S_i are all expressed within a theory admitting algorithmic quantifier elimination, such as Presburger arithmetic or the theory of real closed fields, then the existence of such invariants is decided by quantifier elimination. For instance, if we consider a program operating over real numbers, and invariants of the form $A(\mathbf{p})\mathbf{x} \leq B(\mathbf{p})$ where $A(\mathbf{p})$ and $B(\mathbf{p})$ have a fixed number r of rows (a polyhedron with at most r faces), then the existence of a suitable invariant is established by quantifier elimination in the theory of real closed fields. It is even possible to specify that one wants the least inductive invariant and obtain it by quantifier elimination [21].

3.3 Imprecise Abstract Transfer Functions

Widening operators are the best known source of non-monotonicity and incompleteness in finding suitable invariants in abstract interpretation, but they are not the only one. There are, and this may be surprising, cases of non-monotonicity and incompleteness due to transfer functions, particularly in combinations of abstractions.

It is well-known that even if each individual step of abstract interpretation computes the least possible post-condition within the abstract domain compatible with the precondition, this property does not extend to composition of steps; in other words, optimality does not compose. Let us see a simple example:

```
y = x ;
z = x - y ;
```

If x is known to lie in $[0, 1]$, then so does y , and this is optimal—non-optimal but sound answers would be intervals containing $[0, 1]$. Then interval arithmetic deduces that z is in $[-1, 1]$, which is sound, but not optimal: the optimal interval is $[0, 0]$. However, in order to reach that conclusion, one must know that $x = y$ at the intermediate control point, that is, a relation between x and y , which is not possible in a non-relational abstraction such as intervals.

One workaround to this weakness is to propagate a system of relations along with the interval analysis. For instance, we had propagated $y = x$, then by substituting $y \mapsto x$ into $x - y$ and simplifying, we would have obtained $z = 0$.

An approach to implement this workaround [8], for instance applied in the Astrée static analyzer, is to propagate a terminating rewriting system (for instance, populated in chronological order) along with the interval information, perform interval analysis on both the original expressions and the expressions obtained by rewriting and simplification, and then take the intersection. Here, the rewriting system would contain $y \mapsto x$, and the rewritten and simplified expression would yield $z = 0$.

This approach however creates non-monotonicity.⁵ Consider this example [8, ex. 4]:

Code	Symbolic computation	Less precise symbolic
<code>int i, j=i+1;</code>	$j \mapsto i + 1$	nothing
<code>int k=j+1;</code>	$j \mapsto i + 1, k \mapsto j + 1 \mapsto i + 2$	$k \mapsto j + 1$
<code>if (j > 0) l=k;</code>	$j \mapsto i + 1, k \mapsto i + 2, l \mapsto i + 2$	$k \mapsto j + 1, l \mapsto j + 1$

Assume we don't know anything about the values of the variables initially, and consider the interval obtained for l at the end of the “then” branch of the test. Simple interval propagation yields no information about l . The rewriting system yields $l \mapsto i + 2$ and since no information is known about i , no information is known about l . Yet, if we forget that $j \mapsto i + 1$, and apply the resulting rewriting system, we obtain $l \mapsto j + 1$ and thus $l > 1$.

Granted, some improvements could be made by more clever propagation.⁶ For instance, since $j \mapsto i + 1$, the guard could be rewritten into $i + 1 > 0$, and this could itself be rewritten into $i > -1$. This amounts to “inverting” $j \mapsto i + 1$ to $i \mapsto j - 1$. More generally, one would need to consider the rewriting system not as a directed system, but as a system of equations. The combination of a system of linear equations and intervals forms the basis of the simplex algorithm for linear programming, and obtaining optimal interval bounds from a combination of equalities and interval bounds amounts to linear programming. In fact, any convex polyhedron can be expressed as the projection of a set defined by the conjunction of linear equalities and interval bounds.⁷ The appropriate domain for dealing with such constraints precisely would thus be that of convex polyhedra [7, 18].

4 Undecidability of an Abstraction

The argument from Section 3.2 for establishing decidability of the existence of an inductive invariant within the domain of convex polyhedra with at most r faces does not extend to the domain convex polyhedra with any number of faces. In fact, there is no known algorithm for deciding the existence of inductive invariants within that domain for any nontrivial class of programs. Invariant inference approaches for that domain, starting from the early proposals from Cousot and Halbwachs [7, 18], are typically based on iterations with widening. An intriguing question is whether it is inevitable to resort to heuristics.

4.1 Polyhedral Abstraction

I have attempted proving that there is no algorithm deciding the existence of polyhedral invariants for linear transition systems (real or integer), and have so far failed. Because of my efforts in raising attention to

⁵This non-monotonicity resulted in some “false alarms”, that is, warnings about non-existent problems, on some industrial programs.

⁶More generally, advanced tools such as Astrée implement complex combinations of domains and iteration strategies, which make many examples of non-monotone behavior and false alarms on simple analyses actually succeed.

⁷Consider the representation of the polyhedron as a system of inequalities $l_i(x_1, \dots, x_n) \leq B_i$, create a new variable y_i for any linear combination l_i for which there does not already exist a variable, keep this inequality as $y_i \leq B_i$ and $y_i = l_i(x_1, \dots, x_n)$, then the set defined by the original system is the projection of the set defined by the transformed system on the x variables

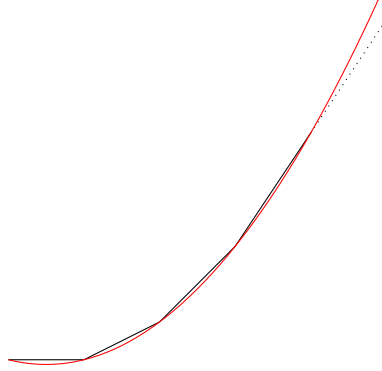


Figure 1: When the process is non terminating, any polyhedron containing the reachable points inevitably goes below the parabola

this issue, some colleagues named this the “Monniaux problem”. Several distinguished researchers⁸ have also reported trying to solve it and failing.

4.1.1 Undecidability with Quadratic Guards

Progress has however been made on variants of the so-called “Monniaux problem” [10]. I was able to prove that this problem becomes undecidable if one is allowed to use a quadratic transition guard: it is then possible to encode a deterministic counter machine reachability problem into the invariant inference problem. Let us recall how (proof sketch) [22].

Let the counter machine M operate over variables z_1, \dots, z_n , to which we add two variables x and y , initialized to 0. The transitions of the counter machine are synchronously combined with steps $(x, y) \mapsto (x + 1, y + x)$. The combined machine thus simulates the counter machine on variables z_i together with a parabola $(n, \frac{1}{2}n(n-1))$ on variables (x, y) , where n is the number of steps taken so far. Now modify the resulting machine by conjoining to all transitions the guard $y = \frac{1}{2}x(x-1)$; clearly this does not modify the behavior of the machine. Add a special control state σ_b , meant to be unreachable, and transitions from any state $y < \frac{1}{2}x(x-1)$ to that state. Call the resulting machine M' .

Assume M terminates in N steps, then so does M' . Consider the convex hull H of the finite family of points such that $x = k$, $y = \frac{1}{2}k(k-1)$, z_1, \dots, z_n is the state reached after k steps in the execution of M , and $0 \leq k \leq N$. The only points in H that lie on the $y = \frac{1}{2}x(x-1)$ parabola project to the reachable states of M . H , in general, contains extra states (integer points) such that $y > \frac{1}{2}x(x-1)$, but due to the nonlinear guards, these states cannot transition to other states. Thus H is an inductive invariant for M' that proves the inaccessibility of σ_b .

Assume M does not terminate, and thus so does M' . Any inductive invariant for M' must contain the infinite family of points such that $x = k$, $y = \frac{1}{2}k(k-1)$, z_1, \dots, z_n is the state reached after k steps, for $k \geq 0$. Any convex polyhedron that contains this infinite family of points must have a point $y < \frac{1}{2}x(x-1)$: assume there is none and consider the vertex V^* with maximal x^* ; this vertex must lie on the parabola; there is for

⁸Including some at the 2022 CSV workshop in Venice, whence this volume comes.

$x \geq x^*$ at least one infinite face of the form $y \geq L(x, z_1, \dots, z_n)$; but then inevitably there will be points of that face lying below the parabola (Fig. 1). However, any convex polyhedron that contains a point where $y < \frac{1}{2}x(x-1)$ is not an inductive invariant capable of proving that σ_b is unreachable. It follows that there is no convex polyhedron that is an inductive invariant for M' capable of showing the unreachability of σ_b .

Thus, M' has an inductive polyhedral invariant suitable for proving the unreachability of σ_b if and only if M terminates, and this for arbitrary M in a class of machines with undecidable termination.

The same reasoning applies whether the state space is considered over integers or over the reals.

4.1.2 Perspectives and Dead Ends

The role of the parabola $y = \frac{1}{2}x(x-1)$ in the above proof could be played by other sets, such as a circle $x^2 + y^2 = 1$: the process that enumerated points on the parabola can be replaced by one that enumerates points on the circle. Initialize $(x, y) = (1, 0)$, and, as next step, apply a rotation matrix defined by a Pythagorean triple $((a, b, c)$ integers such that $a^2 + b^2 = c^2$, for instance $(3, 4, 5)$): $\begin{pmatrix} \frac{a}{c} & \frac{b}{c} \\ \frac{b}{c} & -\frac{a}{c} \end{pmatrix}$. This however leads to more complex proof arguments, and does not gain anything: we still need a nonlinear guard.

Essentially, what we used in both cases is a strictly convex set S ($y \geq \frac{1}{2}x(x-1)$) with boundary definable by a guard ($y = \frac{1}{2}x(x-1)$), with exterior ($y < \frac{1}{2}x(x-1)$) definable by a guard, and with a way to enumerate points in the boundary while still remaining in the class of deterministic transitions under consideration. Strict convexity is used so that taking the convex hull of points in S , as when using polyhedral invariants, does not add “parasitic” points on the boundary. This ensures that the guard that keeps only points on the boundary removes “parasitic” points.

Could we achieve similar goals using only linear constraints? Suppose we can express such a guard with a formula, possibly containing disjunctions. Then, by distributivity, this formula expresses a finite union of polyhedra $P_1 \cup \dots \cup P_n$. Then there is a i such that the enumeration process eventually picks two points (in fact, an infinity of them) in P_i ; and thus the guard cannot exclude “parasitic” points between these two, at least over the reals.

A possible course of action could be to set the problem over integers and take advantage of the implicit guard that non-integer points are discarded. This however was also a dead end.

4.2 Richer Domains

If the abstract domain (class of invariants) under consideration is rich enough to express exactly the set of reachable states of the programs to be analyzed, then obviously the problem is undecidable: a suitable inductive invariant (the set of reachable states) exists in the domain if and only if the safety property is true.

This happens for instance if the domain includes Σ_1^0 formulas of Peano arithmetic, that is, formulas of the form $\exists v_1 \dots \exists v_n F$ where F only contains bounded quantifiers and the usual arithmetic operations and comparisons. By a form of Gödel’s encoding, for any integer program with m variables built using the usual arithmetic operators, it is possible to

build a Σ_1^0 formula $F(k, s_1, \dots, s_m)$ satisfied if and only if (s_1, \dots, s_m) is the state of the program after k steps of [28, Ch. 7][3].

Such domains are obviously too rich. In fact, with the domain of Σ_1^0 formulas described above, it is not even possible in general to check for inductiveness, because the formulas belong to an undecidable class.

However, even semilinear sets (sets defined by Boolean combinations of linear inequalities with algebraic coefficients, thus disjunctions of convex polyhedra) are sufficient to obtain undecidability even for a very restricted class of programs (one single loop control state, nondeterministic choice between two linear transformations, no guards) [10, Th. 9].

5 Perspectives and Conclusion

The question of the completeness of the domain with respect to the properties to prove, and the class of programs under consideration, is distinct from the problem of algorithmically finding suitable invariants within that domain—or the related problem of computing the least inductive invariant within the domain.

The traditional approach for abstract interpretation in domains with infinite ascending chains is iterations with widening. This approach has known weaknesses: the analysis may miss the invariants necessary for the proof and the analysis behavior is non-monotone (increasing knowledge about preconditions or transitions may decrease the precision of the analysis), which is a source of “brittleness” (a minor change in the program causes the analysis to fail to prove the property). Many tricks are thus used to improve the invariants obtained with widening: narrowing iterations [6, 4], lookahead widening [17], guided static analysis [16]. Though they help in practice, none of them guarantees that analysis will succeed.

Over time, completely different approaches were designed to compute suitable arguments for proving safety properties. While widening is a form of extrapolation (look at sets of reachable states in $0, 1, 2 \dots$ steps and try to extrapolate for arbitrary number of steps), these methods are based on *interpolation*: given some under-approximation of the set of reachable states, and the property to prove, find a simple separation between the two (a Craig interpolant) and hope that the interpolant, or components thereof, or formulas constructed from components of successive interpolants, becomes inductive. One such approach is *property-directed reachability* [1, 9], later enriched with a number of extensions and improvements (dealing with arithmetic theories, dealing with non-linear Horn clauses and not just transition systems). Variants of this approach are in particular implemented in the popular Z3 solver.⁹ Unfortunately, this approach, too, is brittle [2]: minor differences in the problem to be solved (names of variables, etc.) may result in dramatic changes in outcome (finding invariants or timing out).

Approaches that are guaranteed to find the least inductive invariant in the chosen abstract domain (or, more generally, an inductive invariant suitable for proving a given property, if it exists) are more resilient; I have listed several of them in Section 3.2. Some of these methods however do not scale up well (those based on quantifier elimination); policy iteration seems to be the most scalable.

In our view, the case for continued use of widening operators, despite their known weaknesses, or other non monotonic and/or brittle meth-

⁹<https://github.com/Z3Prover/z3>

ods, would be strengthened by a proof that the existence of a suitable invariant in the domain is undecidable, or at least has high complexity. Unfortunately, such undecidability proofs are hard.

In contrast, complete methods have many desirable properties. We have illustrated this with the example of LRU cache analysis (Section 2). Since the end result has a unique definition, there is no brittleness, and several algorithms or implementations can be used to compute the same result, which allows performance comparisons (the meaning of performance differences between methods producing non comparable results is unclear) and validation by testing that results are identical.

References

- [1] Aaron R. Bradley. “SAT-Based Model Checking without Unrolling”. In: *Verification, Model Checking, and Abstract Interpretation - 12th International Conference, VMCAI 2011, Austin, TX, USA, January 23-25, 2011. Proceedings*. Ed. by Ranjit Jhala and David A. Schmidt. Vol. 6538. Lecture Notes in Computer Science. Springer, 2011, pp. 70–87. DOI: 10.1007/978-3-642-18275-4_7.
- [2] Julien Braine, Laure Gonnord, and David Monniaux. *Verifying Programs with Arrays and Lists*. Internship report. École normale supérieure de Lyon, June 2016. URL: <https://hal.archives-ouvertes.fr/hal-01337140>.
- [3] Stephen A. Cook. “Soundness and Completeness of an Axiom System for Program Verification”. In: *SIAM Journal on Computing* 7.1 (1978), pp. 70–90. DOI: 10.1137/0207005.
- [4] Patrick Cousot. “Méthodes itératives de construction et d’approximation de points fixes d’opérateurs monotones sur un treillis, analyse sémantique de programmes”. Thèse d’état ès sciences mathématiques. Grenoble, France: Université scientifique et médicale de Grenoble, Mar. 21, 1978.
- [5] Patrick Cousot and Radhia Cousot. “Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints”. In: *Conference Record of the 4th ACM Symposium on Principles of Programming Languages*. Los Angeles, CA, Jan. 1977, pp. 238–252.
- [6] Patrick Cousot and Radhia Cousot. “Abstract Interpretation Frameworks”. In: *J. Log. Comput.* 2.4 (1992), pp. 511–547. DOI: 10.1093/logcom/2.4.511. URL: <https://doi.org/10.1093/logcom/2.4.511>.
- [7] Patrick Cousot and Nicolas Halbwachs. “Automatic Discovery of Linear Constraints among Variables of a Program”. In: *Proceedings of the Fifth Conference on Principles of Programming Languages*. ACM Press, 1978.
- [8] Patrick Cousot et al. “Combination of Abstractions in the ASTRÉE Static Analyzer”. In: *Advances in Computer Science — ASIAN 2006. Secure Software and Related Issues*. Lecture Notes in Computer Science 4435. Springer, 2008, pp. 272–300. ISBN: 978-3-540-77504-1. DOI: 10.1007/978-3-540-77505-8_23.

- [9] Niklas Eén, Alan Mishchenko, and Robert K. Brayton. “Efficient implementation of property directed reachability”. In: *International Conference on Formal Methods in Computer-Aided Design, FMCAD ’11, Austin, TX, USA, October 30 - November 02, 2011*. Ed. by Per Bjesse and Anna Slobodová. FMCAD Inc., 2011, pp. 125–134.
- [10] Nathanaël Fijalkow et al. “On the Monniaux Problem in Abstract Interpretation”. In: *SAS*. Vol. 11822. Lecture Notes in Computer Science. Springer, 2019, pp. 162–180.
- [11] Thomas Gawlitza and Helmut Seidl. “Precise Fixpoint Computation Through Strategy Iteration”. In: *Programming Languages and Systems, 16th European Symposium on Programming, ESOP 2007, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2007, Braga, Portugal, March 24 - April 1, 2007, Proceedings*. Ed. by Rocco De Nicola. Vol. 4421. Lecture Notes in Computer Science. Springer, 2007, pp. 300–315. DOI: 10.1007/978-3-540-71316-6_21.
- [12] Thomas Martin Gawlitza and Helmut Seidl. “Precise Program Analysis through Strategy Iteration and Optimization”. In: *Software Safety and Security*. Vol. 33. NATO Science for Peace and Security Series - D: Information and Communication Security. IOS Press, 2012, pp. 348–384.
- [13] Roberto Giacobazzi and Francesco Ranzato. “Completeness in Abstract Interpretation: A Domain Perspective”. In: *AMAST*. Vol. 1349. Lecture Notes in Computer Science. Springer, 1997, pp. 231–245.
- [14] Roberto Giacobazzi, Francesco Ranzato, and Francesca Scozari. “Making abstract interpretations complete”. In: *J. ACM* 47.2 (2000), pp. 361–416.
- [15] Damien Gille. “Study of Different Cache Line Replacement Algorithms in Embedded Systems”. MA thesis. KTH, Mar. 2007. URL: <https://people.kth.se/~ingo/MasterThesis/ThesisDamienGille2007.pdf>.
- [16] Denis Gopan and Thomas W. Reps. “Guided Static Analysis”. In: *Static Analysis, 14th International Symposium, SAS 2007, Kongens Lyngby, Denmark, August 22-24, 2007, Proceedings*. Ed. by Hanne Riis Nielson and Gilberto Filé. Vol. 4634. Lecture Notes in Computer Science. Springer, 2007, pp. 349–365. DOI: 10.1007/978-3-540-74061-2_22.
- [17] Denis Gopan and Thomas W. Reps. “Lookahead Widening”. In: *Computer Aided Verification, 18th International Conference, CAV 2006, Seattle, WA, USA, August 17-20, 2006, Proceedings*. Ed. by Thomas Ball and Robert B. Jones. Vol. 4144. Lecture Notes in Computer Science. Springer, 2006, pp. 452–466. DOI: 10.1007/11817963_41.

- [18] Nicolas Halbwachs. “Détermination automatique de relations linéaires vérifiées par les variables d’un programme”. Theses. Institut National Polytechnique de Grenoble - INPG ; Université Joseph-Fourier - Grenoble I, Mar. 1979. URL: <https://tel.archives-ouvertes.fr/tel-002>
- [19] Reinhold Heckmann et al. “The influence of processor architecture on the design and the results of WCET tools”. In: *Proceedings of the IEEE* 91.7 (2003), pp. 1038–1054. DOI: 10.1109/JPROC.2003.814618.
- [20] Michael Karr. “Affine Relationships Among Variables of a Program”. In: *Acta Informatica* 6 (1976), pp. 133–151.
- [21] David Monniaux. “Automatic Modular Abstractions for Linear Constraints”. In: *POPL (Principles of programming languages)*. ACM, 2009, pp. 140–151. ISBN: 978-1-60558-379-2. DOI: 10.1145/1480881.1480899. arXiv: 0811.0166.
- [22] David Monniaux. “On the decidability of the existence of polyhedral invariants in transition systems”. In: *Acta Informatica* 56.4 (May 18, 2018), pp. 385–389. ISSN: 0001-5903. DOI: 10.1007/s00236-018-0324-y. arXiv: 1709.04382. URL: <https://hal.archives-ouvertes.fr/hal-01>
- [23] David Monniaux. “The complexity gap in the static analysis of cache accesses grows if procedure calls are added”. In: *Formal Methods in System Design* (2022). DOI: 10.1007/s10703-022-00392-w. HAL: hal-03545774.
- [24] David Monniaux and Julien Le Guen. “Stratified Static Analysis Based on Variable Dependencies”. In: *Electron. Notes Theor. Comput. Sci.* 288 (2012), pp. 61–74.
- [25] David Monniaux and Valentin Touzeau. “On the Complexity of Cache Analysis for Different Replacement Policies”. In: *Journal of the ACM* 66.6 (Nov. 2019), 41:1–41:22. ISSN: 0004-5411. DOI: 10.1145/3366018. arXiv: 1811.01740.
- [26] Valentin Touzeau et al. “Ascertaining Uncertainty for Efficient Exact Cache Analysis”. In: *Computer aided verification (CAV)*. Ed. by Viktor Kuncak and Rupak Majumdar. Vol. 10427. Cham: Springer, 2017, pp. 22–17. DOI: 10.1007/3-540-63141-0_10. arXiv: 1709.10008.
- [27] Valentin Touzeau et al. “Fast and Exact Analysis for LRU Caches”. In: *Proc. ACM Program. Lang.* 3 (Jan. 2019), 54:1–54:29. ISSN: 2475-1421. DOI: 10.1145/3290367. arXiv: 1811.01670.
- [28] Glynn Winskel. *The formal semantics of programming languages - an introduction*. Foundation of computing series. MIT Press, 1993. ISBN: 978-0-262-23169-5.