Abstract Interpretation: Past, Present and Future

Patrick Cousot  
CIMS ‡, NYU, USA  
p cousot@cims.nyu.edu

Radhia Cousot  
CNRS Emeritus, ENS **, France  
r cousot@ens.fr

Abstract

Abstract interpretation is a theory of abstraction and constructive approximation of the mathematical structures used in the formal description of complex or infinite systems and the inference or verification of their combinatorial or undecidable properties. Developed in the late seventies, it has been since then used, implicitly or explicitly, to many aspects of computer science (such as static analysis and verification, contract inference, type inference, termination inference, model-checking, abstraction/refinement, program transformation (including watermarking, obfuscation, etc), combination of decision procedures, security, malware detection, database queries, etc) and more recently, to system biology and SAT/SMT solvers. Production-quality verification tools based on abstract interpretation are available and used in the advanced software, hardware, transportation, communication, and medical industries.

The talk will consist in an introduction to the basic notions of abstract interpretation and the induced methodology for the systematic development of sound abstract interpretation-based tools. Examples of abstractions will be provided, from semantics to typing, grammars to safety, reachability to potential/definite termination, numerical to protein-protein abstractions, as well as applications (including those in industrial use) to software, hardware and system biology.

This paper is a general discussion of abstract interpretation, with selected publications, which unfortunately are far from exhaustive both in the considered themes and the corresponding references.

Categories and Subject Descriptors D.2.4 [Software/Program Verification]; D.3.1 [Formal Definitions and Theory]; F.3.1 [Specifying and Verifying and Reasoning about Programs].


Keywords Abstract interpretation, Semantics, Proof, Verification, Static Analysis.

1. Abstraction

No reasoning on complex systems, including computer systems, can be done without abstracting the behavior, i.e. the semantics, of the system. Because reasoning on a system involves determining or proving its properties, the central concept is the abstraction of properties of the system, starting from the strongest one, as specified by the system semantics (and called the collecting semantics¹). This is the purpose of abstract interpretation (where “interpretation” stands both for “meaning” and “execution”). A few gentle introductions to abstract interpretation [44, 79] can be consulted for a first approach, including some publicly available on the web (e.g. web.mit.edu/16.399/www/).

2. Scope

Abstract interpretation comprehends undecidable problems (hence is also applicable to decidable but complex ones). This implies that any tool (prover, checker, analyzer) designed by abstract interpretation will fail on infinitely many counter-examples. For example a finiteness or decidability hypothesis will only be applicable to a very restricted class of programs with finite behavior, hence will fail on infinitely many other ones. This is inherent to undecidable problems hence inescapable. By failure we understand being unsound/incorrect, non-terminating, using a human oracle to assist the computer, etc. Although abstract interpretation also applies to these cases², it is usually used for sound, terminating, and fully automatic program analysis/verification, including the inference of sound inductive arguments (like invariants) to deal with infinite recurrences for unbounded/non-terminating executions, which makes the problem particularly difficult, with a very high complexity.

3. Static Analysis

The origin of abstract interpretation is in static program analysis [50, 51] where reachable states are abstracted by local interval numerical invariants understood as a generalization of type inference [49, 54]. The abstraction from the collecting semantics was formalized by a Galois insertion and convergence acceleration of the iterates by widening, later improved by narrowing. The main innovations at the time were to consider infinite non-Noetherian abstractions of infinite systems and to prove rigorously the correctness of the static analysis with respect to a formal semantics (see more details in footnote 6).

¹ The collecting (or static in [52]) semantics is the strongest property of the standard semantics.

² For example, some commercial products do consider only two iterations in loops without widening, which is an under-approximation of an over-approximating abstraction of program executions which can be formalized by abstract interpretation theory. The theory also proves beyond doubt that the result cannot be claimed to be a sound over-approximation of the program behavior, a conclusion which is not always stated clearly enough for practitioners to have a precise understanding of the scope of commercial static analyzers.
4. Acceleration of fixpoint iteration

The next step was to consider arbitrary fixpoints in ordered structures, order duality for approximations from above and below, ascending and descending iterations, as well as fixpoint iteration with convergence acceleration from above and below with over- and under-approximation using the following extrapolation operators [39, Ch. 4, [52]

<table>
<thead>
<tr>
<th>Extrapolation operators</th>
<th>iteration from below</th>
<th>iteration from above</th>
</tr>
</thead>
<tbody>
<tr>
<td>over-approximation</td>
<td>widening</td>
<td>narrowing</td>
</tr>
<tr>
<td>under-approximation</td>
<td>dual narrowing</td>
<td>dual widening</td>
</tr>
</tbody>
</table>

Chaotic [53] and asynchronous iterations [38, 39] (including at higher-order [55, 149]) paved the way for demand-driven [203] and parallel implementations of static analyzers [86]. Moreover this approach includes alternative, seemingly different, presentations of static analysis (like constraint solving, deductive rules-based (type) systems, etc which nicely generalize in the abstract [68]). The extrapolation operators, which abstract the inference of an inductive hypothesis in iteration are the more powerful, and difficult to design, aspects of abstract interpretation. The design of powerful extrapolation operators pertinent to the degree of approximation of an abstraction is still an art which remains to be much better understood, formalized, and improved. Not enough work has been devoted to the systematic design of extrapolation operators and the fact that they cannot be increasing [146]. Similarly non-iterative methods are just appearing for specific abstract operators that abstract the inference of an inductive hypothesis in iteration are the more powerful, and difficult to design, aspects of abstract interpretation. The design of powerful extrapolation operators pertinent to the degree of approximation of an abstraction is still an art which remains to be much better understood, formalized, and improved. Not enough work has been devoted to the systematic design of extrapolation operators and the fact that they cannot be increasing [146].

5. Semantics

Transition systems [39, 56] and recursive procedures [39, 55] (see also [67]) departed from the initial consideration of flowcharts as an intermediate program representation [49–51]. Similarly for forward/backward transformers and fixpoint duality (see the figure on page 241) prefiguring the μ-calculus. Which semantics to be preferred, small-step operational [39, 56], denotational [148, 180], big-step natural [205], etc became subject to controversies [206]. The solution was ultimately to relate all of them by abstract interpretation.

Another alternative is to design a meta-language to define semantics and to design a general purpose abstract interpreter taking semantics as a parameter, either that of the program to be analyzed or that of a programming language [42, 152, 168].

6. Hierarchies of semantics

The choice of a standard (or collecting) semantics may look fundamental in the design of abstract interpretations. On one hand, the most concrete/collecting semantics dictates an initial specific abstraction specifying the most precise expressible properties. If the collecting semantics is too abstract, it has a limited expressive power which also limits the scope of the static analysis/verification methods that can be formally derived by abstract interpretation of this collecting semantics. For example, with the big-step natural semantics [205], it is possible to abstract into a partial correctness axiomatic semantics but not to reason on (non-)termination. On the other hand, if the collecting semantics is too precise, then further abstractions are more complex to express (in that they result in more complex formulae in the Galois connection calculus [82]).

Nevertheless, the importance of the choice of a universal standard/collecting semantics should be greatly relativized. The reason is that all known semantics are completely defined by an abstraction of the concrete operational trace semantics and form a hierarchy of semantics where all known semantics are recovered by abstractions [64] (some being Galois isomorphisms like between natural, angelic, or respectively demonic denotational, relational, predicate transformer, and axiomatic semantics). An abstraction based on any semantics in the hierarchy will be comparable to any other abstraction based on a different semantics in the hierarchy by considering the greatest lower bound in the hierarchy, which is the most abstract semantics more precise than these two semantics.

This hierarchical semantic approach is valid for transition systems [46] (including transfinite behaviors [120]), grammars [80], resolution-based languages like Prolog [61, 88], higher-order functional languages [67, 78], etc.

7. Combinations of semantics

Sometimes, the collecting semantics cannot be expressed as a single fixpoint but as a combination of fixpoints. An example in [39] is the intersection of forward and backward reachability. In that case, the best abstraction is not the trivial intersection of abstract forward and backward abstract reachability but an iteration of forward and backward abstract reachability analyzes [39, 61]. This generalizes to more complex combinations of fixpoints as found e.g. in temporal logic [94, 158, 159]. This point is not well understood when reasoning on (abstract) models rather than on properties and their different levels of abstraction.

8. Proof methods, verification and inference

Proof methods all consider an abstraction of the program semantics and so can be explained by abstract interpretation using basic induction principles such as fixpoint/post-fixpoint induction.

The essential difference between computer-aided verification and static analysis is that in verification an inductive argument is assumed to be provided by the end-user (e.g. loop invariants that must be expressed in some logical language) whereas in static analysis, this same inductive argument must be automatically inferred (and expressed in some combination of abstract domains). Notice that incompleteness in proof methods [41] follows from the abstraction of the invariant expressible in the collecting semantics which may be inexpressible in the abstract domain defined by the logical language [91]. Inference is of course more difficult than verification, for which the only problem is to prove implications, not to guess what to prove. Nevertheless, verification has rapidly difficulties to scale up when the size of the information to be provided by the end-user is exponential in the size of the program [219].

9. Over- and under-approximation

The theory of abstract interpretation is usually only presented for over-approximations. An over-approximation means that all executions (and maybe a few spurious ones) have to be considered. An example is automatic reachability analysis with invariant inference [52]. Right at the origin of the theory, under-approximation was considered order or complement dual so there is no need to rephrase...
the dual theory \cite{52}. An under-approximation means that some executions (and no spurious ones) have to be considered. However it is significantly more difficult in practice. An example is the automatic inference of necessary preconditions \cite{83} (whereas an over-approximation would yield sufficient preconditions).

10. Abstract domains

In abstract interpretation, properties are sets of objects with that property \cite{56}. For example, the collection semantics is the singleton whose element is the standard semantics. So a concrete program property is the set of its possible semantics. If executions are traces, non-deterministic semantics are sets of traces, so properties are sets of sets of traces (later called “hyper-properties”). So concrete properties can be expressed as sets of sets of objects formalizing the effect of an execution (e.g. traces). This point of view unifies various versions of abstract interpretation which were first thought to be different because they abstract different kind of semantics (operational, denotational, predicate transformers, axiomatic).

Abstract domains encode a subset of the possible properties of a program/system/etc. They are (pre-)ordered by a relation which encodes logical implication. Because of iteration/recursion, they must also encode some notion of limit (of computations, including unbounded or infinite non-terminating ones). Such limits are often expressed as least upper bounds of increasing chains for a computational order, which is often, but not always, the same as the logical order. Strictness analysis \cite{180} was the first example where the logical and computational orders differ, but iterative fixpoint computation and acceleration results extend naturally by considering different ways of passing to the limit \cite{66,67}.

It is quite remarkable that abstract domains are completely determined by the concrete properties and a Galois connection. This leads to the idea of a Galois connection calculus \cite{82}, which allows to specify the essence of static analysis/verification methods by defining the considered semantic domain and abstraction with a single formula of the calculus. Another consequence is that abstract domains are complete lattices since Galois connections preserve existing least upper bounds.

11. Refinement of abstract domains

The refinement of an abstract domain consists in adding the minimal new abstract properties to the domain so as to be able to express exactly in the refined domain some of the concrete properties which had to be strictly over-approximated in the original domain.

The first example was the disjunctive completion \cite{56,102.116} of an abstract domain \(A\): this is a refinement \(\rho(A)\) of this abstract domain \(A\) adding to the disjunctive completion \(\rho(A)\) all the concrete disjunctions of elements of \(A\) originally missing in \(A\) (hence which had to be strictly over-approximated in \(A\)). It is the most abstract domain that is exact on disjunctions of abstract properties in \(A\). It was used in \cite{56,102.10.2} to prove that the merge-over-all paths data flow analysis can be expressed in fixpoint equational form \cite{65}. Similarly, the complementation \cite{34} adds complements that are missing in the abstract.

The transformer/fixpoint completion of an abstract domain is a program-dependent refinement which minimally enriches this abstract domain by adding to the abstract domain all the concrete properties that would make the abstract transformer/fixpoint precise (so that all counter-examples are considered altogether \cite{124}). Various weaker notions of transformer/fixpoint completion can also be considered, e.g. ignoring termination \cite{45}, relative to an observation of program behaviors \cite{6}, or complete with respect to unification for resolution-based concrete semantics \cite{128}.

12. Combinations of abstract domains

The complex design of abstract domains/Galois connections is better performed compositionally, starting from basic abstract domains which are then refined (Sect. 11), combined (this Sect. 12), and composed to get very expressive abstractions \cite{56,10.1}, \cite{82}. This consists in defining a functor on abstract domains/Galois connections than takes existing domains as parameter and creates a new one.

The most popular abstract domain functor is certainly the reduced product \cite{56,10.1}, and its numerous variants (such as the more economical \cite{33,85} and scope extensions \cite{114,215}). The reduced product functor creates a new domain out of existing domains that essentially performs the conjunction of the information carried by each domain by transferring commonly expressible information between domains. It is the most abstract domain which is more precise than the given domains. A recent example is the understanding of Nelson-Oppen combination of logical theories with equality in SMT solvers as a reduced product \cite{92,96} where the commonly expressible information between domains are uninterpreted in/equals between variables.

The reduced power functor \cite{56,125,10.2} is used to express conditional abstract properties that rely upon abstract conditions, as e.g. in an analysis by cases. The reduced power functor has many more specific instances such as polyvariance for context-sensitive analyzes \cite{130} or trace partitioning \cite{137,202} where a local invariant at a program point may depend upon an abstraction of the history of computations reaching that point.

13. Equational design of abstract domains

The refinement (Sect. 11) and combination (Sect. 12) functors on abstract domains can be used to recursively define abstract domains equationally \cite{115,123,209}. This has been exploited for semantics \cite{122} but remains to be exploited in static analysis e.g. in the automatic design of static analyzers, which is the objective of the Galois connection calculus \cite{82}.

14. Galois connections for best abstraction

If the concept of Galois connection is central to abstract interpretation, it is equivalent to many other formalisms including closure operators \cite{39,56}, Moore families and principal ideals, \cite{56}, soundness/logical relations \cite{182}, topologies \cite{55,208}, etc which noted in \cite{71}. Moreover the finiteness hypothesis to compute the solution of the data flow equations using a fixpoint checker eliminate trivial infinite but Noetherian analyses such as constant propagation. The correctness argument in \cite{151} is based upon a distributivity hypothesis not satisfied by the main example of the paper which is constant propagation. Other precursors \cite{183,184,211} understood (non-relational) static analysis as a specification calculus on non-standard values or (relational) static analysis as recursive definitions \cite{199} i.e. sound by definition. For example the sign analysis in \cite{211} is unsound in that negative \(\ominus\) is interpreted as \(< 0\) and positive \(\oplus\) as \(\geq 0\) so that \(-1 \times 0 = 0\) is a counter-example to the classical rule \(\ominus \times \ominus = \ominus, 0\) has no best abstraction and is handled specifically in mathematics by simplification rules. Intuition is not always right.
can express the concept of best abstraction: any concrete property has a most precise approximation in the abstract [56]. An interesting consequence is that the analysis/verification in the abstract is completely determined by the collecting semantics and the abstraction (nothing else has to be invented apart from convergence acceleration operators). This leads to the idea of calculational (even automatic) design [44, 210], or at least automatic certification of analyzers/verifiers (such as [44] and its extensions which has been formally checked in Coq [18]).

15. In absence of best abstraction

It was recognized early that in many cases there is no best abstraction. For example in [95], there is no best/smallest, even no minimal polyhedron enclosing a disk. Then only half of the Galois connection can be used [62], the most popular one being a concretization function only [95]. Because Galois connection preserves arbitrary joins the power set structure of concrete properties is preserved as complete lattices in abstract domains. This is no longer the case in absence of best abstraction, in which case joins/meets, even finite ones may not exist (so e.g. [151, 221] are no longer applicable). This can be compensated by using widening/narrowing 7. Another example is the abstraction of languages by grammars [69] or of properties by a logical theory [91] where there is no best abstraction. In that case properties and fixpoint transformers may have infinitely many sound abstractions without a best/most precise one and so arbitrary choices have to be made in the abstract.

16. Abstraction of syntax

The syntax of languages as described by grammars provide varied examples of abstractions. Grammars can be given a fixpoint operational semantics of derivations which abstractions yield e.g. the Chomsky-Schützenberger fixpoint characterization of the language generated by the grammar, as well as top-down and bottom-up parsers and numerous algorithms used in compilation [75, 77, 80]. This generalizes to resolution-based languages [88] where essentially, text substitution is replaced by unification.

17. Syntactic abstractions

By syntactic abstraction we mean an abstraction of a collecting semantics to a meta-formalism describing languages, such as regular expressions, context-free grammars [69], or even theories in logic [93]. Typically, for syntactic abstractions, unless a property is exactly describable in the meta-formalism, there is no best abstraction. For example, if an over approximation of non-regular language of finite unbounded strings on a finite alphabet is described by a regular expression, there exists a sentence, of say length n, deriving from the regular expression which is not in the non-regular language to be over-approximated. A better description of the non-regular language consists in the regular expression enumerating the finitely many sentences of length less than or equal to n, and over approximating only those of greater length. There is no best and even minimal abstraction since, at least in theory, n can be indefinitely increased to get better and better abstractions. Of course the description of the concrete collecting semantics must not be subject to such expressivity restrictions which explains that abstract interpretation describes concrete domains using set/model theory rather than proof theory for a given first-order logic. Category theory would also be a mathematically reasonable choice, although it might not help in practice and would considerably restrict the readership [1].

18. Abstraction of programs versus languages, and the power of extrapolation operators

That verification/analysis formal method have to be designed for programming languages and not for checking a given program (or a given model) makes a significant difference by adding a universal dependence on all programs of the language. So an abstract domain collects the part of the abstraction which is common to all programs and must be instantiated when analyzing a specific program. A first example is set-based analysis which is often presented as acting on an infinite domain, which is true for the language, whereas a finite subset only is used for any given program. A consequence is that the set-based constraint-solving methods turned out to be mere finite fixpoint iterations in a finite domain [69].

A second example is on the use of Noetherian/finite or non-Noetherian/infinite domains. The case of finite abstractions is easy since always equivalent to predicate abstraction and reciprocally [47]. On the contrary, abstract interpretation stresses infinite non-Noetherian abstractions of infinitary properties. For a given program, a complete finite abstraction does exist for proving any desired property of this particular program [45]. In fact finding this abstraction or the proof is mathematically equivalent (and undecidable except e.g. for finite states). This is no longer the case for programming languages, for which any finite abstraction, or finite refinement [94], will fail or not terminate on infinitely many programs for which an infinite abstraction would have succeeded [45, 60]. This means that the relevant abstraction can only be determined during the analysis, which is the rôle of extrapolation operators.

Moreover, in practice, a finitude hypothesis is not of much help, since it does not prevent the combinatorial explosion of the set of cases to be considered, so that extrapolation operators must be used (or replaced by specific fixpoint computation methods when they exist). Of course trivial extrapolation operators (like any execution is to be over-approximated by chaos after a given number of steps) lead to very poor results, which are inconclusive but of a finite number of cases. Others like Milner’s idea that all instances of a recursive procedure should be the same polymorphic type, which is a widening [43, 177], is quite satisfactory in practice.

19. Temporal abstraction

Although initially designed for reachability analysis [51], abstract interpretation applies sensu stricto to temporal properties, including time-symmetric trace-based ones referring both to the past, present, and future [71] (for which incompleteness results can be proved even for finite systems [126]).

20. Languages

Almost all families of programming languages (imperative languages [52], functional languages [180], Prolog [13, 26, 104, 142, 178, 179], constraint logic programs [119, 142], constraint solvers [189], database query languages [35], object-oriented languages [154], multithreaded programs [112], byte code [12], machine code [108, 109, 153, 222], etc.), but also specification languages [129], grammars [175], algebraic polynomial systems [70], graph rewriting systems [138], logics [5], games [141], synchronous languages [135], continuous systems [37], hybrid automata [140], quantum computing [190], etc. have been subject to static analysis by abstract interpretation.

21. Control-flow analysis

Besides data, static analysis by abstract interpretation should take into account the control structure of programs.

Control-flow (closure, escape, binding-time, etc.) analysis are an abstract interpretation [165–167, 187].
22. Parallelism
Parallelism was considered early as a challenge to scalability [57, 58, 97]. One important aspect was which collecting semantics should be considered, otherwise stated which semantic model and proof methods should be abstracted. It turned out that the proof methods known at the time were just different abstractions of the same interleaving semantics [97]. Unfortunately the interest of funding studies of parallelism faded in the late eighties because the speed of sequential processors was progressing spectacularly. But any exponential process must stop by exhaustion of resources, which is the case nowadays, so the interest in parallelism is revived, by necessity. Fortunately, principles are immutable, so that more recent proof methods such as assume-guarantee, are also, like all proof methods, abstract interpretations, which can serve as an effective basis for designing scalable analyzers of parallel programs [172, 173]. Other aspects of parallelism are scheduling [132], parallelization of sequential or non-deterministic programs, where abstract interpretation can be used to take semantic aspects into account [25, 217].

23. Types
Type theory that developed around operational semantics, subject reduction, type inference by unification and typing rules, long appeared hardly comparable to abstract interpretation. This is not the case since [43] showed that polymorphic types and type inference are abstract interpretations with widening (to exclude recursive calls of functions with infinitely many different principle types). The approach was based on a denotational semantics, which was intentionally unusual but not surprising in the hierarchy of semantics [46]. The same approach applies to other unification-based analyzers [3].

24. Binary abstraction and hardware analysis
Binary abstractions abstract infinite sets of sequences with finite elements, sometimes of bounded length hence finite but still very large which requires widening for scalability [161, 171]. A trivial abstraction where “0 or 1” cases are abstracted into $\top$ is used for hardware analysis [223] and is one of the rare formal methods that scales on complex hardware. A more sophisticated abstract domain is in Astrea used to analyze integers and floats in their machine representation [171].

25. Numerical abstractions
Numerical abstractions abstract infinite sets of numerical vectors of finite dimension. There are numerous examples, including convex abstractions such as the non-relational intervals [52], and the relational linear inequalities [95], linear interval inequalities [30, 31], zonotopes [134] and a few non-convex abstractions such as linear absolute value inequalities [32]. Very good compromises such as octagons [170], pentagons [156], or parallelotopes [7] provide both scalability and enough precision for most common situations (like buffer- overrun checking in case the array bounds are symbolic).

26. Symbolic abstractions
Symbolic abstractions abstract infinite sets of in/finite functions, graphs, etc. as found in most programming languages such as arrays [89], trees [162], infinitary relations [163], and heaps (like non-sharing [54, 55] or shape analysis [29, 198]). This research area where directly reusable mathematics essentially does not exist is quite important and still in infancy (except in a few easier particular cases, such as linear lists).

27. Simulations
(Bi-)simulations are abstract interpretations [194]. This new understanding has lead to more efficient simulation [196] and partitioning [195] algorithms.

28. Probabilistic abstractions
Probabilistic abstraction is another relatively underdeveloped application domain of abstract interpretation [2, 96, 99, 176, 204]. The reasons are that probabilistic behaviors are often only known with imprecision, that they mix with non-determinism on data, and that inference for infinite systems is much more complex than enumeration in the finite case since very subtle extrapolations are requested for precision.

29. Program transformation
The main applications of abstract interpretation beyond static analysis (e.g. to automatically infer loop invariants), are program transformations such as partial evaluation [146], abstract debugging [22, 40] abstract slicing [201], parallelization (see Sect. 22), compilation [200], etc.

Program transformations can be formalized as an abstraction, where a syntactic transformation of a source program is understood as first performed on the source program semantics to get a generally more efficient although equivalent transformed semantics which is then abstracted into the object program [74]. This is another example of syntactic abstraction (of the transformed semantics into the object program). More examples include partial evaluation [74], dead-end elimination [72], and refactoring [90].

30. Termination
Termination is an abstract interpretation [81] where the transfinite variant functions abstract in each program state what remains to be executed before potential or definite termination. This termination abstraction yields known proof methods [59, 65, 97] while further abstractions with widenings (e.g. piecewise linear [218] or ordinal-valued [218]) or policy iteration ([160]) yield static analyzers automatically inferring abstractions of variant functions. Another aspect of time abstraction is the determination of bounds on the worst-case execution times of programs [108, 109, 222].

31. Modularity
Static analyzers are complex programs, certainly much more complex than compilers, so they must be modular and extensible. The design by combination of abstractions is fundamental for scalability and extensibility [85]. Another aspect is the modular analysis of programs by parts [73]. The sound analyzers can be classified into the ones that perform a global program analysis (like Astrea [16, 17, 84]) which are precise but maybe costly and an analysis by part (like cccheck [107]) where cheap analyzes can be performed at compile time, but may require user-interaction to design code contracts (an activity which may also be automatized [83]); but not loop invariants which in all cases are inferred automatically.

32. Generalist versus domain-aware static analyzers
Sound generalist analyzers like Polyspace Code Prover\textsuperscript{TM} by MathWorks\textsuperscript{TM}\textsuperscript{TM}, Global Surveyor [220], Checkmate [113], CodeSonar for machine code [197], or Julia [188] produce good results for a large variety of codes in a given programming language. Most often they lack extreme precision. Using heuristics to sort out proliferating
false alarms may be a fake commercial argument but does not have the rigor necessary to verify very large safety or security critical software.

This leads to the idea of analyzers aware of an application domain such as the precision of floating-point computations for Fluctuat [102, 133], control-command for Astrée [16, 17, 84]. In addition to generalist abstractions, filters, integrators, relation of variables to the clock, etc must be handled with greater precision thanks to specifically adapted abstract domains [110, 111]. Initially designed for avionic software, Astrée [16, 17, 84] had to be extended with a non-linear abstract domain to analyze precisely quaternions used for satellite positioning [20].

33. Industrial applications

The use of a principled approach to static analysis based on abstract interpretation allow for the rigorous and incremental design of precise and scalable static analyzers, with very few false alarms.

Astrée [14, 16, 17, 48, 84, 85, 87, 150] is a static analyzer commercialized by AbsInt and used in the medical, transportation, and communications industry for analyzing the absence of runtime errors in control command software, e.g. to control medical monitoring equipment, planes or satellites. This is certainly one of the first completely automatic tools based on formal methods that did scale up with enough precision to allow, after avoiding a few potentially catastrophic software failures, to become voluntarily mandatory in the design of e.g. avionic safety critical systems (see [103, 213] for reports on preliminary experiences).

ccheck [107, 155] is an example of general purpose, modular, precise, and efficient analyzer relying on the use of abstract code contracts. Being based on an intermediate language used by compilers it is applicable to several different programming languages that compile to this intermediate language. This is a possible approach to cope with the proliferation of programming languages. Other complementary approaches include the implementation of reusable libraries of abstract domains, like APRON [144] or the Parma library [9] for numerical properties.

34. Security

Security is typical of computer science lack of responsibility with respect to customers. One can leave a door open and take no responsibility at all when a burglar comes in, whereas this would be excluded by a restriction clause in any theft home insurance. This situation cannot go on forever, and the computer industry will have to take responsibility for its provable errors. Such a proof of error can be given by a counter-example. Better, a proof of absence of error can be given by a tool proving the absence of error, provided the tool is sound and precise (otherwise it could be qualified of time-consuming and too restrictive, as for security types). Fully automatic tools are already able to catch such security bugs by taking values and objects into account (e.g. [216]), at least the common ones, but this is not yet a current practice. Many security properties cannot be checked at runtime so static analysis and verification tools have an outstanding domain of application.

Other areas in security where the semantic-based approach of abstract interpretation is much more powerful than syntactic matching or type inference is abstract non-interference [121], steganography [76], obfuscation [191, 192], and malware detection [193].

35. Unexpected applications

35.1 Biology

We anticipated in [39, Ch. 3] that basing abstract interpretation collecting semantics on transition systems “is obviously very general. It applies not only to computer systems but also to economic or biological systems, provided that the model of the system to study evolves according to discrete time steps.”. In biology, it took some time before abstract interpretation was applied in practice in the context of rule-based modeling of protein-protein interaction networks [28, 100, 101]. The widening for economy was a bit too optimistic!

35.2 SAT and SMT solvers

The fact that SAT and SMT solvers are abstract interpretations of the semantics of logics came more as an unanticipated surprise [23, 24, 105, 106, 214]. This understanding yields new decision algorithms [136, 214].

36. Conclusion

Any semantic based-program manipulation, analysis, and verification in the scope of abstract interpretation theory can also be designed in the form of a specific paradigm/method/algorithm, illustrated by examples, and justified by an informal reasoning. This is the usual problem-specific approach of computer science which leads to a proliferation of disparate results lacking a unifying point of view. The consequence is that only one or two paradigms/methods/algorithms/examples are taught, remembered, and later extended, which completely inhibits the creative power. This empirical approach will certainly disappear over a long period of time in favor of more abstract foundations. This took centuries in mathematics were, by combining fundamental ideas, one reaches a definite result. This is indispensable to make computer science teachable beyond tricky know-how rapidly becoming outdated in the short term. In particular, it is not possible to assimilate, understand, and compare thousand of approaches to verification or inference in static analysis without a reduction to a few abstract concepts, explaining concisely how they can be (re-)designed. This is the objective of abstract interpretation, and requires a perpetual evolution, renewal, and rethinking of the foundations and practice, with very few basic principles remaining stable. Principled computer science is the only path to cost-effective safety and security.

Acknowledgments

Our acknowledgments go to the many contributors to abstract interpretation including those who are not referenced here by obvious lack of exhaustivity. They also go to Roberto Giacobazzi and Francesco Ranzato for their comments.

References


