Complete Abstract Interpretations Made Constructive

Roberto Giacobazzi[1], Francesco Ranzato[2], and Francesca Scozzari[1]

[1] Dipartimento di Informatica, Università di Pisa, Italy
{giaco,scozzari}@di.unipi.it
[2] Dipartimento di Matematica Pura ed Applicata, Università di Padova, Italy
franz@math.unipd.it

Abstract. Completeness is a desirable, although uncommon, property of abstract interpretations, formalizing the intuition that, relatively to the underlying abstract domains, the abstract semantics is as precise as possible. We consider here the most general form of completeness, where concrete semantic functions can have different domains and ranges, a case particularly relevant in functional programming. In this setting, our main contributions are as follows. (i) Under the weak and reasonable hypothesis of dealing with continuous semantic functions, a constructive characterization of complete abstract interpretations is given. (ii) It turns out that completeness is an abstract domain property. By exploiting (i), we therefore provide explicit constructive characterizations for the least complete extension and the greatest complete restriction of abstract domains. This considerably extends previous work by the first two authors, who recently proved results of mere existence for more restricted forms of least complete extension and greatest complete restriction. (iii) Our results permit to generalize, from a natural perspective of completeness, the notion of quotient of abstract interpretations, a tool introduced by Cortesi et al. for comparing the expressive power of abstract interpretations. Fairly severe hypotheses are required for Cortesi et al.'s quotients to exist. We prove instead that continuity of the semantic functions guarantees the existence of our generalized quotients.

1 Introduction and Motivation

Within the classical and widely adopted Cousot and Cousot framework for approximating generic semantic definitions [7,8], it is well known that *completeness* for an abstract interpretation is a much richer property than plain mandatory soundness. In fact, roughly speaking, a complete abstract interpretation turns out to be as precise as possible, relatively to its underlying abstract domains where approximate computations are encoded. This simple intuition explains why, although being a rather uncommon property in practice, notably in static program analysis, completeness is a highly desirable feature for an abstract interpretation, especially in abstract model checking (indeed, some authors arguably term it "optimality"). Examples of complete abstract interpretations can be found, e.g., when comparing algebraic polynomial systems [10] and program semantics [9].

Luboš Prim et al. (Eds.): MFCS'98, LNCS 1450, pp. 366–377, 1998.

In recent years, there has been a number of papers dealing with various theoretical issues related to completeness in abstract interpretation (cf. [6,12], [16,17,18,20]). Among them, Giacobazzi and Ranzato's paper [12] points out that completeness for an abstract interpretation only depends on the underlying abstract domain, and therefore is an abstract domain property. In view of this basic observation, the following problem is then considered: Given an abstract interpretation with underlying abstract domain A, do there exist the least extension and the greatest restriction of A making the whole abstract interpretation complete? Giacobazzi and Ranzato [12] give an affirmative answer, by showing that greatest complete restrictions (called *complete kernels*) always exist, and, for continuous concrete semantic operations, *least complete extensions* exist as well. According to [11], these two operators on abstract domains are, resp., instances of generic abstract domain simplifications and refinements. Following the standard notation, let us denote resp. by $\alpha_{X,Y}$ and $\gamma_{Y,X}$ the abstraction and concretization maps for a concrete domain X and an abstract domain Y. In [12], given a semantic operation $f : C^n \rightarrow C$, an abstract interpretation $I = \langle A, f^\sharp \rangle$, with $f^\sharp : A^n \rightarrow A$, is complete w.r.t. $\langle C, f \rangle$ when $\alpha_{C,A} \circ f = f^\sharp \circ \alpha_{C^n,A^n}$. Thus, functions of generic type $C \rightarrow D$, occurring frequently in denotational semantics for functional programming, cannot be handled. Moreover, Giacobazzi and Ranzato's results, in general, only prove the existence of least complete extensions and complete kernels, and give a constructive iterative methodology for obtaining least complete extensions only when the semantic operations are additive. However, additivity is a fairly restrictive hypothesis to be widely applicable in practice. By contrast, the present work deals with the most general formulation of completeness for abstract interpretations – no hypothesis on the type of semantic functions is assumed – and fully solves the limitations of Giacobazzi and Ranzato's approach, in particular on the side of complete domain construction.

Let us explain more in detail the general approach pursued in this paper. Firstly, given any concrete domain C, we denote by $\mathcal{L}_C$ the so-called *lattice of abstract interpretations* of C [7,8]. Let $f : C \rightarrow D$ be any concrete semantic function occurring in some complex semantic specification, and assume that an abstract semantics is given by $f^\sharp : A \rightarrow B$, where $A \in \mathcal{L}_C$ and $B \in \mathcal{L}_D$. The concept of soundness is standard and well-known: $\langle A, B, f^\sharp \rangle$ is a sound abstract interpretation – or $f^\sharp$ is a correct approximation of f relatively to A and B – when $\alpha_{D,B} \circ f \sqsubseteq f^\sharp \circ \alpha_{C,A}$ ($\sqsubseteq$ denotes pointwise ordering). On the other hand, $\langle A, B, f^\sharp \rangle$ is complete when equality holds, i.e. $\alpha_{D,B} \circ f = f^\sharp \circ \alpha_{C,A}$. Since $\alpha_{D,B} \circ f \sqsubseteq f^\sharp \circ \alpha_{C,A} \Leftrightarrow \alpha_{D,B} \circ f \circ \gamma_{A,C} \sqsubseteq f^\sharp$, the canonical best correct approximation $f^{b_{A,B}} : A \rightarrow B$ of f relatively to the abstract domains A and B is defined by $f^{b_{A,B}} \stackrel{\text{def}}{=} \alpha_{D,B} \circ f \circ \gamma_{A,C}$. In this scenario, the following observation still holds: Given A and B, there exists $f^\sharp$ such that $\langle A, B, f^\sharp \rangle$ is complete iff $\langle A, B, f^{b_{A,B}} \rangle$ is complete. This means that, even in this general context, *completeness is an abstract domain property*, and gives rise to the question whether abstract domains can be minimally refined and/or simplified so that completeness is achieved. Let us give a simple example concerning Mycroft's strictness analysis for functional programs [3,15]. Consider the following function F of type $\mathsf{Nat} \times \mathsf{Nat} \rightarrow \mathsf{Bool}$:

$$F(\langle x,y\rangle) \overset{\text{def}}{=} if\ (x=3\ and\ y=3)\ then\ true\ else\ \bot$$

Following Burn et al. [3], from F one gets in the most natural way its denotational "collecting" semantics $f : \mathbf{P}(\mathbb{N}_\bot \times \mathbb{N}_\bot) \to \mathbf{P}(Bool_\bot)$, where $\mathbf{P}$ is the Hoare powerdomain operator and $\bot$ denotes undefinedness (i.e., both nontermination and error). Let $S = \{0 < 1\}$ be the basic strictness domain, abstracting both $\mathbf{P}(\mathbb{N}_\bot)$ and $\mathbf{P}(Bool_\bot)$, and such that $S \times S$ abstracts $\mathbf{P}(\mathbb{N}_\bot \times \mathbb{N}_\bot)$. Concretization and abstraction maps are the usual ones, e.g. $\gamma(\langle 0,0\rangle) = \{\langle\bot,\bot\rangle\}$ and $\gamma(\langle 0,1\rangle) = \{\langle\bot,x\rangle \mid x \in \mathbb{N}_\bot\}$. Then, the best correct approximation $f^b : S \times S \to S$ of f is as follows: $f^b = \{\langle 0,0\rangle \mapsto 0,\ \langle 0,1\rangle \mapsto 0,\ \langle 1,0\rangle \mapsto 0,\ \langle 1,1\rangle \mapsto 1\}$. Clearly, f^b is not complete: For instance, $\alpha(f(\{\langle\bot,\bot\rangle,\langle 4,5\rangle\})) = \alpha(\{\bot\}) = 0$, whilst $f^b(\alpha(\{\langle\bot,\bot\rangle,\langle 4,5\rangle\})) = f^b(\langle 1,1\rangle) = 1$. These phenomena of incompleteness in strictness analysis are analyzed in depth in [17,18], which, however, do not investigate the issue of achieving completeness by minimally modifying the abstract domains. Moreover, because the range and domain of f are different, the method of [12] is not applicable here. Instead, the methodology proposed here allows to constructively derive the least extension $\mathcal{E}(S \times S)$ of $S \times S$ which induces a complete abstract interpretation. It should be clear that by adding a point to $S \times S$ which is able to represent the information that the first and second components are surely not simultaneously equal to $3 \in \mathbb{N}_\bot$, one gets a domain inducing a complete abstract interpretation. Indeed, our methodology allows to constructively derive that $\mathcal{E}(S \times S) = (S \times S) \cup \{\langle\neq,\neq\rangle\}$, where $\gamma(\langle\neq,\neq\rangle) = (\mathbb{N}_\bot \times \mathbb{N}_\bot) \setminus \{\langle 3,3\rangle\}$. In this way, one gets a best correct approximation $f^{b*} : \mathcal{E}(S \times S) \to S$ such that $f^{b*}(\langle\neq,\neq\rangle) = 0$, and therefore completeness has been achieved.

Let us illustrate the main contributions of the paper. In Section 3, the concept of completeness is formalized by resorting to the Cousot and Cousot *closure operator* approach to abstract interpretation [5,8]. This allows us to be independent from specific representations of abstract domain's objects. It is shown that completeness is an abstract domain property, which gives rise to a mathematically compact equation between closures, studied in later sections. Moreover, we observe that if an abstract interpretation $f^\sharp : A \to B$ is complete, and therefore $f^\sharp = f^{b_{A,B}}$, then for all the abstract domains A' more concrete than A and B' more abstract than B, it turns out that $f^{b_{A',B'}} : A' \to B'$ is still complete. This implies that it is not meaningful to search for the complete kernel of A and the least complete extension of B, because, e.g., if the complete kernel of A would exist then A itself would already be complete. Instead, one should try to solve the converse problems. Under the working hypothesis of dealing with continuous semantic functions, a key constructive characterization of the domains inducing complete abstract interpretations is given in Section 4. More precisely, given a continuous semantic function $f : C \to D$, we show that $f^{b_{A,B}} : A \to B$ is complete iff A is more concrete than a certain domain $R_f(B)$ depending on B iff B is more abstract than a certain domain $L_f(A)$ depending on A. Thus, the mappings $L_f : \mathcal{L}_C \to \mathcal{L}_D$ and $R_f : \mathcal{L}_D \to \mathcal{L}_C$ form an adjunction. By exploiting these results, we are able to characterize: (1) the least complete extension of A relative to B as the least domain which contains both A and $R_f(B)$, and (2) the complete kernel of B relative to A as the greatest domain contained in both B and $L_f(A)$. As a further consequence, we subsume the more restrictive notions of least complete extension and greatest complete restriction studied in [12] and the corresponding results of existence as well as the constructive characterization given for additive semantic functions. In Section 5, we investigate the relationship between completeness and the concept of *quotient* of an abstract interpretation, recently introduced by Cortesi et al. [4] for comparing the precision of abstract interpretations in computing a given property. Informally, the quotient of a complex abstract domain A w.r.t. a property P of A (i.e., a further abstraction of A) represents which part of A contributes in computing the property P. We show that, in general, Cortesi et al.'s quotients do not always exist: In particular, the basic assumption of continuity of the semantic functions does not ensure their existence. However, we observe that quotients, when they exist, turn out to be certain least complete extensions, which naturally formalize the intuition behind the notion of quotient. Thus, a simple and natural generalization of the notion of quotient is proposed, which retains the advantage of being always well-defined, under the hypothesis of continuity of the semantic functions.

2 Preliminaries

Basic Notation. If S is any set, P a poset, and $f, g : S \rightarrow P$ then we write $f \sqsubseteq g$ if for all $x \in S$, $f(x) \leq_P g(x)$. If $S \subseteq P$ then $max(S) \overset{\text{def}}{=} \{s \in S \mid \forall t \in S.\, s \leq t \Rightarrow s = t\}$. Given two posets C and D, $C \xrightarrow{\text{m}} D$, $C \xrightarrow{\text{c}} D$, and $C \xrightarrow{\text{a}} D$ denote, resp., the set of all monotone, continuous (i.e. preserving lub's of chains), and (completely) additive (i.e. preserving all lub's, empty set included) functions from C to D. ω denotes the first infinite ordinal. For a complete lattice C, given $f : C \rightarrow C$, for any $i \in \mathbb{N}$, the i-th power $f^i : C \rightarrow C$ of f is inductively defined, for any $x \in C$, as x if $i = 0$, and as $f(f^{i-1}(x))$ if i is a successor.

The Lattice of Abstract Interpretations. In standard Cousot and Cousot's abstract interpretation theory, abstract domains can be equivalently specified either by Galois connections, i.e. adjunctions, or by closure operators (see [5,8]). In the first case, the concrete domain C and the abstract domain A are related by an adjunction (α, C, A, γ). It is generally assumed that (α, C, A, γ) is a Galois insertion (GI), i.e. α is onto or, equivalently, γ is 1-1. In the second case instead, an abstract domain is specified as an (*upper*) *closure operator* (shortly uco or closure) on the concrete domain C, i.e., a monotone, idempotent and extensive operator on C. These two approaches are equivalent, modulo isomorphic representations of domain's objects. In the following, $\langle uco(C), \sqsubseteq\rangle$ denotes the poset of all uco's on C. Let us recall that each $\rho \in uco(C)$ is uniquely determined by the set of its fixpoints, which is its image, i.e. $\rho(C) = \{x \in C \mid \rho(x) = x\}$, and that $\rho \sqsubseteq \eta$ iff $\eta(C) \subseteq \rho(C)$. Also, when $\langle C, \leq, \vee, \wedge, \top, \bot\rangle$ is a complete lattice, $\langle uco(C), \sqsubseteq, \sqcup, \sqcap, \lambda x.\top, \lambda x.x\rangle$ is a complete lattice, and $X \subseteq C$ is the set of fixpoints of a uco iff X is meet-closed, i.e. $X = \mathcal{M}(X) \overset{\text{def}}{=} \{\wedge Y \mid Y \subseteq X\}$ (where $\wedge\emptyset = \top \in X$). Moreover, given $\rho \in uco(C)$, $\langle \rho(C), \leq\rangle$ is a complete meet subsemilattice of C. Hence, for a concrete domain C which is a complete lattice, we will identify $uco(C)$ with the lattice $\mathcal{L}_C$ of abstract interpretations of C, i.e. the complete lattice of all possible abstract domains of C. Often, we will find convenient to identify closures with their sets of fixpoints. This does not give rise to ambiguity, since one can distinguish their use as functions or sets according to the context. The ordering on $uco(C)$ corresponds precisely to the standard order used in abstract interpretation to compare abstract domains with regard to their precision: A_1 is more precise than A_2 (i.e., A_1 is more concrete than A_2 or A_2 is more abstract than A_1) iff $A_1 \sqsubseteq A_2$ in $uco(C)$. Lub and glb of $uco(C)$ have therefore the following reading as operators on abstract domains. Let $\{A_i\}_{i \in I} \subseteq uco(C)$: (i) $\sqcup_{i \in I} A_i$ is the most concrete among the domains which are abstractions of all the A_i's, i.e. it is their least (w.r.t. $\sqsubseteq$) common abstraction; (ii) $\sqcap_{i \in I} A_i$ is the most abstract among the domains (abstracting C) which are more concrete than every A_i; this domain is known as reduced product of all the A_i's.

3 Completeness by Closures

Let $f : C \xrightarrow{\text{m}} D$ be any monotone semantic function, where C and D are complete lattices playing the rôle of concrete semantic domains. Let an abstract interpretation $\langle A, B, f^\sharp \rangle$ of $\langle C, D, f \rangle$ be specified by the GIs $(\alpha_{C,A}, C, A, \gamma_{A,C})$ and $(\alpha_{D,B}, D, B, \gamma_{B,D})$, and by an abstract function $f^\sharp : A \xrightarrow{\text{m}} B$. It is known [8] that $f^\sharp$ is a correct approximation of f, i.e. $\alpha_{D,B} \circ f \sqsubseteq f^\sharp \circ \alpha_{C,A}$, if and only if $\alpha_{D,B} \circ f \circ \gamma_{A,C} \sqsubseteq f^\sharp$. Thus, $f^{b_{A,B}} \overset{\text{def}}{=} \alpha_{D,B} \circ f \circ \gamma_{A,C} : A \rightarrow B$ is called the canonical best correct approximation of f relatively to the abstract domains A and B. $\langle A, B, f^\sharp \rangle$ is called *complete* when $\alpha_{D,B} \circ f = f^\sharp \circ \alpha_{C,A}$. In this case, $f^\sharp = f^\sharp \circ \alpha_{C,A} \circ \gamma_{A,C} = \alpha_{D,B} \circ f \circ \gamma_{A,C} = f^{b_{A,B}}$, i.e. $f^\sharp$ indeed is the best correct approximation $f^{b_{A,B}}$. This means that, given two abstract domains A and B, there exists $f^\sharp$ such that $\langle A, B, f^\sharp \rangle$ is complete iff $\langle A, B, f^{b_{A,B}} \rangle$ is complete. Since $f^{b_{A,B}}$ only depends on A and B, we get that *completeness is an abstract domain property.* Thus, given A and B, we refer to completeness of A and B in order to refer to completeness of the whole abstract interpretation $\langle A, B, f^{b_{A,B}} \rangle$. By using closure operators, if $\rho = \gamma_{A,C} \circ \alpha_{C,A} \in uco(C)$ and $\eta = \gamma_{B,D} \circ \alpha_{D,B} \in uco(D)$ are the uco's associated, resp., with A and B, one can extend an analogous result in [12] by showing that A and B are complete iff $\eta \circ f = \eta \circ f \circ \rho$. This justifies the following general definition of completeness.

Definition 1. Let C and D be complete lattices, $f : C \xrightarrow{\text{m}} D$, $\rho \in uco(C)$, and $\eta \in uco(D)$. Then, the pair $\langle \rho, \eta \rangle$ is *complete* for f if $\eta \circ f = \eta \circ f \circ \rho$. Also, if $F \subseteq C \xrightarrow{\text{m}} D$ then $\langle \rho, \eta \rangle$ is complete for F whenever $\forall f \in F.\ \eta \circ f = \eta \circ f \circ \rho$. □

First, let us notice that, equivalently, one can define $\langle \rho, \eta \rangle$ complete for f when $f \circ \rho \sqsubseteq \eta \circ f$. Further, it is worth remarking that our definition encompasses the case where $f : C \rightarrow C$ and one is interested in two different abstractions of input and output, i.e. $\rho, \eta \in uco(C)$ with $\rho \neq \eta$. Whenever $f : C \rightarrow C$ and $\rho = \eta$, the above definition of completeness boils down to the equation $\rho \circ f = \rho \circ f \circ \rho$ considered in [7,12]. Also, it would not be too difficult (although notationally heavy) to develop the whole theory by considering semantic functions of type $C^n \to D^m$.

For any given set of functions $F \subseteq C \xrightarrow{\mathrm{m}} D$, we will use the following helpful notation: $\Gamma(C,D,F) \overset{\mathrm{def}}{=} \{\langle \rho, \eta \rangle \in uco(C) \times uco(D) \mid \forall f \in F.\ \eta \circ f = \eta \circ f \circ \rho\}$. Whenever $F = \{f\}$, we simply write $\Gamma(C,D,f)$. The following result lists some interesting properties of completeness, where points (i)–(iii) generalize an analogous result given in [12].

Proposition 1.
(i) $\langle \lambda x.x, \eta \rangle, \langle \rho, \lambda x.\top_D \rangle \in \Gamma(C,D,F)$.
(ii) $\forall d \in D.\ \Gamma(C,D,\lambda x.d) = uco(C) \times uco(D)$.
(iii) *If* $\langle \rho, \eta \rangle \in \Gamma(C,D,f)$ *and* $\langle \eta, \mu \rangle \in \Gamma(D,E,g)$ *then* $\langle \rho, \mu \rangle \in \Gamma(C,E,g \circ f)$.
(iv) *If* $\langle \rho, \eta \rangle \in \Gamma(C,D,F)$, $\delta \sqsubseteq \rho$ *and* $\beta \sqsupseteq \eta$, *then* $\langle \delta, \beta \rangle \in \Gamma(C,D,F)$.

Given $F \subseteq C \xrightarrow{\mathrm{m}} D$ and $\eta \in uco(D)$, let us now introduce the following operators transforming abstract domains of C (as usual, we follow the standard conventions $\sqcap \emptyset = \top_{uco(C)}$ and $\sqcup \emptyset = \bot_{uco(C)}$).

- $\mathcal{K}^{\eta}_{F}(\rho) \overset{\mathrm{def}}{=} \sqcap \{\varphi \in uco(C) \mid \rho \sqsubseteq \varphi,\ \langle \varphi, \eta \rangle \in \Gamma(C,D,F)\}$;
- $\mathcal{E}^{\eta}_{F}(\rho) \overset{\mathrm{def}}{=} \sqcup \{\varphi \in uco(C) \mid \varphi \sqsubseteq \rho,\ \langle \varphi, \eta \rangle \in \Gamma(C,D,F)\}$.

Also, given $\rho \in uco(C)$, analogous operators $\mathcal{K}^{\rho}_{F}$ and $\mathcal{E}^{\rho}_{F}$ of type $uco(D) \to uco(D)$ are introduced. Thus, e.g., $\mathcal{E}^{\eta}_{F}(\rho)$ is the least common abstraction of all the domains φ more concrete than ρ and such that $\langle \varphi, \eta \rangle$ is complete for F. As a consequence of Proposition 1 (iv), one can draw the following two important remarks: (i) If $\langle \mathcal{K}^{\eta}_{F}(\rho), \eta \rangle \in \Gamma(C,D,F)$ then $\mathcal{K}^{\eta}_{F}(\rho) = \rho$; (ii) If $\langle \rho, \mathcal{E}^{\rho}_{F}(\eta) \rangle \in \Gamma(C,D,F)$ then $\mathcal{E}^{\rho}_{F}(\eta) = \eta$. This means that it does not make sense to search for the greatest restriction ρ^g of $\rho \in uco(C)$ such that $\langle \rho^g, \eta \rangle$ is complete, and, dually, the least extension η^l of $\eta \in uco(D)$ such that $\langle \rho, \eta^l \rangle$ is complete, because either they coincide with their arguments or they do not exist. That is why we introduce just the following notions.

Definition 2. If $\langle \rho, \mathcal{K}^{\rho}_{F}(\eta) \rangle \in \Gamma(C,D,F)$ then $\mathcal{K}^{\rho}_{F}(\eta)$ is called the *complete kernel* of η relative to ρ. Dually, if $\langle \mathcal{E}^{\eta}_{F}(\rho), \eta \rangle \in \Gamma(C,D,F)$ then $\mathcal{E}^{\eta}_{F}(\rho)$ is called the *least complete extension* of ρ relative to η. □

As far as complete kernels are concerned, it is an easy task to show that they always exist, although no explicit characterization can be given.

Proposition 2. *Let* $F \subseteq C \xrightarrow{\mathrm{m}} D$, $\rho \in uco(C)$ *and* $\eta \in uco(D)$. *There exists the complete kernel of* η *relative to* ρ.

Let us now consider the case where $C = D$ and $\rho = \eta$. It is important to remark that if $\langle \mathcal{E}^{\rho}_{F}(\rho), \rho \rangle \in \Gamma(C,C,F)$, then $\mathcal{E}^{\rho}_{F}(\rho)$ is the most abstract among the domains $\varphi \sqsubseteq \rho$ such that $\rho \circ f = \rho \circ f \circ \varphi$. Thus, we stress that ρ thought of as output abstraction is considered fixed. We will see in Section 5 how this concept can be usefully exploited. Moreover, let us recall that in Giacobazzi and Ranzato's approach [12], the least complete extension of ρ, when it exists, is instead defined as the most abstract among the domains $\varphi \sqsubseteq \rho$ such that $\varphi \circ f = \varphi \circ f \circ \varphi$. Hence, by defining $E_F(\rho) \overset{\text{def}}{=} \sqcup \{\varphi \in uco(C) \mid \varphi \sqsubseteq \rho,\ \langle \varphi, \varphi \rangle \in \Gamma(C, C, F)\}$, this latter least complete extension exists whenever $\langle E_F(\rho), E_F(\rho) \rangle \in \Gamma(C, C, F)$. Therefore, in this case, ρ considered as output abstraction is not fixed. Thus, we remark that this latter concept of least complete extension is different from that introduced in Definition 2. In the next section, we will study both these interesting notions. In order to distinguish them, when $\langle E_F(\rho), E_F(\rho) \rangle \in \Gamma(C, C, F)$, we will call $E_F(\rho)$ the *absolute* least complete extension of ρ. Moreover, analogous dual considerations hold for complete kernels: We will call them absolute complete kernels.

4 Constructive Characterization of Completeness

The following key result characterizes complete abstract interpretations in a "constructive" way: In fact, it shows that a completeness equation $\eta \circ f = \eta \circ f \circ \rho$ holds iff ρ contains a certain set of points depending on η, and, in a dual fashion, iff η is contained in a certain set of points which depends on ρ. The proof makes use of a variant of the axiom of choice, known as Hausdorff's Maximal Principle [2, pag. 192]. We will exploit largely the following compact notation: For any $f : C \to D$ and $y \in D$, $H^f_y \overset{\text{def}}{=} \{x \in C \mid f(x) \leq y\}$.

Theorem 1. *Let $F \subseteq C \xrightarrow{c} D$, $\rho \in uco(C)$ and $\eta \in uco(D)$. Then,*
$\langle \rho, \eta \rangle \in \Gamma(C, D, F) \Leftrightarrow \eta \subseteq \{y \in D \mid \cup_{f \in F} max(H^f_y) \subseteq \rho\} \Leftrightarrow \cup_{f \in F, y \in \eta} max(H^f_y) \subseteq \rho$.
Moreover, $\{y \in D \mid \cup_{f \in F} max(H^f_y) \subseteq \rho\} \in uco(D)$.

It is then useful to observe that, for any arbitrary set of points S and any uco ρ, the following equivalence holds: $S \subseteq \rho \Leftrightarrow \rho \sqsubseteq \mathcal{M}(S)$. Thus, as the above theorem suggests, given any set of continuous functions $F \subseteq C \xrightarrow{c} D$, we define two mappings $L_F : uco(C) \to uco(D)$ and $R_F : uco(D) \to uco(C)$ as follows:

$$L_F(\rho) \overset{\text{def}}{=} \{y \in D \mid \cup_{f \in F} max(H^f_y) \subseteq \rho\}; \qquad R_F(\eta) \overset{\text{def}}{=} \mathcal{M}(\cup_{f \in F, y \in \eta} max(H^f_y)).$$

In this way, Theorem 1 can be restated as follows:

$$\langle \rho, \eta \rangle \in \Gamma(C, D, F) \Leftrightarrow L_F(\rho) \sqsubseteq \eta \Leftrightarrow \rho \sqsubseteq R_F(\eta).$$

In particular, $(L_F, uco(C), uco(D), R_F)$ is an adjunction. Consequently, for any $\rho \in uco(C)$ and $\eta \in uco(D)$, one gets the following characterizations for the operators $\mathcal{K}^\rho_F$ and $\mathcal{E}^\eta_F$:

- $\mathcal{K}^\rho_F(\beta) = \sqcap\{\mu \in uco(D) \mid \beta, L_F(\rho) \sqsubseteq \mu\} = \beta \sqcup L_F(\rho)$;
- $\mathcal{E}^\eta_F(\delta) = \sqcup\{\varphi \in uco(C) \mid \varphi \sqsubseteq \delta, R_F(\eta)\} = \delta \sqcap R_F(\eta)$.

Hence, since $L_F(\rho) \sqsubseteq \beta \sqcup L_F(\rho)$ and $\delta \sqcap R_F(\eta) \sqsubseteq R_F(\eta)$, by Theorem 1, we obtain that $\langle \rho, \beta \sqcup L_F(\rho) \rangle, \langle \delta \sqcap R_F(\eta), \eta \rangle \in \Gamma(C, D, F)$, and therefore, according to Definition 2, we can draw the following consequences:

- The complete kernel of η rel. to ρ is the least common abstraction of η and $L_F(\rho)$;
- The least complete extension of ρ rel. to η is the reduced product of ρ and $R_F(\eta)$.

For any $\rho \in uco(C)$ and $\eta \in uco(D)$, it is helpful to define two dual mappings $\mathcal{F}_F^\rho : uco(D) \to uco(C)$ and $\mathcal{G}_F^\eta : uco(C) \to uco(D)$ as follows:

$$\mathcal{F}_F^\rho(\mu) \stackrel{\text{def}}{=} \rho \sqcap R_F(\mu); \qquad \mathcal{G}_F^\eta(\varphi) \stackrel{\text{def}}{=} \eta \sqcup L_F(\varphi).$$

Summing up, we have shown the following result, which explicitly states what one must add to ρ in order to get its least complete extension relative to η and, dually, what one must subtract from η in order to get its complete kernel relative to ρ.

Theorem 2. *Let $F \subseteq C \xrightarrow{c} D$, $\rho \in uco(C)$ and $\eta \in uco(D)$.*

- $\mathcal{F}_F^\rho(\eta) = \mathcal{M}(\rho \cup (\cup_{f \in F, y \in \eta} max(H_y^f)))$
is the least complete extension of ρ rel. to η;
- $\mathcal{G}_F^\eta(\rho) = \eta \cap \{y \in D \mid \cup_{f \in F} max(H_y^f) \subseteq \rho\}$
is the complete kernel of η rel. to ρ.

Example 1. Consider the example sketched in Section 1. Let $\rho \in uco(\mathbf{P}(\mathbb{N}_\perp \times \mathbb{N}_\perp))$ be the uco associated to the input abstract domain $S \times S$, and $\eta \in uco(\mathbf{P}(Bool_\perp))$ be the uco associated to the output abstract domain S. Therefore, $\rho = \{\{\langle \perp, \perp \rangle\}, \{\perp\} \times \mathbb{N}_\perp, \mathbb{N}_\perp \times \{\perp\}, \mathbb{N}_\perp \times \mathbb{N}_\perp\}$ and $\eta = \{\{\perp\}, Bool_\perp\}$. The semantic function f is obviously continuous and hence, by Theorem 2, the least complete extension of ρ for f relative to η does exist, and it is given by the reduced product $\mathcal{F}_f^\rho(\eta) = \rho \sqcap R_f(\eta)$. Thus, for $y \in \eta$, let us compute $max(H_y^f)$. We have that:

- $max(H_{Bool_\perp}^f) = \{\mathbb{N}_\perp \times \mathbb{N}_\perp\}$;
- $max(H_{\{\perp\}}^f) = max(\{Z \in \mathbf{P}(\mathbb{N}_\perp \times \mathbb{N}_\perp) \mid f(Z) \subseteq \{\perp\}\})$
$= max(\{Z \in \mathbf{P}(\mathbb{N}_\perp \times \mathbb{N}_\perp) \mid \langle 3,3 \rangle \notin Z\} = \{(\mathbb{N}_\perp \times \mathbb{N}_\perp) \setminus \{\langle 3,3 \rangle\}\}$.

Hence, $\mathcal{F}_F^\rho(\eta) = \mathcal{M}(\rho \cup \{(\mathbb{N}_\perp \times \mathbb{N}_\perp) \setminus \{\langle 3,3 \rangle\}\}) = \rho \cup \{(\mathbb{N}_\perp \times \mathbb{N}_\perp) \setminus \{\langle 3,3 \rangle\}\}$. Thus, as announced in Section 1, and as one naturally expects, this shows that the least complete extension of $S \times S$ can be obtained by adding a point $\langle \neq, \neq \rangle$ with concrete meaning $(\mathbb{N}_\perp \times \mathbb{N}_\perp) \setminus \{\langle 3,3 \rangle\}$, i.e. denoting that first and second components are surely not simultaneously equal to the value 3. It should be clear that this refined input abstract domain induces now a complete abstract interpretation. □

Let us now turn to absolute complete kernels and absolute least complete extensions, as formally introduced at the end of Section 3. What follows generalizes the results in [12, Section 6], where the hypothesis consisted of dealing with additive semantic functions. Assume that $C = D$, i.e. $F \subseteq C \xrightarrow{c} C$, and let $\rho \in uco(C)$. By Theorem 1, for any $\varphi \in uco(C)$, we have that:

- $\varphi \sqsubseteq \mathcal{F}_F^\rho(\varphi) \Leftrightarrow \varphi \sqsubseteq \rho$ and $\langle \varphi, \varphi \rangle \in \Gamma(C, C, F)$;
- $\mathcal{G}_F^\rho(\varphi) \sqsubseteq \varphi \Leftrightarrow \rho \sqsubseteq \varphi$ and $\langle \varphi, \varphi \rangle \in \Gamma(C, C, F)$.

Therefore, for the operator E_F introduced at the end of Section 3, we obtain that $E_F(\rho) = \sqcup\{\varphi \in uco(C) \mid \varphi \sqsubseteq \mathcal{F}_F^\rho(\varphi)\}$. Then, since $\mathcal{F}_F^\rho : uco(C) \to uco(C)$ is clearly monotone for any ρ, and hence admits the greatest fixpoint, we get $E_F(\rho) = gfp(\mathcal{F}_F^\rho)$. Moreover, by Theorem 2, $\langle gfp(\mathcal{F}_F^\rho), gfp(\mathcal{F}_F^\rho)\rangle \in \Gamma(C, C, F)$. This means that the absolute least complete extension of ρ exists, and it is $gfp(\mathcal{F}_F^\rho)$. Dual considerations hold for complete kernels. Thus, we get the following constructive characterization for absolute completeness.

Theorem 3. *Let $F \subseteq C \xrightarrow{c} C$ and $\rho \in uco(C)$. Then, $gfp(\mathcal{F}_F^\rho)$ and $lfp(\mathcal{G}_F^\rho)$ are, resp., the absolute least complete extension and absolute complete kernel of ρ.*

5 Generalized Quotients of Abstract Interpretations

The concept of *quotient* of an abstract interpretation has been recently introduced by Cortesi et al. [4] in order to formalize the least amount of information of a complex abstract domain A that is useful for computing some property that A is able to represent. Cortesi et al. [4] show how to exploit this notion for comparing the precision of two abstract interpretations in computing a given common property. Notably, they compare the well-known Jacobs and Langen *Sharing* [13] and Marriott and Søndergaard *Pos* [14] Prolog abstract interpretations, by demonstrating that *Pos* is strictly more precise than *Sharing* for computing variable groundness information. Further, Bagnara et al. [1] show, also experimentally, that in order to compute pair-sharing information, the use of the quotient of *Sharing* w.r.t. the pair-sharing Søndergaard domain [19] leads to remarkable gains of efficiency, when compared with the full domain *Sharing*.

Let us recall from [4] the definition of quotient. Let A be any complete lattice, $f : A \xrightarrow{m} A$ be a monotone semantic function on A, and $\rho \in uco(A)$ be an abstraction of A. Here, $\langle A, f\rangle$ models any abstract interpretation of some reference semantic definition, while ρ plays the rôle of the property (i.e. the abstraction of A) one is interested in. The equivalence relation $r_\rho \subseteq A \times A$ is defined as follows:[1]

$$\langle a_1, a_2\rangle \in r_\rho \text{ iff } \forall i \in \omega.\ \rho(f^i(a_1)) = \rho(f^i(a_2)).$$

Roughly speaking, $\langle a_1, a_2\rangle \in r_\rho$ when A views a_1 and a_2 as "equivalent" w.r.t. the computation of the property ρ. Thus, according to this intuition, the quotient $\mathcal{Q}_\rho(A)$ of A w.r.t. ρ is defined (cf. [4, Definition 3.5]) as the subset of A of the lub's of all equivalence classes of r_ρ: That is, if $[a]$ denotes a generic equivalence class for r_ρ, then $\mathcal{Q}_\rho(A) \stackrel{\text{def}}{=} \{\vee[a] \mid a \in A\}$, and the ordering is that inherited from A. Cortesi et al. [4, Theorem 3.6] show that if the equivalence r_ρ is additive, i.e. $\forall i \in I.\ \langle a_i, b_i\rangle \in r_\rho \Rightarrow \langle \vee_{i \in I} a_i, \vee_{i \in I} b_i\rangle \in r_\rho$, then $\mathcal{Q}_\rho(A)$ is well-defined, namely

[1] This definition considers the case of the first limit ordinal ω for practical purposes – a generalization to any (possibly transfinite) ordinal would be straightforward.

it is in turn an abstraction of A, i.e. the set of fixpoints of a uco on A, and ρ is an abstraction of $\mathcal{Q}_\rho(A)$.

Cortesi et al.'s results can be sharpened as follows. Firstly, it is useful to recall (see [8, Section 6.3]) that, in general, given an equivalence relation R on a complete lattice L, R is additive iff $\lambda x. \vee_L [x]_R \in uco(L)$. Thus, the hypothesis that the equivalence relation r_ρ is additive is indeed equivalent to the fact that $\lambda a. \vee [a] \in uco(A)$, i.e. that the quotient $\mathcal{Q}_\rho(A)$ is well-defined. In this case, that $\mathcal{Q}_\rho(A) \sqsubseteq \rho$ (i.e. [4, Theorem 3.6 (ii)]) is an immediate consequence: In fact, if $a = \rho(a)$ and $b \in [a]$ then $a = \rho(a) = \rho(f^0(a)) = \rho(f^0(b)) = \rho(b)$, and hence $b \leq a$, i.e. $\vee[a] = a$.

En passant, we observe that the additivity of f is an obvious sufficient condition guaranteeing that the quotient exists. Actually, the quotients presented in [1,4] exist just because the involved semantic functions are additive.

Lemma 1. *If $f : A \xrightarrow{a} A$ and $\rho \in uco(A)$ then $\mathcal{Q}_\rho(A) \in uco(A)$.*

It turns out that the quotient abstract domain satisfies the following remarkable property of "minimality": When a quotient $\mathcal{Q}_\rho(A)$ exists, if $\phi_\rho = \lambda a. \vee [a] \in uco(A)$ is the uco associated to $\mathcal{Q}_\rho(A)$, then ϕ_ρ is the most abstract solution in $uco(A)$ of the system of equations $\{\rho \circ f^i = \rho \circ f^i \circ \psi \mid i \in \omega\}$.

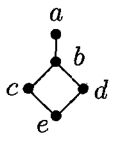
Lemma 2. *Let $\rho \in uco(A)$ such that $\mathcal{Q}_\rho(A) \in uco(A)$. Then, $\forall i \in \omega. \, \rho \circ f^i = \rho \circ f^i \circ \mathcal{Q}_\rho(A)$, and for any $\psi \in uco(A)$, $\forall i \in \omega. \, \rho \circ f^i = \rho \circ f^i \circ \psi$ implies $\psi \sqsubseteq \mathcal{Q}_\rho(A)$.*

Since the system of equations $\{\rho \circ f^i = \rho \circ f^i \circ \psi \mid i \in \omega\}$ is clearly equivalent to the system $\{\psi \sqsubseteq \rho\} \cup \{\rho \circ f^i = \rho \circ f^i \circ \psi \mid i > 0\}$, the above lemma says that, when a quotient $\mathcal{Q}_\rho(A)$ exists (i.e. $\mathcal{Q}_\rho(A) \in uco(A)$), it is characterized as follows:

$$\mathcal{Q}_\rho(A) = \sqcup\{\psi \in uco(A) \mid \psi \sqsubseteq \rho, \; \forall i > 0. \, \rho \circ f^i = \rho \circ f^i \circ \psi\}.$$

Of course, in the terminology of this paper, this means that when it exists, $\mathcal{Q}_\rho(A)$ is the least complete extension of ρ for the set of functions $\{f^i\}_{i>0}$ relative to ρ itself. However, it may well happen that, for some A, f and ρ, such least complete extension exists, whilst the quotient $\mathcal{Q}_\rho(A)$ does not exist, as the following simple example shows.

Example 2. Let A be the lattice depicted in the figure. Also, let $f : A \to A$ be defined as $f = \{a \mapsto a, \; b \mapsto a, \; c \mapsto e, \; d \mapsto e, \; e \mapsto e\}$, and let $\rho \in uco(A)$ such that $\rho(A) = \{a, b, e\}$. Trivially, f is monotone (and therefore continuous) but not additive. Moreover, f is idempotent, and therefore, for any $i \geq 1$, $f^i = f$. It turns out that, for any $i \geq 1$, $\rho \circ f^i = f$. As a consequence, r_ρ is not an additive equivalence relation. In fact, for any $i \geq 0$, $\rho(f^i(c)) = \rho(f^i(d))$: If $i = 0$ then, $\rho(c) = \rho(d) = b$; if $i \geq 1$ then, $\rho(f^i(c)) = f(c) = e = f(d) = \rho(f^i(d))$. But, $\rho(f(c \vee d)) = \rho(f(b)) = f(b) = a$. Hence, this means



that the quotient $Q_\rho(A)$ does not exist. Instead, as each f^i is monotone, by Theorem 2, the least complete extension of ρ for $\{f^i\}_{i>0}$ relative to ρ does exist. Moreover, this is given by the following reduced product: $\rho \sqcap (\cup_{i>0,\, y\in\rho} max(H_y^{f^i}))$. It is then a routine task to check that this is the domain A itself, i.e. the identity uco $\lambda x.x$. □

Then, Lemma 2 and Example 2 hint to generalize the notion of quotient as the least complete extension of ρ for $\{f^i\}_{i>0}$ relative to ρ, whenever this exists.

Definition 3. Given a complete lattice A, $f : A \xrightarrow{m} A$, and $\rho \in uco(A)$, the *generalized quotient* of A w.r.t. ρ is well-defined when there exists the least complete extension $\Phi_\rho(A)$ of ρ for $\{f^i\}_{i>0}$ relative to ρ; in such a case, the generalized quotient is defined to be $\Phi_\rho(A)$. □

It is here worth noting that the above definition naturally extends the intuitive meaning of the concept of quotient: In fact, the abstract domain $\Phi_\rho(A)$ is the most abstract domain which is more concrete than the property ρ and which is as good as A for propagating the information through the semantic function f. In other words, $\Phi_\rho(A)$ encodes exactly the least amount of information of A that is useful for computing the property ρ. Thus, this exactly formalizes the clear intuition behind the concept of quotient. As an immediate consequence of Theorem 2, we are then able to give the following theorem ensuring that, when the semantic function f is continuous, generalized quotients always exist.

Theorem 4. *If $f : A \xrightarrow{c} A$ then, for any $\rho \in uco(A)$, the generalized quotient $\Phi_\rho(A)$ exists.*

Acknowledgments. We wish to thank Enea Zaffanella for his helpful remarks on quotients of abstract interpretations and an anonymous referee for his/her useful comments. The work of Francesco Ranzato has been supported by an individual grant no. 202.12199 from Comitato 12 "Scienza e Tecnologie dell'Informazione" of Italian CNR.

References

1. R. Bagnara, P.M. Hill, and E. Zaffanella. Set-sharing is redundant for pair-sharing. In *Proc. 4th Int. Static Analysis Symp.*, LNCS 1302:53–67, 1997.
2. G. Birkhoff. *Lattice Theory.* AMS Colloq. Publications vol. XXV, 3rd ed., 1967.
3. G.L. Burn, C. Hankin, and S. Abramsky. Strictness analysis for higher-order functions. *Sci. Comput. Program.*, 7:249–278, 1986.
4. A. Cortesi, G. Filé, and W. Winsborough. The quotient of an abstract interpretation. *Theor. Comput. Sci.*, 202(1-2):163–192, 1998.
5. P. Cousot. *Méthodes itératives de construction et d'approximation de points fixes d'opérateurs monotones sur un treillis, analyse sémantique des programmes.* PhD thesis, Université Scientifique et Médicale de Grenoble, 1978.
6. P. Cousot. Completeness in abstract interpretation (Invited Lecture). In *Proc. 1995 Joint Italian-Spanish Conference on Declarative Programming*, pp. 37–38, 1995.

7. P. Cousot and R. Cousot. Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints. In *Proc. 4th ACM POPL*, pp. 238–252, 1977.
8. P. Cousot and R. Cousot. Systematic design of program analysis frameworks. In *Proc. 6th ACM POPL*, pp. 269–282, 1979.
9. P. Cousot and R. Cousot. Inductive definitions, semantics and abstract interpretation. In *Proc. 19th ACM POPL*, pp. 83–94, 1992.
10. P. Cousot and R. Cousot. Abstract interpretation of algebraic polynomial systems. In *Proc. 6th AMAST Conf.*, LNCS 1349:138–154, 1997.
11. R. Giacobazzi and F. Ranzato. Refining and compressing abstract domains. In *Proc. 24th ICALP*, LNCS 1256:771–781, 1997.
12. R. Giacobazzi and F. Ranzato. Completeness in abstract interpretation: a domain perspective. In *Proc. 6th AMAST Conf.*, LNCS 1349:231–245, 1997.
13. D. Jacobs and A. Langen. Static analysis of logic programs for independent AND-parallelism. *J. Logic Program.*, 13(2-3):154–165, 1992.
14. K. Marriott and H. Søndergaard. Precise and efficient groundness analysis for logic programs. *ACM Lett. Program. Lang. Syst.*, 2(1-4):181–196, 1993.
15. A. Mycroft. *Abstract interpretation and optimising transformations for applicative programs.* PhD thesis, CST-15-81, Univ. of Edinburgh, 1981.
16. A. Mycroft. Completeness and predicate-based abstract interpretation. In *Proc. ACM PEPM Conf.*, pp. 179–185, 1993.
17. U.S. Reddy and S.N. Kamin. On the power of abstract interpretation. *Computer Languages*, 19(2):79–89, 1993.
18. R.C. Sekar, P. Mishra, and I.V. Ramakrishnan. On the power and limitation of strictness analysis. *J. ACM*, 44(3):505–525, 1997.
19. H. Søndergaard. An application of abstract interpretation of logic programs: occur check reduction. In *Proc. ESOP '86*, LNCS 213:327–338, 1986.
20. B. Steffen. Optimal data flow analysis via observational equivalence. In *Proc. 14th MFCS Symp.*, LNCS 379:492–502, 1989.