



**HAL**  
open science

## Semantic foundations and inference of non-null annotations

Laurent Hubert, Thomas Jensen, David Pichardie

► **To cite this version:**

Laurent Hubert, Thomas Jensen, David Pichardie. Semantic foundations and inference of non-null annotations. [Research Report] 2008. inria-00266171v1

**HAL Id: inria-00266171**

**<https://inria.hal.science/inria-00266171v1>**

Submitted on 21 Mar 2008 (v1), last revised 25 Mar 2008 (v2)

**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.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

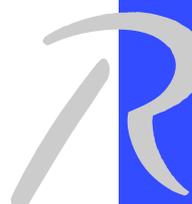
*Semantic foundations and inference of non-null annotations*

Laurent Hubert — Thomas Jensen — David Pichardie

N° ????

March 2008

Thème SYM

 *Rapport  
de recherche*



## Semantic foundations and inference of non-null annotations

Laurent Hubert<sup>\*</sup>, Thomas Jensen<sup>\*</sup>, David Pichardie<sup>\*\*</sup>

Thème SYM — Systèmes symboliques  
Équipe-Projet Lande

Rapport de recherche n° ???? — March 2008 — 37 pages

**Abstract:** This paper proposes a semantics-based automatic null pointer analysis for inferring non-null annotations of fields in object-oriented programs. The analysis is formulated for a minimalistic OO language and is expressed as a constraint-based abstract interpretation of the program which for each field of a class infers whether the field is definitely non-null or possibly null after object initialization. The analysis is proved correct with respect to an operational semantics of the minimalistic OO language. This correctness proof has been machine checked using the Coq proof assistant. We also prove the analysis complete with respect to the non-null type system proposed by Fähndrich and Leino, in the sense that for every typable program the analysis is able to prove the absence of null dereferences without any hand-written annotations. Experiments with a prototype implementation of the analysis show that the inference is feasible for large programs.

**Key-words:** program analysis, automated program proving, machine checked proofs, operational semantics

<sup>\*</sup> CNRS/IRISA, France

<sup>\*\*</sup> INRIA Rennes - Bretagne Atlantique/IRISA, France

# Foundations sémantiques et inférence d'annotations de type @NotNull

**Résumé :** Ce rapport propose une analyse automatique pour inférer des annotations @NotNull pour les champs des programmes orienté objet. L'analyse est proposée pour un langage orienté objet minimaliste et est présentée comme de l'interprétation abstraite. Pour chaque champs d'une classe, elle infère si le champs est toujours différent de nul ou bien si il peut être nul après l'initialisation de l'objet. Cette analyse est prouvée correcte vis-à-vis d'une sémantique opérationnelle du langage proposé et cette preuve a été vérifiée par l'assistant de preuve Coq. Nous prouvons aussi que l'analyse est complète vis-à-vis du système de type proposé par Fähndrich et Leino, c'est à dire que pour tout programme typable, l'analyse est capable de prouver l'absence de déréférencement de pointeur nul sans aucune annotation manuelle. Les expériences faites avec notre prototype montrent que l'inférence est possible pour de gros programmes.

**Mots-clés :** analyse de programmes, preuve automatique de programmes, preuve vérifiée par ordinateur, sémantique opérationnelle

## 1 Introduction

A common source of exceptional program behaviour is the dereferencing of null references (also called null pointers), resulting in segmentation faults in C or null pointer exceptions in Java. Even if such exceptions are caught, the presence of exception handlers creates an additional amount of potential branching which in turn implies that: 1) fewer optimizations are possible and 2) verification is more difficult (bigger certification conditions, implicit flow in information flow verification, etc.). Furthermore, the Java virtual machine is obliged to perform run-time checks for non-nullness of references when executing a number of its bytecode instructions, thereby incurring a performance penalty.<sup>1</sup> For all these reasons, a static program analysis which can guarantee before execution of the program that certain references will definitely be non-null is useful.

Some non-null type systems for the Java language have been proposed [6, 14]. Although non-null annotations are not used by the compiler yet, this could likely change in the future, in which case it becomes crucial to prove the soundness of the logic underlying the type checker. Furthermore, although non-null annotations are not yet mandatory, automatic type inference would help to retrofit legacy code, to lower the annotation burden on the programmer, to document the code and to verify existing code.

While some object oriented languages ensure all fields are initialized and objects are not read before being fully initialized, this is not the case in Java. More precisely, there are three aspects that complicate non-nullness analysis: 1) fields can be accessed during object construction (before or after their initialization) which means all fields contain null value before their first assignment; 2) no guarantee is given on the initialization of fields at the end of the constructor so if a field has not been initialized it contains a null value; and 3) an object being initialized can be stored in fields or passed as an argument of methods.

The first aspect means a naive flow-insensitive heap abstraction is not sufficient as all fields are null at some stage and hence would be annotated as possibly null. The second aspect makes a special initialization analysis necessary in order to track fields that are certain to be initialized during the construction. Those fields can safely be considered as non-null unless they are explicitly assigned a null value. All other fields might not have been initialized and must be considered as possibly null.

The third aspect was observed in [6] and concerns the problem where a virtual method `A.m()` is being redefined in a sub-class `B` by a method referencing a field `f` that is local to the class `B`. If the constructor of a super-class `A` executes `this.m()` on an object of class `B`, it calls a method that dereferences field `f` before the constructor of `B` has had the possibility of initializing this field. To solve this problem, references to objects under construction need to be tracked, e.g. by a tag indicating their state.

---

<sup>1</sup> Although hardware traps are used for free whenever possible, explicit non-nullness tests are still required as explained in [12]

*Related Work.* Freund and Mitchell have proposed in [9] a *type system* combined with a data-flow analysis to ensure the correct initialization of objects. The goal is to formalize the initialization part of Java bytecode verification. In this respect it is different from our analysis, which is focused on field initialization and on their nullness property. Fähndrich and Leino in [6] have proposed another *type system* also combined with a data-flow analysis to ensure a correct manipulation of references with respect to a nullness property. This system is presented in Sect. 6 where we compare our inference analysis to their type system and give examples of how our analysis infers more precise types than what their system is able to check. More recently, Fähndrich and Xia have proposed another *type system* introducing *delayed* initialization [7]. It generalizes the previous one and allows to prove some properties that our analysis cannot, like initialization of circular data structures, but it has also the same loss of precision discussed in Sect. 6 .

Some works are focused on *local type inference*, *i.e.* inferring nullness property for blocks of code from the guards. This is notably the case of FindBugs [11, 10] and of the work by Male *et al.* [14]. They rely on *path-sensitive* analysis and the treatment of field initialization is very weak.

To *infer type annotations*, Houdini [8] generates a set of possible annotations (non-null annotations among others) for a given program and uses ESC/Java [13] to refute false assumptions. CANAPA [3] lowers the burden of annotating a program by propagating some non-null annotations. It also relies on ESC/Java to infer where annotations are needed. Those two annotation assistants have a simplified handling of objects under construction and are intended to be used by the developer to debug and not to certify programs. Indeed, they rely on ESC/Java [13], which is not sound (nor complete). JastAdd [5] is a tool to infer annotations for a simplified version of Fähndrich and Leino’s type system.<sup>2</sup>

Finally, another approach to lower the amount of annotations is proposed by Chalin and James [2]. They suggest to consider the references as non-null by default, so the developer has only to explicitly annotate nullable references. They claim that, *in Java programs, at least 2/3 of declarations of reference types are meant to be non-null, by design*. Despite the lower amount of annotation needed, about 1/3 of declarations still need annotations which can represent a substantial amount of annotations in legacy code.

*Contributions.* The non-null reference analysis presented here makes the following contributions.

- A fully automatic analysis.
- A proof of soundness with respect to an operational semantics. This is the first formal correctness proof for this kind of analyses and it has been machine checked using the Coq proof assistant.
- A detailed comparison with Fähndrich and Leino’s type system, which is a reference among the nullness program analyses. We prove the completeness

<sup>2</sup> The treatment of objects under initialization is simplified and initializations done by methods (called from the constructor) are not taken in account.

with respect to their type system. In this way, the correctness proof of our analysis also provides a formal proof of correctness of their type system. We also show that our analysis can be more precise than their type system.

- The analysis is modular: a program can be analysed without analyzing the libraries if they have already been annotated.

*Outline.* Section 2 presents the syntax and semantics of the simple OO language we use to formalize our analysis. Section 3 presents the system of constraints of the analysis and Sect. 4 gives the proof of soundness. Section 5 proves the constraint system has a least fixpoint and discusses the modularity. Section 6 then presents Fähndrich and Leino’s type system and proves the completeness of our analysis with respect to their type system. Section 7 gives some details on the adaptation of the analysis to the Java bytecode level and Sect. 8 concludes.

## 2 Syntax and Semantics

We define a minimalistic language<sup>3</sup> to analyze the flow of references in object-based languages. A program is a collection of classes, arranged in a class hierarchy via a transitive subclass relation  $\prec$  (we write  $\preceq$  for its reflexive closure). We only consider single inheritance (where  $\preceq$  can be embedded in a sup-semilattice structure). The language, as described in Fig. 1 has two kinds of expressions  $E$ : variables and references to fields. Assignments are either to variables or to fields. New objects are created using the **new** instruction which takes as arguments the name  $C$  of a class, the vector  $\alpha$  of the types of the parameters (used to deal with overloading) and a list of expressions. There is a conditional instruction which may non-deterministically branches to a given program label. Finally, the instructions  $x.ms(E, \dots, E)$  and **return** represent method invocation and return, respectively. Methods invoked are found with a lookup procedure that looks for a method depending on its signature and the class of the current object. A method descriptor  $ms$  and a method identifier  $m$  are both of the form  $\{C, mn, \alpha \rightarrow \beta\}$ , where  $mn$  is a method name,  $\alpha \rightarrow \beta$  is a type signature (to simplify the presentation we here restrict  $\beta$  to **void**) and  $C$  is, for the method descriptor, the type of the reference on which the method is called and, for the method identifier, the class where the method has been defined. We use  $name(ms)$  and  $name(m)$  to retrieve  $mn$  and  $class(ms)$  and  $class(m)$  to retrieve  $C$ . A method signature is the couple  $(mn, \alpha \rightarrow \beta)$ . A method signature can correspond to many methods, a method descriptor corresponds to at most one method and a method identifier corresponds to only one method. As a notation abuse,  $m$  will be used as the method identified instead of the identifier itself. A method is composed of a method signature, a list of parameters and a method body consisting of a labelled list of instructions. A class is composed of fields and method definitions. For a given field  $f$   $class(f)$  gives the class in which  $f$  is defined. The set of

<sup>3</sup> The language is closed to the Java bytecode language but without any operand stack. Removing the operand stack avoid to introduce an alias analysis which is needed at the bytecode level, for instance, to know if a stack variable is an alias of **this**.

fields declared in a class  $C$  is written  $\text{fields}(C)$ . Each program contains a special method **main** used to start the execution.

The operational semantics of our minimalistic language is defined by the inference rules in Fig. 2 which specify a (small-step) transition relation  $\rightarrow_m$  for each method  $m$  and a (big-step) evaluation relation  $\Downarrow_{h,l}$  for expressions relative to a heap  $h$  and local variables  $l$ . The semantics uses an explicit error element  $\Omega$  to signal the dereferencing of a null reference. Alternatively, we could have let the semantics “get stuck” when it encounters an error but our choice of propagating an error element to method boundaries facilitates the correctness proof of the null pointer analysis to follow. For space reasons, we only detail the error case for method invocation; all other instructions may lead to similar error steps. The analysis keeps track of how objects are initialized and, in order to prove its correctness, the semantics instruments the fields of objects with flags **def** or **undef** to track the history of field initializations. A field flagged as **undef** contains necessarily `MaybeNull`. This instrumentation is transparent in the sense that it does not affect the behavior of a program and will be used later to prove the correctness of the analysis. In Java bytecode, unlike fields, local variables do not have a default value and the bytecode verifier ensures uninitialized local variables are never read. We formalized this by using  $\perp$  as the default value for local variables and by ensuring with the semantics  $\perp$  is never read. We also model object initialization to ensure a constructor of class  $C$  terminates only when at least one constructor of each ancestor has been called. To do so, we keep for each object the set of classes whose constructor has been called. The semantics is stuck if a constructor violates that policy. This means we only consider programs where objects are correctly initialized.

### Syntax

$$\begin{aligned}
 E &::= x \mid E.F \\
 I &::= x \leftarrow E \mid x.f \leftarrow E \mid x \leftarrow \mathbf{new} (C, \alpha)(E, \dots, E) \mid \mathbf{if} (\star) \mathbf{jmp} \mid x.\mathbf{ms}(E, \dots, E) \mid \mathbf{return} \\
 M &::= \text{mn}(x_1, \dots, x_n) \{ I; \dots; I \} \\
 C &::= \{\mathbf{fields} : \{f; \dots; f\}; \mathbf{methods} : \{M; \dots; M\}\} \\
 P &::= (C \cdots C, \prec)
 \end{aligned}$$

### Domains

$$\begin{aligned}
 \text{Val} &= \text{Loc} + \{\mathbf{null}\} & \text{Object} &= \mathbb{F} \rightarrow \text{Val} \times \{\mathbf{def}, \mathbf{undef}\} \\
 \text{LocalVar} &= \mathbb{V} \rightarrow \text{Val} + \{\perp\} & \text{Heap} &= \text{Loc} \rightarrow \text{Object} \times \mathbb{C} \times \wp(\mathbb{C}) \\
 \text{State} &= (\mathbb{N} \times \text{LocalVar} \times \text{Heap}) + \Omega
 \end{aligned}$$

### Expression Evaluation

$$\frac{l(x) \neq \perp}{x \Downarrow_{h,l} l(x)} \quad \frac{e \Downarrow_{h,l} r \quad r \in \text{dom}(h) \quad f \in \text{dom}(h(r))}{e.f \Downarrow_{h,l} h(r)(f)} \quad \frac{e \Downarrow_{h,l} v \quad v \in \{\mathbf{null}, \Omega\}}{e.f \Downarrow_{h,l} \Omega}$$

**Fig. 1.** Syntax of the language

*Notation:* We write  $\mathbb{F}$ ,  $\mathbb{V}$ ,  $\text{Loc}$  and  $\mathbb{N}$  for fields, variables, memory addresses and program points, respectively. We consider that a field includes the class

## Operational Semantics: Normal Cases

$$\begin{array}{c}
 \frac{\boxed{P_m[i] = x \leftarrow e} \quad e \Downarrow_{h,l} v \quad x \neq \mathbf{this}}{\langle i, l, h \rangle \rightarrow_m \langle i + 1, l[x \mapsto v], h \rangle} \quad \frac{\boxed{P_m[i] = x.f \leftarrow e} \quad l(x) \in \text{dom}(h) \quad f \in \text{dom}(h(l(x))) \quad e \Downarrow_{h,l} v}{\langle i, l, h \rangle \rightarrow_m \langle i + 1, l, \text{upd}(h, l(x), f, (v, \mathbf{def})) \rangle} \\
 \\
 \frac{\boxed{P_m[i] = x \leftarrow \mathbf{new} (C, \alpha)(e_1, \dots, e_n)} \quad \forall i, e_i \Downarrow_{h,l} v_i \quad r \notin \text{dom}(h) \quad x \neq \mathbf{this} \quad \langle 0, \perp[\mathbf{this} \mapsto r, \{i \mapsto v_i\}_{i=1..n}], h[r \mapsto \text{default}(C)] \rangle \rightarrow_{\llbracket C, \text{init}, \alpha \rightarrow \text{void} \rrbracket}^* h'}{\langle i, l, h \rangle \rightarrow_m \langle i + 1, l[x \mapsto r], h' \rangle} \\
 \\
 \frac{\boxed{P_m[i] = x.\text{ms}(e_1, \dots, e_n)} \quad \forall i, e_i \Downarrow_{h,l} v_i \quad l(x) \in \text{dom}(h) \quad m' = \text{lookup}(\text{class}(h(l(x))), \text{ms}) \quad \text{class}(h(l(x))) \preceq \text{class}(\text{ms}) \quad \langle 0, \perp[\mathbf{this} \mapsto l(x), \{i \mapsto v_i\}_{i=1..n}], h \rangle \rightarrow_{m'}^* h' \quad \text{name}(m) = \mathbf{init} \Rightarrow x = \mathbf{this} \quad \text{name}(m) \neq \mathbf{init} \Rightarrow h' = h}{\langle i, l, h \rangle \rightarrow_m \langle i + 1, l, h' \rangle} \\
 \\
 \frac{\boxed{P_m[i] = \mathbf{if jmp}}}{\langle i, l, h \rangle \rightarrow_m \langle \text{jmp}, l, h \rangle} \quad \frac{\boxed{P_m[i] = \mathbf{if jmp}}}{\langle i, l, h \rangle \rightarrow_m \langle i + 1, l, h \rangle} \\
 \\
 \frac{\boxed{P_m[i] = \mathbf{return}} \quad C = \text{class}(m) \quad \text{name}(m) = \mathbf{init} \Rightarrow \forall A, C \prec A \Rightarrow A \in \text{history}(l(\mathbf{this})) \quad \text{name}(m) = \mathbf{init} \Rightarrow h' = h[l(\mathbf{this}) \mapsto \text{addHistory}_C(\text{set2Def}_C(h(l(\mathbf{this}))))] \quad \text{name}(m) \neq \mathbf{init} \Rightarrow h' = h}{\langle i, l, h \rangle \rightarrow_m h'}
 \end{array}$$

## Operational Semantics: Error Cases

$$\begin{array}{c}
 \frac{\boxed{P_m[i] = x \leftarrow e} \quad e \Downarrow \Omega}{\langle i, l, h \rangle \rightarrow_m \Omega} \quad \frac{\boxed{P_m[i] = x.f \leftarrow e} \quad l(x) = \mathbf{null} \vee e \Downarrow \Omega}{\langle i, l, h \rangle \rightarrow_m \Omega} \\
 \\
 \frac{\boxed{P_m[i] = x \leftarrow \mathbf{new} C(e_1, \dots, e_n)} \quad \exists i, e_i \Downarrow_{h,l} \Omega}{\langle i, l, h \rangle \rightarrow_m \Omega} \\
 \\
 \frac{\boxed{P_m[i] = x \leftarrow \mathbf{new} C(e_1, \dots, e_n)} \quad \forall i, e_i \Downarrow_{h,l} v_i \quad r \notin \text{dom}(h) \quad \langle 0, \perp[\mathbf{this} \mapsto r, \{i \mapsto v_i\}_{i=1..n}], h[r \mapsto \text{default}(C)] \rangle \rightarrow_{\llbracket C, \text{init}, \alpha \rightarrow \text{void} \rrbracket} \Omega}{\langle i, l, h \rangle \rightarrow_m \Omega} \\
 \\
 \frac{\boxed{P_m[i] = x.\text{ms}(e_1, \dots, e_n)} \quad \forall i, e_i \Downarrow_{h,l} v_i \quad l(x) \in \text{dom}(h) \quad m' = \text{lookup}(\text{class}(h(l(x))), \text{ms}) \quad \text{class}(h(l(x))) \preceq \text{class}(\text{ms}) \quad \langle 0, \perp[\mathbf{this} \mapsto l(x), \{i \mapsto v_i\}_{i=1..n}], h \rangle \rightarrow_{m'}^* \Omega \quad \text{name}(m) = \mathbf{init} \Rightarrow x = \mathbf{this}}{\langle i, l, h \rangle \rightarrow_m \Omega} \\
 \\
 \frac{\boxed{P_m[i] = x.\text{ms}(e_1, \dots, e_n)} \quad l(x) = \mathbf{null} \vee \exists i, e_i \Downarrow_{h,l} \Omega}{\langle i, l, h \rangle \rightarrow_m \Omega}
 \end{array}$$

Fig. 2. Semantics of the language

name where it has been defined. We note  $h(r)(f)$ ,  $\text{class}(h(r))$  and  $\text{history}(h(r))$  the accesses to the first, second and third components of the heap cell  $h(r)$ , respectively. Function  $\text{upd}(h, r, f, (v, \mathbf{def}))$  sets the field  $f$  of the object at location  $r$  to the value  $v$  and marks it as defined. The expression  $\text{default}(C)$  denotes a new object of class  $C$  where all fields contain null values and are **undef** (the history of such object is empty). The function  $\text{set2Def}_C$  sets all fields (in  $\text{fields}(C)$ ) of an object to **def**. The function  $\text{addHistory}_C$  adds a class name  $C$  to the initialization history of an object. The star transition  $\rightarrow_m^*$  corresponds to the transitive closure of  $\rightarrow_m$ .

A program is considered *null-pointer error safe* if the execution of the method **main** never reaches an error state.

**Definition 1.** *A program  $P$  is said to be null-pointer error safe if for all local variables  $l$  and heap  $h$ ,*

$$\langle 0, l, h \rangle \rightarrow_{\text{main}}^* s \text{ implies } s \neq \Omega$$

Note that since error states are propagated after method calls, this definition of safe program implies that no dereferencing of a null reference will occur during the execution of the program.

### 3 Null-Pointer Analysis

We will now present the analysis that is able to prove a program is null-pointer error safe. Without field annotations, this analysis would be the naive intra-procedural flow-sensitive analysis inferring pre- and post-conditions for methods.

#### 3.1 Abstract Domains

In this section we define an analysis that for each class of a well-typed program infers annotations about nullity of its fields together with pre- and post-conditions for its methods. The basic properties of interest are **NotNull** — meaning “definitely not-null” — and **MaybeNull** — meaning possibly a null reference. The field annotations computed by the analysis are represented as a heap abstraction  $H^\# \in \text{Heap}^\#$  which provides an abstraction for all fields of all initialized objects and all initialized fields of objects being initialized.<sup>4</sup>

As explained in the Introduction, object initialization requires a special treatment because all fields are null when the object is created so this would lead a simple-minded analysis to annotate all fields as **MaybeNull**. However, we want to infer non-null annotations for all fields which are initialized explicitly to be so in a constructor. In other words, fields that are not initialized in a constructor are annotated as **MaybeNull** and all other fields have a type compatible with the value they have been initialized with.

<sup>4</sup> We consider fields *initialized* when they have been assigned a value, whereas we consider an object *initialized* when it has returned from its constructor.

## Abstract Domains

$\text{Val}^\sharp = \{\text{Raw}(Y) \mid Y \in \text{Class}\} \cup \{\text{Raw}^-, \text{NotNull}, \text{MaybeNull}\}$			
$\text{Def}^\sharp = \{\text{Def}, \text{UnDef}\}$	$\text{TVal}^\sharp = \mathbb{F} \rightarrow \text{Def}^\sharp$	$\text{Heap}^\sharp = \mathbb{F} \rightarrow \text{Val}^\sharp$	$\text{LocalVar}^\sharp = \mathbb{V} \rightarrow \text{Val}^\sharp$
$\text{Method}^\sharp = \mathbb{M} \rightarrow \{\{\text{this} \in \text{TVal}^\sharp; \text{args} \in (\text{Val}^\sharp)^*; \text{post} \in \text{TVal}^\sharp\}\}$			
$\text{State}^\sharp = \text{Method}^\sharp \times \text{Heap}^\sharp \times (\mathbb{M} \times \mathbb{N} \rightarrow \text{TVal}^\sharp) \times (\mathbb{M} \times \mathbb{N} \rightarrow \text{LocalVar}^\sharp)$			

## Selected Partial Orders

$\text{Val}^\sharp$		$\text{Def}^\sharp$
$x \in \text{Val}^\sharp$	$\text{Raw}(X) \sqsubseteq \text{Raw}^-$	
$\text{NotNull} \sqsubseteq x$	$X \preceq Y$	$\text{Def} \sqsubseteq \text{UnDef}$
$x \in \text{Val}^\sharp$	$x \sqsubseteq \text{MaybeNull}$	
$\text{Raw}(X) \sqsubseteq \text{Raw}(Y)$		

Fig. 3. Abstract domains and selected partial orders

To this end, the analysis tracks field initializations of the current object in constructors (and methods called from constructors). This is done via an abstraction of the **this** reference by a domain  $\text{TVal}^\sharp$  which maps each of the fields declared in the current class to  $\text{Def}$  or  $\text{UnDef}$ . To allow strong updates, we need a flow-sensitive abstraction so to each program point we map a such abstraction ( $T^\sharp$ ).

References are then abstracted by a domain  $\text{Val}^\sharp$  which incorporate the “raw” references from [6]. A  $\text{Raw}^-$  value denotes a non-null reference of an object being initialized, which does not yet respect his invariant (e.g.  $\text{Raw}^-$  can be used as a property of **this** when it occurs in constructors). If a reference is known to have all its fields declared in class  $X$  and in the parents of  $X$  initialized, then the reference is  $\text{Raw}(X)$ . The inclusion of “raw” references allows the manipulation of objects during initialization because the analysis can use the fact that for an object of type  $\text{Raw}(X)$ , only fields declared in  $X$  and above have a valid annotation in the abstract heap  $H^\sharp$ . A  $\text{NotNull}$  value denotes a non-null reference that has finished its initialization.

Figure 3 defines formally each abstract domain. The flow of references through local variables is analysed with a flow-sensitive abstraction of each variable ( $L^\sharp \in \text{LocalVar}^\sharp$ ). The analysis also infers method annotations  $M^\sharp \in \text{Method}^\sharp$ . For any method  $m$ ,  $M^\sharp(m)[\text{this}]$  is an approximation of the initialization state of **this** before the execution of  $m$ , while  $M^\sharp(m)[\text{post}]$  gives the corresponding approximation at the end of the execution.  $M^\sharp(m)[\text{args}]$  approximates the parameters of the method, taking into account all the context in which  $m$  may actually be invoked. We only give the definition of the partial order for  $\text{Val}^\sharp$  and  $\text{Def}^\sharp$ . The other orders are defined in a standard way using the canonical orders on partial functions, products and lists. The final domain  $\text{State}^\sharp$  is hence equipped with a straightforward lattice structure.

### 3.2 Inference Rules

The analysis is specified via a set of inference rules, shown in Fig. 4, which define a judgment

$$M^\#, H^\#, T^\#, L^\# \models (m, i) : instr$$

for when the abstract state  $(M^\#, H^\#, T^\#, L^\#)$  is coherent with instruction  $instr$  at program point  $(m, i)$ . For each such program point, this produces a set of constraints over the abstract domain  $\text{State}^\#$ , whose solutions constitute the correct analysis results. The rules make use of an abstract evaluation function for expressions (explained below) that we write  $\llbracket e \rrbracket^\#$ . An example of a program with the corresponding constraints are given in Sect. 3.3.

Assignment to a local variable (rule (1)) simply assigns the abstract value associated with the expression to the local variable in the abstract environment  $L^\#$ . Assignment to a field (2) can either be to a field of the current object, in which case the field becomes “defined”, or to another object. In both cases, the abstract heap  $H^\#$  is augmented with the value of the expression as a possible value for the field. When a return value is encountered (3) in a constructor, all still-undefined fields are explicitly set to `MaybeNull`. For a `new` instruction (4), a new abstraction  $\top_C$  is built for the pre-condition on `this` of the constructor. The first argument (`this`) is set to  $\text{Raw}^-$ . The result of the constructor is known to be `NotNull` (the object is fully initialized). The method call (6) uses conditional constraints to distinguish a number of cases, depending on whether the call is to a method in the current class and the receiving object is the current object `this` or not. We use  $\text{Raw}(\text{super}(C))$  to denote the `Raw` type just above  $\text{Raw}(C)$ .<sup>5</sup>

The whole constraint system of a program  $P$  is then formally defined by the judgement  $M^\#, H^\#, T^\#, L^\# \models P$ . We write  $\text{overrides}(m', m)$  when the method  $m'$  overrides  $m$ . In such case we require a contravariant property between the parameters of  $m$  and  $m'$ . Any overriding  $\text{overrides}(m', m)$  need also to invalidate (in term of precision)  $M^\#(m)[\text{post}]$  because a virtual call to  $m$  could lead to the execution of  $m'$  which is not able (by definition of  $T^\#$ ) to track the initialization of fields declared in  $\text{class}(m)$ . A similar constraint is required for  $M^\#(m')[\text{this}]$  because for a virtual call to  $m$  we have only constrained  $M^\#(m)[\text{this}]$  and not  $M^\#(m')[\text{this}]$ .

Finally an abstract state is said safe (noted  $\text{safe}_P^\#(M^\#, H^\#, T^\#, L^\#)$ ) when for each program point  $(m, i)$ , if  $P_m[i]$  is of the form  $x.f \leftarrow e$  or  $x.\text{ms}(\dots)$  then  $L^\#(m, i)(x) \neq \text{MaybeNull}$ , and for all expressions  $e$  which appear in the instruction  $P_m[i]$ , any dereferenced sub-expression  $e'$  has an abstract evaluation strictly lower than `MaybeNull`.

The analysis uses a function  $\rho$  to transfer information from domain  $\text{Val}^\#$  to domain  $\text{TVal}^\#$ :  $\rho$  transforms an abstract reference  $\text{Val}^\#$  to a  $\text{TVal}^\#$  abstraction of a current object `this` in the class  $C$ . The notations  $\perp_C$ , respectively  $\top_C$ , correspond to the function that maps all fields defined in  $C$  to `Def`, respectively `UnDef`. The analysis also relies on an abstract evaluation  $\llbracket e \rrbracket_{C, H^\#, t^\#, l^\#}^\#$  of expressions parameterised by the current class  $C$ , an abstract heap  $H^\#$ , an abstraction

<sup>5</sup> If  $C$  is the root of the class hierarchy (Object in Java), then  $\text{Raw}(\text{super}(C)) = \text{Raw}^-$

## Inference Rules

$$\begin{array}{c}
 \frac{T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1) \quad L^\sharp(m, i)[x \mapsto \llbracket e \rrbracket^\sharp] \sqsubseteq L^\sharp(m, i+1)}{M^\sharp, H^\sharp, T^\sharp, L^\sharp \models (m, i) : x \leftarrow e} \quad (1) \\
 \dots \\
 \frac{\text{if } x = \mathbf{this} \wedge f \in \text{fields}(\text{class}(m)) \text{ then } T^\sharp(m, i)[f \mapsto \text{Def}] \text{ else } T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1)}{L^\sharp(m, i) \sqsubseteq L^\sharp(m, i+1)} \quad \frac{\llbracket e \rrbracket^\sharp \sqsubseteq H^\sharp(f)}{M^\sharp, H^\sharp, T^\sharp, L^\sharp \models (m, i) : x.f \leftarrow e} \quad (2) \\
 \dots \\
 \frac{\text{name}(m) = \mathbf{init} \Rightarrow \forall f \in \text{fields}(\text{class}(m)). (T^\sharp(m, i)(f) = \text{UnDef} \Rightarrow \text{MayBeNull} \sqsubseteq H^\sharp(f))}{M^\sharp, H^\sharp, T^\sharp, L^\sharp \models (m, i) : \mathbf{return}} \quad (3) \\
 \dots \\
 \frac{\begin{array}{c} \top_C \sqsubseteq M^\sharp(m')[\mathbf{this}] \quad \text{Raw}^- :: \llbracket e_1 \rrbracket^\sharp :: \dots :: \llbracket e_j \rrbracket^\sharp \sqsubseteq M^\sharp(m')[\text{args}] \\ L^\sharp(m, i)[x \mapsto \text{NotNull}] \sqsubseteq L^\sharp(m, i+1) \quad T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1) \end{array}}{M^\sharp, H^\sharp, T^\sharp, L^\sharp \models (m, i) : x \leftarrow \mathbf{new}(C, \alpha)(e_1, \dots, e_j)} \quad (4) \\
 \dots \\
 \frac{\begin{array}{c} L^\sharp(m, i) \sqsubseteq L^\sharp(m, i+1) \quad L^\sharp(m, i) \sqsubseteq L^\sharp(m, \text{jmp}) \\ T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1) \quad T^\sharp(m, i) \sqsubseteq T^\sharp(m, \text{jmp}) \end{array}}{M^\sharp, H^\sharp, T^\sharp, L^\sharp \models (m, i) : \mathbf{if jmp}} \quad (5) \\
 \dots \\
 \frac{\begin{array}{c} m' = \text{lookup}(\text{class}(\text{ms}), \text{ms}) \quad C' = \text{class}(m') \\ \left( \begin{array}{l} \text{if } x = \mathbf{this} \wedge \text{class}(m) = C' \wedge \llbracket x \rrbracket^\sharp \sqsubseteq \text{Raw}(\text{super}(C')) \wedge M^\sharp(m')[\mathbf{post}] \sqcap T^\sharp(m, i) \sqsubseteq \perp_C \\ \text{then } L^\sharp(m, i)[x \mapsto \text{Raw}(C') \sqcap \llbracket x \rrbracket^\sharp] \sqsubseteq L^\sharp(m, i+1) \\ \text{else if } \text{name}(m') = \mathbf{init} \text{ then } L^\sharp(m, i)[x \mapsto \text{Raw}(C')] \sqsubseteq L^\sharp(m, i+1) \\ \text{else } L^\sharp(m, i) \sqsubseteq L^\sharp(m, i+1) \end{array} \right) \\ \left( \begin{array}{l} \text{if } x = \mathbf{this} \wedge \text{class}(m) = C' \text{ then } M^\sharp(m')[\mathbf{post}] \sqcap T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1) \\ \text{else } T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1) \end{array} \right) \\ \left( \begin{array}{l} \text{if } x = \mathbf{this} \wedge \text{class}(m) = C' \text{ then } T^\sharp(m, i) \sqsubseteq M^\sharp(m')[\mathbf{this}] \\ \text{else } \rho(C', \llbracket x \rrbracket^\sharp) \sqsubseteq M^\sharp(m')[\mathbf{this}] \end{array} \right) \\ \llbracket x \rrbracket^\sharp :: \llbracket e_1 \rrbracket^\sharp :: \dots :: \llbracket e_j \rrbracket^\sharp \sqsubseteq M^\sharp(m')[\text{args}] \end{array}}{M^\sharp, H^\sharp, T^\sharp, L^\sharp \models (m, i) : x.\text{ms}(e_1, \dots, e_j)} \quad (6) \\
 \dots \\
 \frac{\begin{array}{c} \forall m, m', \text{overrides}(m', m) \Rightarrow M^\sharp(m)[\text{args}] \sqsubseteq M^\sharp(m')[\text{args}] \\ \forall m, m', \text{overrides}(m', m) \Rightarrow \top_{\text{class}(m)} \sqsubseteq M^\sharp(m)[\mathbf{post}] \\ \forall m, m', \text{overrides}(m', m) \Rightarrow \top_{\text{class}(m')} \sqsubseteq M^\sharp(m')[\mathbf{this}] \\ \forall m, M^\sharp(m)[\text{args}] \sqsubseteq L^\sharp(m, 0) \\ \forall m, M^\sharp(m)[\mathbf{this}] \sqsubseteq T^\sharp(m, 0) \\ \forall m, \forall i, M^\sharp, H^\sharp, T^\sharp, L^\sharp \models (m, i) : P_m[i] \end{array}}{M^\sharp, H^\sharp, T^\sharp, L^\sharp \models P} \quad (7)
 \end{array}$$

## Auxiliary Operators

$$\begin{aligned}
 \rho(C, \text{NotNull}) &= \perp_C \\
 \rho(C, \text{Raw}(X)) &= \text{if } X \preceq C \text{ then } \perp_C \text{ else } \top_C \\
 \rho(C, \text{MayBeNull}) &= \rho(C, \text{Raw}^-) = \top_C \\
 \llbracket x \rrbracket_{C, H^\sharp, t^\sharp, l^\sharp}^\sharp &= \begin{cases} \text{Raw}(C) & \text{if } x = \mathbf{this} \wedge t^\sharp = \perp_C \wedge l^\sharp(x) = \text{Raw}(\text{super}(C)) \\ L^\sharp(x) & \text{otherwise} \end{cases} \\
 \llbracket e.f \rrbracket_{C, H^\sharp, t^\sharp, l^\sharp}^\sharp &= \begin{cases} H^\sharp(f) & \text{if } \llbracket e \rrbracket_{C, H^\sharp, t^\sharp, l^\sharp}^\sharp = \text{NotNull} \text{ or } (e = \mathbf{this} \wedge t^\sharp(f) = \text{Def}) \\ & \text{or } \llbracket e \rrbracket_{C, H^\sharp, t^\sharp, l^\sharp}^\sharp = \text{Raw}(X) \text{ with } X \preceq \text{class}(f) \\ \text{MayBeNull} & \text{otherwise} \end{cases}
 \end{aligned}$$

Fig. 4. Analysis specification

deletekeywords=not,prefix,value,where,morekeywords=assert,asr,class,closed,constraint,external,false,functor,inclu

```

class A {
    private Object f;
    private Object g;

    A(Object o){ f = o;}

    void m(){g=f;}

    public static void main(String args []){
        Object o = new Object ();
        A a = new A(o);
        a.m();
    }
}

```

**Fig. 5.** Java source

$t^\#$  of the fields (declared in class  $C$ ) of the **this** object and an abstraction  $l^\#$  of the local variables. The first equation states that the type of a local variable is obtained from the  $L^\#$  function and can be refined from  $\text{Raw}(\text{super}(C))$  to  $\text{Raw}(C)$  if **this** is sufficiently initialized. The second equation states that the type of a field is obtained from the heap abstraction if the type of the reference is `NotNull` or if it is a reference to an object sufficiently deeply initialized. Otherwise, it is `MaybeNull`. In the inference rules we write  $\llbracket e \rrbracket^\#$  for  $\llbracket e \rrbracket_{\text{class}(m), H^\#, T^\#(m,i), L^\#(m,i)}$

### 3.3 Example

The Java source code of our example is provided in Fig. 5, while Fig. 6 shows the code in the syntax defined in Sect. 2 and the constraints obtained from the rules defined in Sect. 3.2. This example is fairly simple and, for conciseness, the code of the `main` method has been omitted, but the generated constraints should be sufficient to give an idea of how the inference works.

The labels in the source starting with the character `@` are the annotations inferred by the analysis for the fields and the methods signatures. An annotation in front of a method corresponds to the property of **this** before the invocations of the method. The other annotations are placed just before the variable they refer to.

Lines 6, 7, 24 and 25 list the constraints obtained from rule (7) in Fig. 4. They “initialize” the flow-sensitive abstractions  $L^\#(m,0)$  and  $T^\#(m,0)$  with the information of the method signatures. All other constraints are directly deduced from the rule corresponding to the instruction where conditional constraints have been simplified where it was possible. Lines 9 to 12 list the constraints obtained from rule (6) when  $\text{name}(m') = \text{init}$  and  $\text{class}(m) \neq \text{class}(m')$ . Lines 14 to 16 correspond to a field assignment on **this** of a field defined in the current class.

```

1: class A {
2:   Object @NotNull f;
3:   Object @MayBeNull g;
4:
5:   @RawTop void init(Object @NotNull o){
6:      $M^\sharp(A.\text{init})[\text{args}] \sqsubseteq L^\sharp(A.\text{init}, 0)$ 
7:      $M^\sharp(A.\text{init})[\text{this}] \sqsubseteq T^\sharp(A.\text{init}, 0)$ 
8:     0: this.{|Object,init,[], -> void|}()
9:        $\rho(\text{Object}, \llbracket \text{this} \rrbracket) \sqsubseteq M^\sharp(\text{Object}.\text{init})[\text{this}]$ 
10:      []  $\sqsubseteq M^\sharp(\text{Object}.\text{init})[\text{args}]$ 
11:       $L^\sharp(A.\text{init}, 0)[\text{this} \mapsto \text{Raw}(\text{Object})] \sqsubseteq L^\sharp(A.\text{init}, 1)$ 
12:       $T^\sharp(A.\text{init}, 0) \sqsubseteq T^\sharp(A.\text{init}, 1)$ 
13:      1: this.f = o
14:         $\llbracket o \rrbracket \sqsubseteq H^\sharp(A.f)$ 
15:         $T^\sharp(A.\text{init}, 1)[A.f \mapsto \text{Def}] \sqsubseteq T^\sharp(A.\text{init}, 2)$ 
16:         $L^\sharp(A.\text{init}, 1) \sqsubseteq L^\sharp(A.\text{init}, 2)$ 
17:      2: return
18:        if  $T^\sharp(A.\text{init}, 2)(A.f) = \text{UnDef}$  then  $\text{MayBeNull} \sqsubseteq H^\sharp(A.f)$ 
19:        if  $T^\sharp(A.\text{init}, 2)(A.g) = \text{UnDef}$  then  $\text{MayBeNull} \sqsubseteq H^\sharp(A.g)$ 
20:         $T^\sharp(A.\text{init}, 2) \sqsubseteq M^\sharp(A.\text{init})[\text{post}]$ 
21:    }
22:
23:   @NotNull void m(){
24:      $M^\sharp(A.m)[\text{args}] \sqsubseteq L^\sharp(A.m, 0)$ 
25:      $M^\sharp(A.m)[\text{this}] \sqsubseteq T^\sharp(A.m, 0)$ 
26:     0: this.g = this.f
27:        $\llbracket \text{this.f} \rrbracket \sqsubseteq H^\sharp(A.g)$ 
28:        $T^\sharp(A.m, 0)[A.g \mapsto \text{Def}] \sqsubseteq T^\sharp(A.m, 1)$ 
29:        $L^\sharp(A.m, 0) \sqsubseteq L^\sharp(A.m, 1)$ 
30:     1: return
31:        $T^\sharp(A.m, 1) \sqsubseteq M^\sharp(A.m)[\text{post}]$ 
    }
  }

```

Fig. 6. Code with constraints

Lines 18 to 20 correspond to a **return** instruction of a constructor, which adds the value `MayBeNull` in the abstract heap for all (maybe) undefined fields, while line 31 corresponds to a **return** instruction of a non-constructor method.

The methods are called from the `main` method of Fig. 5. The method `init` is annotated with  $\text{Raw}^-$  (`@RawTop`) as it is called on a completely uninitialized value. The constraint line 11 then refine  $\text{Raw}^-$  to  $\text{Raw}(\text{Object})$  so if a field of `Object` were accessed in the rest of the method, the abstraction of the heap would be used. The `NotNull` (`@NotNull`) annotation before the argument of the method `init` is first transferred to a local variable at line 6 and then moved from `o` to  $H^\sharp(A.f)$  at line 14. At the same time, line 15 records that the field `f`

is defined. Then, line 18, as  $\mathbf{f}$  has been defined  $H^\sharp(\mathbf{A.f})$  is not modified whereas, line 19,  $T^\sharp(\mathbf{A.init}, 2)(\mathbf{A.g}) = \text{UnDef}$  so  $H^\sharp(\mathbf{A.g})$  is constrained to  $\text{MayBeNull}$ .

## 4 Correctness

In this section we prove the correctness of the analysis. We first define (see Fig. 7) the logical link between concrete and abstract domains via a correctness relation. Then we prove (Theorem 1) that any solution of the constraint system which verifies the predicate  $\text{safe}_p^\sharp$  enforces the null-pointer error safety property for the set of preconditions inferred by the analysis. The result mainly relies on a suitable subject reduction lemma (Lemma 1). The theorem has been mechanically proved with the Coq proof assistant<sup>6</sup>.

$$\begin{array}{c}
\frac{v \in \text{Val}}{\text{Def} \sim (v, \mathbf{def})} \quad \frac{v \in \text{Val} \quad d \in \{\mathbf{def}, \mathbf{undef}\}}{\text{UnDef} \sim (v, d)} \\
\frac{v \in \text{dom}(h) \quad \forall f \in \text{dom}(h(v)), \text{IsDef}(h(v)(f))}{\text{NotNull} \sim_h v} \quad \frac{v \in \text{Val}}{\text{MayBeNull} \sim_h v} \\
\frac{v \in \text{dom}(h) \quad \forall f \in \bigcup_{A \leq C} \text{fields}(C) \cap \text{dom}(h(v)), \text{IsDef}(h(v)(f))}{\text{Raw}(A) \sim_h v} \quad \frac{v \neq \mathbf{null}}{\text{Raw}^- \sim_h v} \\
\frac{\forall x, l(x) = \perp \vee L^\sharp(x) \sim_h l(x)}{L^\sharp \sim_h l} \quad \frac{r \in \text{dom}(h) \quad \text{fields}(C) \subseteq \text{dom}(T^\sharp) \cap \text{dom}(h(r))}{\forall f \in \text{fields}(C), T^\sharp(f) \sim_h(r)(f)} \\
\frac{\forall r \in \text{dom}(h), \forall f \in \text{dom}(h(r)), h(r)(f) = (v, d) \Rightarrow d = \mathbf{undef} \vee H^\sharp(f) \sim_h v}{L^\sharp \sim_h l \quad T^\sharp \sim_{h,C} r} \\
\frac{o = l(\mathbf{this}) \quad L^\sharp(m, i) \sim_h l \quad H^\sharp \sim_h}{T^\sharp(m, i) \sim_{h, \text{class}(m)} l(\mathbf{this}) \quad H^\sharp \sim_h} \\
\frac{M^\sharp(m)[\text{post}] \sim_{h, \text{class}(m)} o \quad H^\sharp \sim_h}{(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m,o} \langle i, l, h \rangle} \\
\frac{}{(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m,o} h} \quad \frac{}{(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m,o} \Omega}
\end{array}$$

**Fig. 7.** Correctness relations

An abstract element  $(M^\sharp, H^\sharp)$  induces a set of method preconditions defined by

$$\text{Pre}(M^\sharp, H^\sharp)(m) = \left\{ (l, h) \mid \begin{array}{l} H^\sharp \sim_h, M^\sharp(m)[\mathbf{this}] \sim_{\text{class}(m), h} l(\mathbf{this}), \\ V_0^\sharp \sim_h l(\mathbf{this}) \text{ and } \forall i = 1..n, V_i^\sharp \sim_h l(i) \end{array} \right\} \\
\text{with } M^\sharp(m)[\mathbf{args}] = V_0^\sharp :: \dots :: V_n^\sharp$$

Note that a method  $m$  which is not called by any other method in the program, will be associated with a non-null assumption on all its parameters. This corresponds to the *non-null by default* approach advocated by Chalin and James [2].

We write  $\rightarrow_m^{(n)}$  (resp  $\rightarrow_m^{(n)*}$ ) for the operational step relation (resp. transitive closure) where exactly  $n$  method calls are performed during step (resp. transitive closure), including sub-calls.

<sup>6</sup> The proof is available at <http://www.irisa.fr/lande/pichardie/np/>

**Lemma 1 (Subject reduction).** *Let  $(M^\#, H^\#, T^\#, L^\#)$  such that  $M^\#, H^\#, T^\#, L^\# \models P$  and  $\text{safe}_P^\#(M^\#, H^\#, T^\#, L^\#)$ . Let  $n$  be an integer such that for all  $k, k < n$ , and all methods  $m$ , local variables  $l$ , heap  $h$  and configuration  $X$  if  $(l, h) \in \text{Pre}(M^\#, H^\#)(m)$  and  $\langle 0, l, h \rangle \xrightarrow{m}^{(k)*} X$  then  $(M^\#, H^\#, T^\#, L^\#) \sim_{m,l(\text{this})} X$  and  $X \neq \Omega$ .*

*Let  $p$  be an integer,  $i$  a program counter,  $l$  local variables,  $h$  a heap and  $X$  a configuration such that  $\langle i, l, h \rangle \xrightarrow{m}^{(p)} X$ ,  $(M^\#, H^\#, T^\#, L^\#) \sim_{m,o} \langle i, l, h \rangle$  and  $p \leq n$  then  $(M^\#, H^\#, T^\#, L^\#) \sim_{m,o} X$  and  $X \neq \Omega$ .*

*Proof.* See Appendix C

**Theorem 1 (constraint system soundness).** *If there exists  $(M^\#, H^\#, T^\#, L^\#)$  such that  $M^\#, H^\#, T^\#, L^\# \models P$  and  $\text{safe}_P^\#(M^\#, H^\#, T^\#, L^\#)$  holds then  $P$  is null-pointer error safe w.r.t the set of preconditions  $\text{Pre}(M^\#, H^\#)$ .*

*Proof.* The proof proceeds by induction on the maximum number of method calls in the execution sequences and then another induction on the length of intra-method derivation, in order to be able to conclude with the subject reduction Lemma 1.

## 5 Inference

Expressing the analysis in terms of constraints over lattices has the immediate advantage that inference can be obtained from standard iterative constraint solving techniques for static analyses. Proposition 1 asserts that there is a decision procedure to detect programs which are verifiable with the analysis.

**Proposition 1.** *For all program  $P$  there exists an algorithm which decides if there exists  $(M^\#, H^\#, T^\#, L^\#)$  such that  $M^\#, H^\#, T^\#, L^\# \models P$  and  $\text{safe}_P^\#(M^\#, H^\#, T^\#, L^\#)$ .*

*Proof.* It is a standard proof since the system of constraint is composed of monotone functions on a finite lattice. The least solution can then be computed and be checked with respect to predicate  $\text{safe}_P^\#$ .

Furthermore, the present analysis is modular in the sense that, rather than performing an analysis of all classes of a program, it is possible to describe certain classes (e.g., classes coming from a library) by providing *interfaces* consisting of some method signatures  $M^\#(m)$ ,  $m \in M^{\text{fix}}$  and field invariants  $H^\#(f)$ ,  $f \in F^{\text{fix}}$  relative to the classes.<sup>7</sup> The validity of these signatures can be established once and for all by the modular type checker proposed by Fähndrich and Leino (which works class by class). The analysis of partial programs would then proceed in several stages. First, the constraint system is generated only for the available classes. The system may refer to the variables  $M^\#(m)$ ,  $m \in M^{\text{fix}}$  and  $H^\#(f)$ ,

<sup>7</sup> Modular inference is less precise: to keep the same precision the analysis would need richer annotations for the libraries.

$f \in F^{\text{fix}}$ . Then the partial system is solved starting the iteration *à la* Kleene from

$$(M_0^\#, H_0^\#, T_0^\#, L_0^\#) = (\perp \sqcup \{m \mapsto M^\#(m)\}_{m \in M^{\text{fix}}}, \perp \sqcup \{f \mapsto H^\#(f)\}_{f \in F^{\text{fix}}}, \perp, \perp)$$

If one of the variables  $M^\#(m)$ ,  $m \in M^{\text{fix}}$  or  $H^\#(f)$ ,  $f \in F^{\text{fix}}$  has to be updated during the iteration, the constraint resolution fails, since this means there does not exist a solution compatible with the proposed signatures.

**Theorem 2 (Relative completeness of the modular solver).** *If the previous algorithm halts then the partial constraint system has no solution compatible with the signatures provided for the unknown classes.*

*Proof.* Let  $S$  denote the set

$$\{(M^\#, H^\#, T^\#, L^\#) \mid M^\#, H^\#, T^\#, L^\# \models P \text{ and } (M_0^\#, H_0^\#, T_0^\#, L_0^\#) \sqsubseteq (M^\#, H^\#, T^\#, L^\#)\}$$

Suppose the previous algorithm halts. Since the iteration is ascending it means the least solution of  $S$  is necessarily strictly greater than  $(M_0^\#, H_0^\#, T_0^\#, L_0^\#)$ . Since any solution compatible with the signatures  $M^\#(m)$ ,  $m \in M^{\text{fix}}$  or  $H^\#(f)$ ,  $f \in F^{\text{fix}}$  is in  $S$ , we can conclude that such a solution does not exist.

## 6 Fährdrich and Leino's Type System

In this section we compare our analysis with the type system proposed by Fährdrich and Leino [6]. This comparison relies on a formal definition<sup>8</sup> of their notion of typable program.

Figure 8 presents the corresponding typing rules. A typing judgment for expression  $e$  is of the form  $\Gamma, L \vdash e : \tau$  with  $\Gamma$  a type annotation for fields and methods,  $L$  a type annotation for local variables and  $\tau$  a type in  $\text{Val}^\#$ . Each instruction  $\text{instr}$  is also associated to a type judgement  $\Gamma, m \vdash \text{instr} : L \rightarrow L'$  where  $m$  is the current method,  $L$  the current annotation for local variables and  $L'$  a valid annotation after the execution of  $\text{instr}$ .  $[\text{inits } F]$  is a method annotation that indicates the method initializes the fields in the set  $F$ . If a method initializes all the fields defined in its class, it is also written  $[\text{inits}]$ . A program is said *well-typed* w.r.t.  $\Gamma$  if there are no overridden methods annotated as  $[\text{inits } F]$ , arguments are contravariant and there exists  $L$  such that for all methods  $m$  and for all program points  $i$ , either  $P_m[i] = \text{return}$  or for all successors points  $j$  of  $i$ , there exists  $L'$  such that  $L' \sqsubseteq L(m, j)$  and  $\Gamma, m \vdash P_m[i] : L(m, i) \rightarrow L'$ .

This type system is coupled with a data flow analysis to ensure that all fields not declared as MayBeNull in class  $C$  are sure to be defined at the end of all constructors of  $C$ . It is a standard data flow analysis, described in Fig. 8.  $F(m, i)$  represents the fields that are not initialized at program point  $(m, i)$ . At

<sup>8</sup> Note that in their paper the authors only propose an informal definition so what we formalise here is only our interpretation of their work.

the beginning of every constructor `m`, it is constrained to the set of fields defined in the class of `m`. For field assignation on `this`, the field is not propagated to the next node. For method calls to methods tagged as `[inits F]`, fields in `F` are not propagated to the next node, but the set of undefined fields is propagated to the first instruction of the callee. It is then checked that those tagged methods initialize their fields.

**Theorem 3.** *If  $P$  is FL-typable then there exists  $(M^\#, H^\#, T^\#, L^\#)$  such that  $M^\#, H^\#, T^\#, L^\# \models P$  and  $\text{safe}_P^\#(M^\#, H^\#, T^\#, L^\#)$  holds.*

*Proof.* We show that if  $P$  is FL-typable for a given  $\Gamma$  and  $L$  then this type annotations represent a valid solution of the constraint system which furthermore satisfies the  $\text{safe}_P^\#$  property.

**Corollary 1 (FL type system soundness).** *If there exists  $\Gamma$  such that  $P$  is FL-typable then  $P$  is null-pointer error safe w.r.t. the preconditions given by  $\Gamma$ .*

*Proof.* Direct consequence of Theorem 3 and Theorem 1.

**Theorem 4.** *There exists  $P$  such that  $P$  is not FL-typable and there exists  $(M^\#, H^\#, T^\#, L^\#)$  such that  $M^\#, H^\#, T^\#, L^\# \models P$  and  $\text{safe}_P^\#(M^\#, H^\#, T^\#, L^\#)$  holds.*

*Proof.* As shown in Fig. 4, the analysis of expressions is parametrized on the abstraction of `this` in order to know if a field has already been defined or not. In Föhndrich and Leino's analysis, the type checking is separated from the data-flow analysis that knows which fields have already been defined. For example, in

```
class C {
  Object f; //NotNull
  Object g; //MaybeNull or NotNull?
  public C(){ this.f = new Object(); this.g = this.f;}
}
```

our analysis benefits from the abstraction of `this` and knows `this.f` has been initialized before it is assigned to `this.g`. In Föhndrich and Leino's analysis, an intermediate local variable set to the new object and affected to `f` and `g` or an explicit cast operator which checks the initialization at run-time would be needed in order to type the program. Our abstraction of `this` can also be passed from a method to another, here also keeping some more information as the intra-procedural data-flow of Föhndrich and Leino.

*Gain in practice* To evaluate the practical gain of our analysis w.r.t the type system of Föhndrich and Leino, we count the number of program points that evaluate  $\llbracket e.f \rrbracket^\#$  where  $\neg(\llbracket e \rrbracket^\# = \text{NotNull} \vee \llbracket e \rrbracket^\# = \text{Raw}(X) \text{ with } X \preceq \text{class}(f))$ , to avoid the cases handled by the type system, and  $e = \text{this} \wedge T^\#(f) = \text{Def}$  where our analysis use a more precise information than the type system, *i.e.* the inferred type instead of `MaybeNull`.

*Expression Typability*

$$\begin{array}{c}
\frac{}{\Gamma, L \vdash x : L(x)} \quad \frac{\Gamma, L \vdash e : \text{NotNull}}{\Gamma, L \vdash e.f : \Gamma[f]} \quad \frac{\Gamma, L \vdash e : \text{Raw}^-}{\Gamma, L \vdash e.f : \text{MaybeNull}} \\
\frac{\text{class}(f) \prec C}{\Gamma, L \vdash e : \text{Raw}(C)} \quad \frac{C \preceq \text{class}(f)}{\Gamma, L \vdash e : \text{Raw}(C)} \quad \frac{\Gamma, L \vdash e : A \quad A \sqsubseteq B}{\Gamma, L \vdash e : B} \\
\frac{}{\Gamma, L \vdash e.f : \text{MaybeNull}} \quad \frac{}{\Gamma, L \vdash e.f : \Gamma[f]} \quad \frac{}{\Gamma, L \vdash e : B}
\end{array}$$

*Instruction Typability*

$$\begin{array}{c}
\frac{\Gamma, L \vdash e : \tau}{\Gamma, m \vdash x \leftarrow e : L \rightarrow L[x \mapsto \tau]} \quad \frac{\Gamma, L \vdash e : \tau \quad \tau \sqsubseteq \Gamma[f] \quad L(x) \neq \text{MaybeNull}}{\Gamma, m \vdash x.f \leftarrow e : L \rightarrow L} \\
\frac{\Gamma[\{C, \mathbf{init}, \alpha \rightarrow \mathbf{void}\}] = [\text{Raw}^-]\beta \rightarrow \mathbf{void} \quad (\forall i, \Gamma, L \vdash e_i : \beta_i)}{\Gamma, m \vdash x \leftarrow \mathbf{new}(C, \alpha)(e_1, \dots, e_n) : L \rightarrow L[x \mapsto \text{NotNull}]} \\
\begin{array}{l}
m' = \text{lookup}(ms) \quad \Gamma[m'] = [\tau]\beta \rightarrow \mathbf{void} \quad (\forall i, \Gamma, L \vdash e_i : \beta_i) \quad \tau \neq \text{MaybeNull} \\
C' = \text{class}(m') \quad \Gamma, L \vdash x : \text{if } \tau \in \text{Val}^\sharp \text{ then } \tau \text{ else Raw}(\text{super}(C')) \\
\tau = \begin{cases} \tau \sqcap \text{Raw}(C') & \text{if name}(m') = \mathbf{init} \\ \text{Raw}(C') & \text{if } \tau = \mathbf{inits} F \wedge \text{fields}(C') \subseteq F \wedge \text{class}(m) = C' \wedge x = \mathbf{this} \\ \tau & \text{otherwise} \end{cases}
\end{array} \\
\frac{}{\Gamma, m \vdash x.ms(e_1, \dots, e_n) : L \rightarrow L[x \mapsto \tau']} \\
\frac{}{\Gamma, m \vdash \mathbf{if} \text{ jmp} : L \rightarrow L}
\end{array}$$

*Data Flow Analysis to Ensure All NotNull Fields Are Defined*

$$\begin{array}{c}
F : \mathbb{M} \times \mathbb{N} \rightarrow \mathcal{P}(\mathbb{F}) \\
\text{inits}_{m, m', x, F} \equiv \Gamma[m'] = [\mathbf{inits} F]\alpha \rightarrow \mathbf{void} \wedge \text{class}(m) = \text{class}(m') \wedge x = \mathbf{this} \\
\frac{F(m, i) \setminus \{f\} \subseteq F(m, i+1) \quad x \neq \mathbf{this} \quad F(m, i) \subseteq F(m, i+1)}{\Gamma, F \models (m, i) : \mathbf{this}.f \leftarrow e} \quad \frac{}{\Gamma, F \models (m, i) : x.f \leftarrow e} \\
\frac{F(m, i) \subseteq F(m, i+1)}{\Gamma, F \models (m, i) : x \leftarrow e} \quad \frac{F(m, i) \subseteq F(m, i+1)}{\Gamma, F \models (m, i) : x \leftarrow \mathbf{new}(C, \alpha)(e_1, \dots, e_n)} \\
\frac{m' = \text{lookup}(ms) \quad \text{inits}_{m, m', x, F'} \quad F(m, i) \setminus F' \subseteq F(m, i+1)}{\Gamma, F \models (m, i) : x.ms(e_1, \dots, e_n)} \\
\frac{m' = \text{lookup}(ms) \quad \neg(\exists F', \text{inits}_{m, m', x, F'}) \quad F(m, i) \subseteq F(m, i+1)}{\Gamma, F \models (m, i) : x.ms(e_1, \dots, e_n)} \\
\frac{F(m, i) \subseteq F(m, i+1) \quad F(m, i) \subseteq F(m, \mathbf{jmp})}{\Gamma, F \models (m, i) : \mathbf{if}(\star) \mathbf{jmp}} \\
\frac{\text{name}(m) = \mathbf{init} \quad \forall f \in F(m, i).(\text{MaybeNull} \sqsubseteq \Gamma[f])}{\Gamma, F \models (m, i) : \mathbf{return}} \quad \frac{\Gamma[m] = [\mathbf{inits} F]\alpha \rightarrow \mathbf{void} \quad F \cap F(m, i) = \emptyset}{\Gamma, F \models (m, i) : \mathbf{return}} \\
\frac{\forall m, m = \{C, \mathbf{init}, \alpha \rightarrow \mathbf{void}\} \vee \Gamma[m] = [\mathbf{inits} F]\alpha \rightarrow \mathbf{void} \Rightarrow \text{fields}(\text{class}(m)) \subseteq F(m, 0) \wedge \forall i, \Gamma, F \models (m, i) : P_m[i]}{\Gamma, F \models P}
\end{array}$$

**Fig. 8.** Fähndrich and Leino's type system

In the Soot suite we have analysed [15] (see the next section), on 800 analyzed constructors, 391 program points have been found in 45 different constructors (of which 16 were part of the run-time). This does not take in account the gain obtained in methods called from a constructor.

## 7 Towards a Null-Pointer Analyzer for Java Bytecode

A prototype of the analysis has been implemented. The analysis works at the Java bytecode (JVML) level but annotations on fields and method can be propagated at the source level with a minor effort.

Working in the fragment of JVML without exceptions does not modify the analysis but some points need to be taken in account. JVML is a **stack language**, we therefore need to track aliases of **this** on the stack to know when a field manipulation or a method call is actually done on **this**. Operations on **numbers** do not add any difficulty as JVML is typed and numbers are guaranteed not to be assigned to references. The **multi-threading** cannot interfere with weak updates, but it could interfere with the strong updates used on the abstraction of **this**. However, as the initialization evolves monotonically (*i.e.* a field once defined cannot be reverted to undefined) it is still correct. **Static methods** do not have an abstraction of **this** and do not add any other difficulty to handle.

**Static fields** are difficult to analyse precisely as they are initialized through `<clinit>` methods, which are not explicitly called in the code but executed by the Java Virtual Machine (at loading-time) in an unspecified order in the general case (in particular in case of circular dependencies between the classes). It should be noted here that strong updates on static fields could be done if there were no multi-threading. **Exceptions** in Java can be Raw object, *i.e.* it is possible in a constructor to throw **this** as an exception, although we have not found examples of this yet. We could imagine a solution where abstract method signatures would include the type of the exceptions a method can throw. Tracking initialization of **array** cells precisely is also challenging as checking that all cells have been initialized requires numerical abstraction. We currently and conservatively annotate static fields and array cells as `MayBeNull` and references to caught exception objects as `Raw-` and left their precise handling for future work.

Other small optimizations (that have not been formalized) are possible. For example, the abstraction of **this** could be done for all other references of type `Raw(X)` (or `Raw-`). It would give a more precise information on fields in general.

The first performance tests are promising as large programs such as Javacc 4.0 and Soot 2.2.4 can be analysed in about 40 seconds and 5 minutes respectively. The analysis still needs to be completed with a path-sensitive analysis [4] to recover non-nullness information from the conditionals. Table 1 gives the execution times for Soot 2.2.4 (520.651 instructions), Jasmin (104.810 instructions), which provided with Soot 2.2.4, and Javacc 4.0 (160.942 instructions) with the Java run-time provided with `gij 4.0.2`. The test machine was a MacBook Pro

with a 2.4GHz Intel Core 2 Duo with 2GB of RAM, MacOS 10.5 and OCaml 3.10.0. Table 2 gives the current results of the inference, but as the tool as not been properly tested yet, results may evolve in the future. The table gives, for each program, the number of fields that have been annotated as `NotNull`, `Raw` or `MayBeNull` and that are actually read somewhere in the program (fields never read are not counted as they mostly correspond to dead code and do not risk to produce any exception). Fields declared in the run-time are not counted.

**Table 1.** Performance of the inference of non-null annotations (in seconds)

	loading and parsing	collecting (alias)	solving (alias)	collecting (NNA)	solving (NNA)	saving	total
jasmin	1.5	1.1	1.2	2.8	5.2	8.6	20.4
javacc	2.0	1.6	2.0	4.7	9.1	20.4	39.8
soot	6.3	21.5	10.5	48.8	63.2	139.5	289.8

**Table 2.** Results: annotations inferred for fields actually read in the code

	NotNull	Raw	MayBeNull
jasmin	70 (44.9%)	4 (2.6%)	82 (55.5%)
javacc	46 (44.2%)	0 (0.0%)	58 (55.8%)
soot	1389 (44.0%)	158 (5.0%)	1608 (51.0%)

## 8 Conclusions and Future Work

We have defined a semantics-based analysis and inference technique for automatically inferring non-null annotations for fields. The analysis has been proved correct and the correctness proof has been machine-checked in the proof assistant Coq. This extends and complements the seminal paper of Fähndrich and Leino in which is proposed an extended type system for verifying non-null type annotations. Fähndrich and Leino’s approach mixes type system and data-flow analysis. In our work, we follow an abstract interpretation methodology to gain strong semantic foundations and a goal-directed inference mechanism to find a minimal (*i.e.* principal) non-null annotation. By the same token we also gained in precision thanks to a better communication between abstract domains. We then proved the correctness of our analysis and its completeness with respect to their type system.

Variations of the present analysis can be envisaged. For example, in our analysis, preconditions for methods are computed as the least upper bounds of the conditions verified at call points. It corresponds more to the *actual* behaviour of the program than to the *intention* of the developer and in some cases, weaker preconditions would still be safe. Another approach would be to infer the weakest preconditions that prevents null-pointer exceptions. It would be specially useful to annotate libraries.

In the future, we also plan to extend our analysis and its correctness proof to the full Java bytecode language. To manage this substantial extension it is important to be able to machine-check the correctness proof, which will necessarily be large. Our previous experience with developing a certified static analyser for Java [1] using the Coq proof assistant leads us to believe that such a formalisation is indeed feasible.

## References

1. D. Cachera, T. P. Jensen, D. Pichardie, and V. Rusu. Extracting a data flow analyser in constructive logic. *Theoretical Computer Science*, 342(1), 2005.
2. P. Chalin and P. R. James. Non-null references by default in Java: Alleviating the nullity annotation burden. In *Proceedings of European Conference on Object-Oriented Programming (ECOOP'07)*.
3. M. Cielecki, J. Fulara, K. Jakubczyk, and L. Jancewicz. Propagation of JML non-null annotations in Java programs. In *Proceedings of the 4th international symposium on Principles and practice of programming in Java (PPPJ '06)*.
4. M. Das, S. Lerner, and M. Seigle. Esp: path-sensitive program verification in polynomial time. In *Proceedings of the ACM SIGPLAN Conference on Programming language design and implementation (PLDI'02)*.
5. T. Ekman and G. Hedin. Pluggable non-null types for Java. In Torbjörn Ekman, editor, *Extensible Compiler Construction*, chapter V. Lund University, June 2006.
6. M. Fähndrich and K. R. M. Leino. Declaring and checking non-null types in an object-oriented language. In *Proceedings of Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA'03)*.
7. M. Fähndrich and S. Xia. Establishing object invariants with delayed types. In *OOPSLA '07: Proceedings of the 22nd ACM conference on Object Oriented Programming Systems and Applications*.
8. C. Flanagan and K. R. M. Leino. Houdini, an annotation assistant for ESC/Java. In *Proceedings of International Symposium of Formal Methods Europe (FME'01)*.
9. S. N. Freund and J. C. Mitchell. A formal framework for the java bytecode language and verifier. In *Proceedings of the 14th ACM conference on Object-oriented programming, systems, languages, and applications (OOPSLA '99)*.
10. D. Hovemeyer and W. Pugh. Finding more null pointer bugs, but not too many. In *PASTE '07: Proceedings of the 7th ACM SIGPLAN-SIGSOFT workshop on Program analysis for software tools and engineering*.
11. D. Hovemeyer, J. Spacco, and W. Pugh. Evaluating and tuning a static analysis to find null pointer bugs. *SIGSOFT Softw. Eng. Notes*, 31(1), 2006.
12. M. Kawahito, H. Komatsu, and T. Nakatani. Effective null pointer check elimination utilizing hardware trap. *SIGPLAN Not.*, 35(11), 2000.

13. K. R. M. Leino, J. B. Saxe, and R. Stata. *ESC/Java user's manual*. Compaq Systems Research Center, technical note 2000-002 edition, October 2000.
14. Chris Male, David J. Pearce, Alex Potanin, and Constantine Dymnikov. Java bytecode verification for @NonNull types. In *Proceedings of the Conference on Compiler Construction (CC'08)*.
15. R. Vallée-Rai, P. Co, E. Gagnon, L. Hendren, P. Lam, and V. Sundaresan. Soot — a Java bytecode optimization framework. In *CASCON '99: Proceedings of the 1999 conference of the Centre for Advanced Studies on Collaborative research*.

## A Abstract safety

$$\begin{array}{c}
 \frac{\text{safe}_{\text{expr}}^{\sharp}(x)_{C,H^{\sharp},T^{\sharp},L^{\sharp}}}{\forall m, \forall i,} \quad \frac{\frac{\llbracket e \rrbracket_{C,H^{\sharp},T^{\sharp},L^{\sharp}}^{\sharp} \neq \text{MaybeNull}}{\text{safe}_{\text{expr}}^{\sharp}(e)_{C,H^{\sharp},T^{\sharp},L^{\sharp}}}}{\text{safe}_{\text{expr}}^{\sharp}(e.f)_{C,H^{\sharp},T^{\sharp},L^{\sharp}}} \\
 \frac{P_m[i] = x \leftarrow e \implies \text{safe}_{\text{expr}}^{\sharp}(e)_{\text{class}(m),H^{\sharp},T^{\sharp}(m,i),L^{\sharp}(m,i)}}{P_m[i] = x.f \leftarrow e \implies L^{\sharp}(m,i)(x) \neq \text{MaybeNull}} \\
 \frac{P_m[i] = x \leftarrow \mathbf{new} (C, \alpha)(e_1, \dots, e_j) \implies \bigwedge \text{safe}_{\text{expr}}^{\sharp}(e)_{\text{class}(m),H^{\sharp},T^{\sharp}(m,i),L^{\sharp}(m,i)}}{\bigwedge \text{safe}_{\text{expr}}^{\sharp}(e_j)_{\text{class}(m),H^{\sharp},T^{\sharp}(m,i),L^{\sharp}(m,i)}} \\
 \frac{P_m[i] = x.m(e_1, \dots, e_n) \implies L^{\sharp}(m,i)(x) \neq \text{MaybeNull}}{\bigwedge \bigwedge \text{safe}_{\text{expr}}^{\sharp}(e_j)_{\text{class}(m),H^{\sharp},T^{\sharp}(m,i),L^{\sharp}(m,i)}} \\
 \hline
 \text{safe}^{\sharp}(M^{\sharp}, H^{\sharp}, T^{\sharp}, L^{\sharp})
 \end{array}$$

For all expression  $e$  we note  $\Gamma[e]$ :

$$\begin{aligned}
 \Gamma[x] &= \Gamma[x] \\
 \Gamma[e.f] &= \Gamma[f]
 \end{aligned}$$

**Lemma 2.**  $e \Downarrow_{h,l} v$  implies  $v \triangleleft_h \Gamma[e]$ .

*Proof.* By inspection of the definition of  $\Downarrow_{h,l}$  and  $\triangleleft_h$ .

**Lemma 3.** If  $e \Downarrow_{h,l} v$ ,  $H^{\sharp} \sim h$ ,  $T^{\sharp} \sim_{\Gamma[\mathbf{this}],h} l(\mathbf{this})$  and  $L^{\sharp} \sim_h l$  then  $\llbracket e \rrbracket^{\sharp} \sim_{h,\Gamma[e]} v$  or  $v = \Omega$ .

By induction on  $e$ .

- If  $e = x$  then  $v = l(x) \neq \perp$  and  $v \triangleleft_h \Gamma[x]$ . Hence  $L^{\sharp}(x) \sim_{h,\Gamma[x]} v$ . We distinguish two cases:
  - if  $x = \mathbf{this}$ ,  $T^{\sharp} = \perp_{\Gamma[\mathbf{this}]}$  and  $L^{\sharp}(x) = \text{Raw}(\text{super}(\Gamma[\mathbf{this}]))$  then we have to show  $\text{Raw}(\Gamma[\mathbf{this}]) \sim_h v$ . First,  $\Gamma[\mathbf{this}] \bowtie \Gamma[e]$  since  $e = \mathbf{this}$ . Then since  $\perp_{\Gamma[\mathbf{this}]} \sim_{\Gamma[\mathbf{this}],h} l(x)$ , all fields in  $\text{fields}(\Gamma[\mathbf{this}])$  are initialized. Since furthermore  $L^{\sharp}(x) = \text{Raw}(\text{super}(\Gamma[\mathbf{this}]))$  we can conclude.
  - otherwise  $\llbracket e \rrbracket_{\Gamma[\mathbf{this}],H^{\sharp},T^{\sharp},L^{\sharp}}^{\sharp} = L^{\sharp}(x)$  and we conclude because  $L^{\sharp}(x) \sim_{h,\Gamma[e]} v$  holds.
- If  $e = e'.f$  we suppose that the property holds for  $e'$ . The case  $v = \Omega$  is easy. Since  $\text{MaybeNull} \sim_h v$  the only interesting case occurs when  $\llbracket e' \rrbracket_{\Gamma[\mathbf{this}],H^{\sharp},T^{\sharp},L^{\sharp}}^{\sharp} = \text{NotNull}$  or ( $e' = \mathbf{this} \wedge T^{\sharp}(f) = \text{Def}$ ) or  $\llbracket e' \rrbracket_{\Gamma[\mathbf{this}],H^{\sharp},T^{\sharp},L^{\sharp}}^{\sharp} = \text{Raw}(X)$  with  $X \preceq \text{class}(f)$ . In all these cases, the field  $f$  is necessarily initialized so  $H^{\sharp}(f) \sim_{h,\Gamma[f]} v$  holds.

**Lemma 4.** *If  $e \Downarrow_{h,l} v$ ,  $\text{safe}_{\text{expr}}^{\#}(e)_{\Gamma[\text{this}],H^{\#},T^{\#},L^{\#}}$ ,  $H^{\#} \sim h$ ,  $T^{\#} \sim_{\Gamma[\text{this}],h} l(\text{this})$  and  $L^{\#} \sim_h l$  then  $v \neq \Omega$ .*

*Proof.* By induction on the hypothesis  $e \Downarrow_{h,l} v$ . If  $e = x$ ,  $v$  is necessarily different from  $\Omega$ . If  $e = e'.f$ ,  $e' \Downarrow_{h,l} r$  and  $e \Downarrow_{h,l} h(r)(f)$ ,  $h(r)(f)$  is necessarily different from  $\Omega$ . Finally if  $e = e'.f$ ,  $e' \Downarrow_{h,l} v'$  with  $v' \in \{\mathbf{null}, \Omega\}$ , we have by definition of  $\text{safe}_{\text{expr}}^{\#}(e)$ ,  $\text{safe}_{\text{expr}}^{\#}(e')_{\Gamma[\text{this}],H^{\#},T^{\#},L^{\#}}$  and  $\llbracket e' \rrbracket_{\Gamma[\text{this}],H^{\#},T^{\#},L^{\#}}^{\#} \neq \text{MayBeNull}$ . Hence by lemma 3,  $v' \neq \mathbf{null}$  or  $v' = \Omega$ . But the case  $v' = \Omega$  is also impossible by induction hypothesis.

## B Technical lemmas

*Monotony of the correctness relations* All the correctness relations can be proved monotone by inspection of their definition and the associated partial order  $\sqsubseteq$ .

**Lemma 5.** *If  $V_1^{\#} \sim_{h,\tau} v$  and  $V_1^{\#} \sqsubseteq V_2^{\#}$  then  $V_2^{\#} \sim_{h,\tau} v$ .*

**Lemma 6.** *If  $L_1^{\#} \sim_h l$  and  $L_1^{\#} \sqsubseteq L_2^{\#}$  then  $L_2^{\#} \sim_h l$ .*

**Lemma 7.** *If  $T_1^{\#} \sim_{h,C} r$  and  $T_1^{\#} \sqsubseteq T_2^{\#}$  then  $T_2^{\#} \sim_{h,C} r$ .*

**Lemma 8.** *If  $H_1^{\#} \sim h$  and  $H_1^{\#} \sqsubseteq H_2^{\#}$  then  $H_2^{\#} \sim h$ .*

*Monotony w.r.t. to preceq*

**Lemma 9.** *If  $V^{\#} \sim_{h,\tau_1} v$  and  $\tau_1 \preceq \tau_2$  then  $V^{\#} \sim_{h,\tau_2} v$ .*

*Heap update*

**Lemma 10.** *If  $r \in \text{dom}(h)$  and  $V^{\#} \sim_{h,\tau} v$  then  $V^{\#} \sim_{\text{upd}(h,r,f,(v,\mathbf{def})),\tau} v$ .*

*Proof.* By case analysis on  $V^{\#} \sim_h v$ . Since  $r \in \text{dom}(h)$ ,  $\text{dom}(\text{upd}(h,r,f,(v,\mathbf{def}))) = \text{dom}(h)$  and we remark that a field which was initialized in  $h$  is still initialized in  $\text{upd}(h,r,f,(v,\mathbf{def}))$ . This allows us to conclude the proof.

**Lemma 11.** *If  $r \in \text{dom}(h)$  and  $L^{\#} \sim_h l$  then  $L^{\#} \sim_{\text{upd}(h,r,f,(v,\mathbf{def}))} l$ .*

*Proof.* Direct consequence of lemma 10 and definition of  $\sim_h$  on local variables.

**Lemma 12.** *If  $r \in \text{dom}(h)$  and  $T^{\#} \sim_{h,C} o$  then  $T^{\#} \sim_{\text{upd}(h,r,f,(v,\mathbf{def})),C} o$ .*

*Proof.* The property holds by definition of  $\sim$ , by remarking that  $\text{dom}(\text{upd}(h,r,f,(v,\mathbf{def}))) = \text{dom}(h)$  (since  $r \in \text{dom}(h)$ ) and by using lemma 10. We must also use the fact that for any  $D^{\#} \in \text{Def}^{\#}$  and  $v \in \text{Val}$ ,  $D^{\#} \sim (v,\mathbf{def})$ .

**Lemma 13.** *If  $r \in \text{dom}(h)$ ,  $f \in \text{dom}(h(r))$ ,  $H^{\#} \sim h$  and  $H^{\#}(f) \sim_{h,\Gamma[f]} v$  then  $H^{\#} \sim_{\text{upd}(h,r,f,(v,\mathbf{def}))}$ .*

*Proof.* By definition of  $\sim$  and lemma 10.

**Lemma 14.** *If  $T^{\#} \sim_{h,C} r$  then  $T^{\#}[f \mapsto \text{Def}] \sim_{\text{upd}(h,r,f,(v,\mathbf{def})),C} r$ .*

*Proof.* By inspection of the definition of  $\sim_{h,C}$ .

*Heap allocation*

**Lemma 15.** *If  $r \notin \text{dom}(h)$  and  $V^\sharp \sim_{h,\tau} v$  then  $V^\sharp \sim_{h[r \mapsto \text{default}(C)],\tau} v$ .*

*Proof.* By definition of  $V^\sharp \sim_h v$ , since it only depends of valid address in the heap.

**Lemma 16.** *If  $r \notin \text{dom}(h)$  and  $L^\sharp \sim_h l$  then  $L^\sharp \sim_{h[r \mapsto \text{default}(C)]} l$ .*

*Proof.* Direct consequence of lemma 15 and definition of  $\sim_h$  on local variables.

**Lemma 17.** *If  $r \notin \text{dom}(h)$  and  $T^\sharp \sim_{h,C} o$  then  $T^\sharp \sim_{h[r \mapsto \text{default}(C)],C} o$ .*

*Proof.* Similar proof as for lemma 16.

**Lemma 18.** *If  $r \notin \text{dom}(h)$ ,  $f \in \text{dom}(h(r))$  and  $H^\sharp \sim h$  then  $H^\sharp \sim h[r \mapsto \text{default}(C)]$ .*

*Proof.* For any  $r' \in \text{dom}(h[r \mapsto \text{default}(C)])$  either  $r \neq r'$  and the property still holds, or  $r = r'$  and the property holds because all fields in  $\text{default}(C)$  are **undef**.

Similar lemmas (for  $L^\sharp$  et  $T^\sharp$ ) hold for the heap transformation used at the end of constructors when all fields of an object are set to **def**. By iteration of the previous lemma we obtain the two following results.

**Lemma 19.** *If  $L^\sharp \sim_h l$  and  $\langle i_0, l_0, h \rangle \rightarrow_m^* h'$  then  $L^\sharp \sim_{h'} l$ .*

**Lemma 20.** *If  $T^\sharp \sim_{C,h} r$  and  $\langle i_0, l_0, h \rangle \rightarrow_m^* h'$  then  $T^\sharp \sim_{C,h'} r$ .*

## C Analysis soundness

**Lemma 21 (Subject reduction).** *Let  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp)$  such that  $M^\sharp, H^\sharp, T^\sharp, L^\sharp \models P$  and  $\text{safe}^\sharp(M^\sharp, H^\sharp, T^\sharp, L^\sharp)$ . Let  $n$  an integer such that for all  $k$ ,  $k < n$ , and all methods  $m$ , local variables  $l$ , heap  $h$  and configuration  $X$  if  $(l, h) \in \text{Pre}(M^\sharp, H^\sharp)(m)$  and  $\langle 0, l, h \rangle \rightarrow_m^{(k)*} X$  then  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m,l(\text{this})} X$  and  $X \neq \Omega$ .*

*Let  $p$  and  $n$  two integers,  $i$  a program counter,  $l$  some local variables,  $h$  a heap and  $X$  a configuration such that  $\langle i, l, h \rangle \rightarrow_m^{(p)} X$ ,  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m,o} \langle i, l, h \rangle$  and  $p \leq n$  then  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m,o} X$  and  $X \neq \Omega$ .*

*Proof.* We assume

$$\forall n, k, m, l, h, X, \quad (1)$$

$$k < n \wedge (l, h) \in \text{Pre}(M^\sharp, H^\sharp)(m) \wedge \langle 0, l, h \rangle \xrightarrow{m}^{(k)*} X \Rightarrow \\ (M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m, l(\mathbf{this})} X \wedge X \neq \Omega$$

$$\langle i, l, h \rangle \xrightarrow{m}^{(p)} X \quad (2)$$

$$M^\sharp, H^\sharp, T^\sharp, L^\sharp \models P \quad (3)$$

$$\text{safe}^\sharp(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \quad (4)$$

$$L^\sharp(m, i) \sim_h l \quad (5)$$

$$T^\sharp(m, i) \sim_{h, \text{class}(m)} l(\mathbf{this}) \quad (6)$$

$$H^\sharp \sim h \quad (7)$$

$$o = l(\mathbf{this}) \quad (8)$$

and make a case analysis on (2).

**Case 1.1:**  $\mathbf{P}_m[i] = \mathbf{x} \leftarrow \mathbf{e}$  and  $\mathbf{X} = \langle \mathbf{i} + 1, l[\mathbf{x} \mapsto \mathbf{v}], \mathbf{h} \rangle$ .

We immediately see that  $X \neq \Omega$ .

Furthermore, by hypothesis,

$$e \Downarrow_{h, l} v \quad (9)$$

$$x \neq \mathbf{this} \quad (10)$$

$$\Gamma[e] \preceq \Gamma[x] \quad (11)$$

and

$$T^\sharp(m, i) \sqsubseteq T^\sharp(m, i + 1) \quad (12)$$

$$L^\sharp(m, i)[x \mapsto \llbracket e \rrbracket_{\text{class}(m), H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}^\sharp] \sqsubseteq L^\sharp(m, i + 1) \quad (13)$$

We must show that  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m, o} \langle i + 1, l[x \mapsto v], h \rangle$  holds.

The property holds for the previous state and since the heap is not modified, we only have to show  $o = l[x \mapsto v](\mathbf{this})$ ,  $L^\sharp(m, i + 1) \sim_h l[x \mapsto v]$  and  $T^\sharp(m, i) \sim_{h, \text{class}(m)} l[x \mapsto v](\mathbf{this})$ . Since  $l[x \mapsto v](\mathbf{this}) = l(\mathbf{this})$ ,  $T^\sharp(m, i) \sim_{h, \text{class}(m)} l[x \mapsto v](\mathbf{this})$  is easily proved by monotonicity of  $\sim_{h, \text{class}(m)}$ , (12) and (6).  $o = l[x \mapsto v](\mathbf{this})$  holds also because of hypotheses (8) and (10). Now for any variable  $y$ , we must establish

$$l[x \mapsto v](y) = \perp \vee L^\sharp(m, i + 1)(y) \sim_{h, \Gamma[y]} l[x \mapsto v](y)$$

knowing that  $\forall y, l(y) = \perp \vee L^\sharp(m, i)(y) \sim_h l(y)$ . By monotonicity of  $\sim_h$  it is sufficient to prove

$$l[x \mapsto v](y) = \perp \vee L^\sharp(m, i)[x \mapsto \llbracket e \rrbracket_{\text{class}(m), H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}^\sharp](y) \sim_{h, \Gamma[y]} l[x \mapsto v](y)$$

The case  $y \neq x$  is trivial so we conclude by proving

$$\llbracket e \rrbracket_{\text{class}(m), H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}^\sharp \sim_{h, \Gamma[x]} v$$

thanks to Lemma 3 and 9 and hypotheses (9) and (11).

**Case 1.2:  $\mathbf{P}_m[i] = \mathbf{x} \leftarrow \mathbf{e}$  and  $\mathbf{X} = \Omega$ .**

We must show that this case is impossible. By hypothesis  $e \Downarrow_{h,l} \Omega$ . But this is contradictory with the result of lemma 4, using hypotheses (4), (5), (6) and (7).

**Case 2.1:  $\mathbf{P}_m[i] = \mathbf{x}.f \leftarrow \mathbf{e}$  and  $\mathbf{X} = \langle \mathbf{i} + 1, \mathbf{l}, \text{upd}(\mathbf{h}, \mathbf{l}(\mathbf{x}), \mathbf{f}, (\mathbf{v}, \mathbf{def})) \rangle$ .**

We immediately see that  $X \neq \Omega$ .

Furthermore, by hypothesis,

$$e \Downarrow_{h,l} v \quad (14)$$

$$l(x) \in \text{dom}(h) \quad (15)$$

$$f \in \text{dom}(h(l(x))) \quad (16)$$

$$\Gamma[e] \preceq \Gamma[f] \quad (17)$$

and

$$\text{if } x = \mathbf{this} \text{ then } T^\sharp(m, i)[f \mapsto \text{Def}] \text{ else } T^\sharp(m, i) \sqsubseteq T^\sharp(m, i + 1) \quad (18)$$

$$L^\sharp(m, i) \sqsubseteq L^\sharp(m, i + 1) \quad (19)$$

$$\llbracket e \rrbracket^\sharp \sqsubseteq H^\sharp(f) \quad (20)$$

We must show that  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m,o} \langle \mathbf{i} + 1, \mathbf{l}, \text{upd}(h, \mathbf{l}(\mathbf{x}), \mathbf{f}, (\mathbf{v}, \mathbf{def})) \rangle$  holds. That is

$$o = l(\mathbf{this}), \quad L^\sharp(m, i + 1) \sim_{\text{upd}(h, \mathbf{l}(\mathbf{x}), \mathbf{f}, (\mathbf{v}, \mathbf{def}))} l, \\ T^\sharp(m, i + 1) \sim_{\text{upd}(h, \mathbf{l}(\mathbf{x}), \mathbf{f}, (\mathbf{v}, \mathbf{def})), \text{class}(m)} l(\mathbf{this}) \text{ and } H^\sharp \sim_{\text{upd}(h, \mathbf{l}(\mathbf{x}), \mathbf{f}, (\mathbf{v}, \mathbf{def}))}$$

$o = l(\mathbf{this})$  holds by (8).

$L^\sharp(m, i + 1) \sim_{\text{upd}(h, \mathbf{l}(\mathbf{x}), \mathbf{f}, (\mathbf{v}, \mathbf{def}))} l$  holds by monotony of  $\sim$  and by lemma 11 and hypothesis (19).

For  $T^\sharp(m, i + 1) \sim_{\text{upd}(h, \mathbf{l}(\mathbf{x}), \mathbf{f}, (\mathbf{v}, \mathbf{def})), \text{class}(m)} l(\mathbf{this})$ , by monotony of  $\sim$  and hypothesis (18), we only need to prove

$$(\text{if } x = \mathbf{this} \text{ then } T^\sharp(m, i)[f \mapsto \text{Def}] \text{ else } T^\sharp(m, i)) \sim_{\text{upd}(h, \mathbf{l}(\mathbf{x}), \mathbf{f}, (\mathbf{v}, \mathbf{def})), \text{class}(m)} l(\mathbf{this})$$

If  $x \neq \mathbf{this}$  the property holds thanks to (6) and lemma 12 else we conclude by lemma 14.

Finally, for  $H^\sharp \sim_{\text{upd}(h, \mathbf{l}(\mathbf{x}), \mathbf{f}, (\mathbf{v}, \mathbf{def}))}$  we first apply lemma 13 and so it is sufficient to prove  $H^\sharp(f) \sim_h v$  to conclude. This holds thanks to lemma 3 and 9, and hypotheses (20) and (17) and monotony of  $\sim_h$ .

**Case 2.2:  $\mathbf{P}_m[i] = \mathbf{x}.f \leftarrow \mathbf{e}$  and  $\mathbf{X} = \Omega$ .**

We must show that this case is impossible. By hypothesis  $l(x) = \mathbf{null}$  or  $e \Downarrow_{h,l} \Omega$ . Both case are contradictory with the result of lemma 4, using hypotheses (4), (5), (6) and (7).

**Case 3.1:  $\mathbf{P}_m[i] = \mathbf{x} \leftarrow \mathbf{new}(\mathbf{C}, \alpha)(\mathbf{e}_1, \dots, \mathbf{e}_n)$  and  $\mathbf{X} = \langle \mathbf{i} + 1, \mathbf{l}[\mathbf{x} \mapsto \mathbf{r}], \mathbf{h}' \rangle$ .**

We immediately see that  $X \neq \Omega$ .

Furthermore, by hypothesis,

$$\forall i, e_i \Downarrow_{h,l} v_i \quad (21)$$

$$r \notin \text{dom}(h) \quad (22)$$

$$\langle 0, \perp[\mathbf{this} \mapsto r, \{i \mapsto v_i\}_{i=1..n}], h[r \mapsto \text{default}(C)] \rangle \rightarrow_{\{\{C, \mathbf{init}, \alpha \rightarrow \mathbf{void}\}\}^{(n-1)*}} h' \quad (23)$$

$$x \neq \mathbf{this} \quad (24)$$

$$C \preceq \Gamma[x] \quad (25)$$

$$\forall i = 1..n, \Gamma[e_i] \preceq \Gamma[i] \quad (26)$$

and

$$m' = \{\{C, \mathbf{init}, \alpha \rightarrow \mathbf{void}\}\} \quad (27)$$

$$\top_C \sqsubseteq M^\sharp(m')[\mathbf{this}] \quad (28)$$

$$\text{Raw}^- :: \llbracket e_1 \rrbracket^\sharp :: \dots :: \llbracket e_j \rrbracket^\sharp \sqsubseteq M^\sharp(m')[\text{args}] \quad (29)$$

$$L^\sharp(m, i)[x \mapsto \text{NotNull}] \sqsubseteq L^\sharp(m, i+1) \quad (30)$$

$$T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1) \quad (31)$$

$$(32)$$

We must show that  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m,o} \langle i+1, l[x \mapsto r], h' \rangle$  holds. That is

$$o = l[x \mapsto r](\mathbf{this}), \quad L^\sharp(m, i+1) \sim_{h'} l[x \mapsto r], \\ T^\sharp(m, i+1) \sim_{h', \text{class}(m)} l[x \mapsto r](\mathbf{this}) \text{ and } H^\sharp \sim h'$$

$o = l[x \mapsto r](\mathbf{this})$  holds by (24) and (8).

$L^\sharp(m, i+1) \sim_{h'} l[x \mapsto r]$  is reduced, by monotony of  $\sim$  and (30) to  $L^\sharp(m, i)[x \mapsto \text{NotNull}] \sim_{h'} l[x \mapsto r]$ . Given a variable  $y$  we hence prove that  $L^\sharp(m, i)(y)[x \mapsto \text{NotNull}] \sim_{h', \Gamma[y]} l[x \mapsto r](y)$ . If  $x = y$ , we have to prove  $\text{NotNull} \sim_{h', \Gamma[x]} r$ . By hypothesis (23) the execution must have ended with an instruction **return** and so all fields of  $h'(r)$  in  $\text{fields}(C)$  (and also those declared in the  $C$ 's ancestor) are now set to **def**.  $h'(r)$  is hence fully initialized since it is of class  $C$ . Together with the remark that  $\text{class}(\text{default}(C)) = C \preceq \Gamma[x]$  (by (25)), this proves  $\text{NotNull} \sim_{h', \Gamma[x]} r$ . If  $x \neq y$ , we have to prove  $L^\sharp(m, i)(y) \sim_{h', \Gamma[y]} l(y)$  and this comes from lemma 19 and hypothesis (5).

$T^\sharp(m, i+1) \sim_{h', \text{class}(m)} l[x \mapsto r](\mathbf{this})$  reduces to  $T^\sharp(m, i+1) \sim_{h', \text{class}(m)} l(\mathbf{this})$  because  $x \neq \mathbf{this}$ . We then conclude by lemma 20, monotony of  $\sim$  and hypothesis (31).

At last, for  $H^\sharp \sim h'$  we invoke (1) on the execution of  $\{\{C, \mathbf{init}, \alpha \rightarrow \mathbf{void}\}\}$  to obtain  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{\{\{C, \mathbf{init}, \alpha \rightarrow \mathbf{void}\}, l(\mathbf{this})\}} h'$  and hence conclude. In order to justify the use of (1) it remains to prove

$$(\perp[\mathbf{this} \mapsto r, \{i \mapsto v_i\}_{i=1..n}], h[r \mapsto \text{default}(C)]) \in \text{Pre}(M^\sharp, H^\sharp)(m')$$

To establish this, it is sufficient to prove the followings, thanks to monotony of  $\sim$  and hypothesis (29):

$$H^\sharp \sim h[r \mapsto \text{default}(C)], \quad \text{Raw}^- \sim_{h[r \mapsto \text{default}(C)], \Gamma[\mathbf{this}]} r, \\ M^\sharp(m')[\mathbf{this}] \sim_{C, h[r \mapsto \text{default}(C)]} r \quad \text{and } \forall i = 1..j, \llbracket e_i \rrbracket^\sharp \sim_{h[r \mapsto \text{default}(C)], \Gamma[i]} v_i$$

$H^\sharp \sim h[r \mapsto \text{default}(C)]$  holds by lemma 18 and  $Raw^- \sim_{h[r \mapsto \text{default}(C)], \Gamma[\mathbf{this}]} r$  holds since  $r \neq \mathbf{null}$ .  $M^\sharp(m')[\mathbf{this}] \sim_{C, h[r \mapsto \text{default}(C)]} r$  is proved by hypothesis (28). By lemmas 3, 15, 9 and hypothesis (26) we at last prove  $\forall p = 1..j, \llbracket e_p \rrbracket^\sharp \sim_{h[r \mapsto \text{default}(C)], \Gamma[i]} v_i$  and conclude.

**Case 3.2:**  $\mathbf{P}_m[\mathbf{i}] = \mathbf{x} \leftarrow \mathbf{new}(\mathbf{C}, \alpha)(\mathbf{e}_1, \dots, \mathbf{e}_n)$  and  $\mathbf{X} = \Omega$ .

Thanks to lemma 4 we know that none of the expression  $e_i$  may lead to an error. The only remaining case is when the execution of the constructors ends with an error, but this contradictory with hypothesis (1) (which can be invoked thanks to the same arguments than in Case 3.1).

**Case 4:**  $\mathbf{P}_m[\mathbf{i}] = \mathbf{return}$  and  $\mathbf{X} = \mathbf{h}'$ .

We immediately see that  $X \neq \Omega$ .

Furthermore, by hypothesis,

$$\text{name}(m) = \mathbf{init} \Rightarrow \forall A, \text{class}(m) \prec A \Rightarrow A \in \text{history}(l(\mathbf{this})) \quad (33)$$

if  $\text{name}(m) = \mathbf{init}$

$$\text{then } h' = h[l(\mathbf{this}) \mapsto \text{addHistory}_{\text{class}(m)}(\text{set2Def}_{\text{class}(m)}(h(l(\mathbf{this}))))] \quad (34)$$

else  $h' = h$

and

$$T^\sharp(m, i) \sqsubseteq M^\sharp(m)[\mathbf{post}] \quad (35)$$

$$m = \{C, \mathbf{init}, \alpha \rightarrow \mathbf{void}\} \Rightarrow \forall f \in \text{fields}(C). (T^\sharp(m, i)(f) = \text{UnDef} \Rightarrow \text{MayBeNull} \sqsubseteq H^\sharp(f)) \quad (36)$$

We must show that  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m, o} h'$  holds. That is

$$M^\sharp(m)[\mathbf{post}] \sim_{h', \text{class}(m)} o \text{ and } H^\sharp \sim h'$$

By (35), (8) and monotony of  $\sim$  we reduces  $M^\sharp(m)[\mathbf{post}] \sim_{h', \text{class}(m)} o$  to  $T^\sharp(m, i) \sim_{h', \text{class}(m)} l(\mathbf{this})$ . If  $h = h'$  we conclude directly with (6) else we add the remark that setting some fields to **def** does not break (6).

At last, for  $H^\sharp \sim h'$  we only have something to prove when  $m$  is a constructor  $\{C, \mathbf{init}, \alpha \rightarrow \mathbf{void}\}$  and all fields in  $h(o)$  have been set to **def**. For all reference  $r$  distinct from  $o$  we use (7) to conclude. Otherwise we must show that for any field  $f \in \text{dom}(h(o))$ ,  $H^\sharp(f) \sim_{h, \Gamma[f]} h'(o)(f)$ . There are two cases to consider in order to conclude:

1.  $f$  was already initialized before, then we conclude by  $H^\sharp \sim h$ .
2.  $f$  was **undef** in  $h$  and is now **def** in  $h'$ . This mean that  $f \in \text{fields}(C)$  and then by (36)  $\text{MayBeNull} \sqsubseteq H^\sharp(f)$  so the property holds.

**Case 5.1:**  $\mathbf{P}_m[\mathbf{i}] = \mathbf{x.ms}(\mathbf{e}_1, \dots, \mathbf{e}_n)$  and  $\mathbf{X} = \langle \mathbf{i} + 1, \mathbf{l}, \mathbf{h}' \rangle$ .

We immediately see that  $X \neq \Omega$ .

Furthermore, by hypothesis,

$$\forall i, e_i \Downarrow_{h,l} v_i \quad (37)$$

$$l(x) \in \text{dom}(h) \quad (38)$$

$$m'' = \text{lookup}(\text{class}(h(l(x))), ms) \quad (39)$$

$$\langle 0, \perp[\mathbf{this} \mapsto l(x), \{i \mapsto v_i\}_{i=1..n}], h \rangle \rightarrow_{m''}^* h' \quad (40)$$

$$\text{name}(ms) = \mathbf{init} \Rightarrow x = \mathbf{this} \quad (41)$$

and

$$ms = \{C, m, \alpha \rightarrow \mathbf{void}\} \quad (42)$$

$$m' = \text{lookup}(C, m, \alpha) \quad (43)$$

$$C' = \text{class}(m') \quad (44)$$

$$m = \mathbf{init} \Rightarrow L^\sharp(m, i)[x \mapsto \text{Raw}(C') \sqcap L^\sharp(m, i)(x)] \sqsubseteq L^\sharp(m, i+1) \quad (45)$$

$$m \neq \mathbf{init} \Rightarrow L^\sharp(m, i) \sqsubseteq L^\sharp(m, i+1)$$

$$x = \mathbf{this} \Rightarrow M^\sharp(m')[\mathbf{post}] \sqcap T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1) \quad (46)$$

$$x \neq \mathbf{this} \Rightarrow T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1)$$

$$\left( \begin{array}{l} \text{if } x = \mathbf{this} \wedge \text{class}(m) = C' \text{ then } T^\sharp(m, i) \sqsubseteq M^\sharp(m')[\mathbf{this}] \\ \text{else } \rho(C', L^\sharp(m, i)(x)) \sqsubseteq M^\sharp(m')[\mathbf{this}] \end{array} \right) \quad (47)$$

$$L^\sharp(m, i)(x) :: \llbracket e_1 \rrbracket^\sharp :: \dots :: \llbracket e_j \rrbracket^\sharp \sqsubseteq M^\sharp(m')[\mathbf{args}] \quad (48)$$

We must show that  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) \sim_{m,o} \langle i+1, l, h' \rangle$  holds. That is

$$\begin{aligned} o &= l(\mathbf{this}), \quad L^\sharp(m, i+1) \sim_{h'} l, \\ T^\sharp(m, i+1) &\sim_{h', \text{class}(m)} l(\mathbf{this}) \text{ and } H^\sharp \sim h' \end{aligned}$$

$o = l(\mathbf{this})$  holds by (8).

For  $L^\sharp(m, i+1) \sim_{h'} l$  we make two cases:

- if  $m = \mathbf{init}$ , by monotony and (45) we show  $L^\sharp(m, i)[x \mapsto \text{Raw}(C') \sqcap L^\sharp(m, i)(x)] \sim_{h'} l$ . Since  $m = \mathbf{init}$ , we have  $x = \mathbf{this}$  by (41). For variable distinct from  $x$  the proof is similar to Case 3.1. Otherwise we show  $\text{Raw}(C') \sqcap L^\sharp(m, i)(x) \sim_{h', \Gamma[x]} l(x)$  which holds because the called constructor has initialized all fields declared in  $C'$  (and its ancestors).
- if  $m \neq \mathbf{init}$ , the proof is similar to Case 3.1.

For  $T^\sharp(m, i+1) \sim_{h', \text{class}(m)} l(\mathbf{this})$  we make two cases:

- $x = \mathbf{this}$  and  $\text{class}(m) = \text{class}(m')$ : by monotony we have to prove  $M^\sharp(m')[\mathbf{post}] \sqcap T^\sharp(m, i) \sim_{h', \text{class}(m)} l(\mathbf{this})$ . By lemma 20, we have  $T^\sharp(m, i) \sim_{h', \text{class}(m)} l(\mathbf{this})$ . By correctness of the execution of the called method,  $M^\sharp(m')[\mathbf{post}] \sim_{h', \text{class}(m'')} l(\mathbf{this})$ . If  $m' = m''$  we can conclude since  $\text{class}(m) = \text{class}(m')$ . Otherwise  $\text{overrides}(m'', m')$  and thus we can affirm that  $\top_C \sqsubseteq M^\sharp(m')[\mathbf{post}]$ . Hence  $M^\sharp(m')[\mathbf{post}] \sqcap T^\sharp(m, i) = T^\sharp(m, i)$  and we can conclude.

- $x \neq \mathbf{this}$ : by monotony we have to prove  $T^\sharp(m, i) \sim_{h', \text{class}(m)}$  which is justified by lemma 20.

For  $H^\sharp \sim h'$  we invoke the hypothesis (1) in a similar way than for case 3.1.

It finally remains to justify the use of hypothesis (1) we have made several times in this part of the proof. That is:

$$(\perp[\mathbf{this} \mapsto l(x), \{i \mapsto v_i\}_{i=1..n}], h) \in \text{Pre}(M^\sharp, H^\sharp)(m'')$$

First we remark that, since  $m' = m''$  or  $\text{overrides}(m'', m')$ ,

$$M^\sharp(m')[\text{args}] \sqsubseteq M^\sharp(m'')[\text{args}] \quad (49)$$

To establish this, it is sufficient to prove the followings, thanks to monotony of  $\sim$  and hypothesis (48) and (49):

$$H^\sharp \sim h, L^\sharp(m, i)(x) \sim_h l(x), \\ M^\sharp(m'')[\mathbf{this}] \sim_{\text{class}(m''), h} l(x) \text{ and } \forall p = 1..j, \llbracket e_p \rrbracket^\sharp \sim_h v_i$$

$H^\sharp \sim h$  holds by hypothesis (7) and  $L^\sharp(m, i)(x) \sim_h l(x)$  by (5). By lemmas 3 and 15 we prove  $\forall p = 1..j, \llbracket e_p \rrbracket^\sharp \sim_h v_i$ .

To finish we have to prove

$$M^\sharp(m'')[\mathbf{this}] \sim_{\text{class}(m''), h} l(x)$$

If  $m' \neq m''$  then  $\text{overrides}(m'', m')$  holds, by consequence  $\top_{\text{class}(m'')} \sqsubseteq M^\sharp(m'')[\mathbf{this}]$  and the property holds. Otherwise we must prove

$$M^\sharp(m')[\mathbf{this}] \sim_{C', h} l(x)$$

We distinguish two cases

1.  $x = \mathbf{this} \wedge \text{class}(m) = C'$ : by hypothesis (47),  $T^\sharp(m, i) \sqsubseteq M^\sharp(m')[\mathbf{this}]$  and thanks to (6) we have  $T^\sharp(m, i) \sim_{h, \text{class}(m)} l(\mathbf{this})$  and we can conclude by monotony.
2. otherwise: hypothesis (47) gives  $\rho(C', L^\sharp(m, i)(x)) \sqsubseteq M^\sharp(m')[\mathbf{this}]$  and it remains to prove:

$$\rho(C', L^\sharp(m, i)(x)) \sim_{C', h} l(x)$$

which holds by definition of  $\rho$  and hypothesis (5).

**Case 5.2:  $\mathbf{P}_m[\mathbf{i}] = \mathbf{x.m}(e_1, \dots, e_n)$  and  $\mathbf{X} = \Omega$ .**

This case is similar to case 3.2.

**Case 6:  $\mathbf{P}_m[\mathbf{i}] = \mathbf{if\ jmp}$**  Trivial case.

## D Comparison to Fähndrich and Leino's type system

*Program typability*

$$\begin{array}{c}
\Gamma[m] = [\tau]\beta \rightarrow \text{void} \quad \tau \neq \text{MaybeNull} \quad L' = \perp[\{x_i \mapsto \beta_i\}_{i=1..|\beta|}] \\
\tau = \text{inits } F \Rightarrow F \subseteq \text{fields}(\text{class}(m)) \\
\tau \in \text{Val}^\sharp \Rightarrow L'[\text{this} \mapsto \tau] \sqsubseteq L(m, 0) \\
\tau = \text{inits } F \Rightarrow L'[\text{this} \mapsto \text{Raw}(\text{super}(\text{class}(m)))] \sqsubseteq L(m, 0) \\
\forall i \left( \begin{array}{l} P_m[i] = \text{return } \vee \\ \forall j. (j \in \text{succ}(i) \wedge \exists L' \sqsubseteq L(m, j) \wedge \Gamma, m \vdash P_m[i] : L(m, i) \rightarrow L') \end{array} \right) \\
\hline
\Gamma, L, m \vdash P \\
\\
\forall m, m', \Gamma[m] = [\tau]\beta \rightarrow \text{void} \wedge \Gamma[m'] = [\tau']\beta' \rightarrow \text{void} \wedge \text{overrides}(m, m') \\
\Rightarrow \tau' \in \text{Val}^\sharp \wedge \tau' \sqsubseteq \tau \wedge \forall i, \beta'_i \sqsubseteq \beta_i \\
\forall m, \Gamma, L, m \vdash P \qquad \Gamma, F \models P \\
\hline
\Gamma, L, F \vdash P
\end{array}$$

*Conversion to our analysis.* We define a function  $\text{Conv} : (\Gamma, L, F) \mapsto (M^\sharp, H^\sharp, T^\sharp, L^\sharp)$  which inject an annotation of Fähndrich and Leino's type system to an abstract value for our constraint system such that:

$$\forall m, M^\sharp(m)[\text{this}] = \top_{\text{class}(m)} \quad (50)$$

$$\forall m, M^\sharp(m)[\text{args}] = \text{Raw}(\text{super}(\text{class}(m))) :: \beta \quad (51)$$

$$\begin{aligned} & \text{if } \Gamma[m] = [\text{inits } F]\beta \rightarrow \text{void} \\ & = \tau :: \beta \quad \text{otherwise, where } \Gamma[m] = [\tau]\beta \rightarrow \text{void} \end{aligned} \quad (52)$$

$$\begin{aligned} \forall m, M^\sharp(m)[\text{post}] &= \top_{\text{class}(m)} \quad \text{if } \Gamma[m] = [\tau]\beta \rightarrow \text{void} \wedge \tau \in \text{Val}^\sharp \\ &= \top_{\text{class}(m)}[\bigcup_{f \in F} \{f \mapsto \text{Def}\}] \end{aligned} \quad (53)$$

$$\begin{aligned} & \text{if } \Gamma[m] = [\text{inits } F]\beta \rightarrow \text{void} \\ \forall f, H^\sharp(f) &= \Gamma[f] \end{aligned} \quad (54)$$

$$\forall m, i, L^\sharp(m, i) = L(m, i) \quad (55)$$

$$\forall m, i, T^\sharp(m, i) = \perp_{\text{class}(m)}[\bigcup_{f \in F(m, i)} \{f \mapsto \text{UnDef}\}] \quad (56)$$

$$\begin{aligned} & \text{if } (\Gamma[m] = [\text{inits } F]\beta \rightarrow \text{void} \vee \text{name}(m) = \text{init}) \\ & = \top_{\text{class}(m)} \quad \text{otherwise} \end{aligned} \quad (57)$$

We then propose to prove that a such value is a safe solution of our constraint system.

**Lemma 22.** *Let  $P$  be a well-typed program w.r.t  $\Gamma, L$  and  $F$  such that  $\Gamma, L, F \vdash P$ , and  $e$  be an expression reachable at program point  $(m, i)$  of  $P$ . Let  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) = \text{Conv}(\Gamma, L, F)$ . Then*

$$\Gamma, L(m, i) \vdash e : \tau \Rightarrow \forall C \in \mathbb{C}, \llbracket e \rrbracket_{C, H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}^\sharp \sqsubseteq \tau$$

*Proof.* We denote in this proof  $L^\sharp(m, i)$  by  $L^\sharp$  and  $T^\sharp(m, i)$  by  $T^\sharp$ . We propose an induction on  $e$ .

- If  $e = x$  then  $\Gamma, L \vdash e : L(x)$ . We distinguish two cases.
  - If  $L^\sharp(x) = \text{Raw}(\text{super}(C))$ , as  $\text{Raw}(C) \sqsubseteq \text{Raw}(\text{super}(C))$ ,  $\text{Raw}(C) \sqsubseteq L(x)$  holds
  - otherwise  $L^\sharp(x) \sqsubseteq L(x)$  holds by definition of  $L^\sharp$  (55).
- If  $e = e'.f$ , by hypothesis, we have  $\Gamma, L \vdash e' : \tau' \Rightarrow \llbracket e' \rrbracket_{C, H^\sharp, T^\sharp, L^\sharp}^\sharp \sqsubseteq \tau'$ . We distinguish four cases.
  - If  $\Gamma, L \vdash e' : \text{Raw}^-$ , then  $\Gamma, L \vdash e'.f : \text{MaybeNull}$  and  $\llbracket e'.f \rrbracket_{C, H^\sharp, T^\sharp, L^\sharp}^\sharp \sqsubseteq \text{MaybeNull}$  holds.
  - If  $\Gamma, L \vdash e' : \text{Raw}(C)$  and  $\text{class}(f) \prec C$ , then  $\Gamma, L \vdash e'.f : \text{MaybeNull}$  and  $\llbracket e'.f \rrbracket_{C, H^\sharp, T^\sharp, L^\sharp}^\sharp \sqsubseteq \text{MaybeNull}$  holds.
  - If  $\Gamma, L \vdash e' : \text{Raw}(C)$  and  $C \preceq \text{class}(f)$ , then  $\Gamma, L \vdash e'.f : \Gamma[f]$  and  $\llbracket e' \rrbracket_{C, H^\sharp, T^\sharp, L^\sharp}^\sharp \sqsubseteq \text{Raw}(C)$ . The analysis gives us  $\llbracket e'.f \rrbracket_{C, H^\sharp, T^\sharp, L^\sharp}^\sharp = H^\sharp(f)$ , so, from (54),  $\llbracket e'.f \rrbracket_{C, H^\sharp, T^\sharp, L^\sharp}^\sharp \sqsubseteq \Gamma[f]$  holds.
  - If  $\Gamma, L \vdash e' : \text{NotNull}$ , then  $\Gamma, L \vdash e'.f : \Gamma[f]$  and  $\llbracket e' \rrbracket_{C, H^\sharp, T^\sharp, L^\sharp}^\sharp \sqsubseteq \text{NotNull}$ , so we can deduce  $\llbracket e'.f \rrbracket_{C, H^\sharp, T^\sharp, L^\sharp}^\sharp = H^\sharp(f)$ . With (54) we can conclude that  $\llbracket e'.f \rrbracket_{C, H^\sharp, T^\sharp, L^\sharp}^\sharp \sqsubseteq \Gamma[f]$  holds. □

**Lemma 23.**

$$L(m, i) \sqsubseteq L(m, j) \Rightarrow L^\sharp(m, i) \sqsubseteq L^\sharp(m, j)$$

*Proof.* Trivial from (55). □

**Lemma 24.**

$$F(m, i) \sqsubseteq F(m, j) \Rightarrow T^\sharp(m, i) \sqsubseteq T^\sharp(m, j)$$

*Proof.* Trivial from (56). □

**Lemma 25.**

$$\Gamma, L, F \vdash P \Rightarrow \forall m, i, \text{Conv}(\Gamma, L, F) \models (m, i) : P_m[i]$$

*Proof.* We do a case analysis on  $P_m[i]$ .

**Case 1:**  $x \leftarrow e$ . We have to prove that:

$$\begin{aligned} & T^\sharp(m, i) \sqsubseteq T^\sharp(m, i + 1) \\ & L^\sharp(m, i)[x \mapsto e \llbracket e \rrbracket_{\text{class}(m), H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}^\sharp}] \sqsubseteq L^\sharp(m, i + 1) \end{aligned}$$

We first prove that  $T^\sharp(m, i) \sqsubseteq T^\sharp(m, i + 1)$ .

**Case 1.1:**  $\Gamma[m] = [\text{inits} \dots]\beta \rightarrow \text{void}$  or  $m = \{C, \text{init}, \alpha \rightarrow \text{void}\}$ . By hypothesis,  $F(m, i) \sqsubseteq F(m, i + 1)$ , which, with Lemma 24, implies  $T^\sharp(m, i) \sqsubseteq T^\sharp(m, i + 1)$ .

**Case 1.2: otherwise.** From (57),  $T^\sharp(m, i + 1) = \top_{\text{class}(m)}$ , so  $T^\sharp(m, i) \sqsubseteq T^\sharp(m, i + 1)$  holds.

We now prove that  $L^\sharp(m, i)[x \mapsto e \llbracket e \rrbracket_{\text{class}(m), H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}^\sharp] \sqsubseteq L^\sharp(m, i + 1)$ . By hypothesis,  $\Gamma, m \vdash x \leftarrow e : L(m, i) \rightarrow L(m, i)[x \mapsto \tau]$  and  $L(m, i)[x \mapsto \tau] \sqsubseteq L(m, i + 1)$  so, from Lemma 22, we have  $\llbracket e \rrbracket_{\text{class}(m), H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}^\sharp \sqsubseteq \tau$ . We can conclude with (55).

**Case 2:**  $x.f \leftarrow e$ . The value of  $T^\sharp$  depends on whether  $x$  equals to **this**.

**Case 2.1:**  $x = \mathbf{this}$ . By hypothesis,  $F(m, i) \setminus \{f\} \subseteq F(m, i + 1)$  so, from (56),  $T^\sharp(m, i)[f \mapsto \text{Def}] \sqsubseteq T^\sharp(m, i + 1)$  holds.

**Case 2.2:**  $x \neq \mathbf{this}$ . By hypothesis,  $F(m, i) \subseteq F(m, i + 1)$  so, from Lemma 24,  $T^\sharp(m, i) \sqsubseteq T^\sharp(m, i + 1)$  holds.

By hypothesis,

$$L(m, i) \sqsubseteq L(m, i + 1) \quad (58)$$

$$\Gamma, L \vdash e : \tau \quad (59)$$

$$\tau \sqsubseteq \Gamma[f] \quad (60)$$

From Lemma 23 and (58)  $L^\sharp(m, i) \sqsubseteq L^\sharp(m, i + 1)$  holds.

From Lemma 22 and (59),  $\llbracket e \rrbracket_{\text{class}(m), H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}^\sharp \sqsubseteq \tau$ . So with (60) and (54),  $\llbracket e \rrbracket_{\text{class}(m), H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}^\sharp \sqsubseteq H^\sharp(f)$  holds.

**Case 3:**  $x \leftarrow \mathbf{new} (C, \alpha)(e_1, \dots, e_j)$ . Let  $m'$  be  $\{C, \mathbf{init}, \alpha \rightarrow \mathbf{void}\}$ . We have to prove:

$$\begin{aligned} \top_{\text{class}(m')} &\sqsubseteq M^\sharp(m')[\mathbf{this}] \\ \text{Raw}^- :: \llbracket e_1 \rrbracket^\sharp :: \dots :: \llbracket e_j \rrbracket^\sharp &\sqsubseteq M^\sharp(m')[\mathbf{args}] \\ L^\sharp(m, i)[x \mapsto \text{NotNull}] &\sqsubseteq L^\sharp(m, i + 1) \\ T^\sharp(m, i) &\sqsubseteq T^\sharp(m, i + 1) \end{aligned}$$

By hypothesis,

$$\Gamma[m'] = [\text{Raw}^-]\beta \rightarrow \mathbf{void} \quad (61)$$

$$L(m, i)[x \mapsto \text{NotNull}] \sqsubseteq L(m, i + 1) \quad (62)$$

$$\forall i, \Gamma, L \vdash e_i : \beta_i \quad (63)$$

$$F(m, i) \subseteq F(m, i + 1) \quad (64)$$

By (50),  $\top_{\text{class}(m')} \sqsubseteq M^\sharp(m')[\mathbf{this}]$  holds.

From Lemma 22 and (63), we have  $\forall i, \llbracket e_i \rrbracket^\sharp \sqsubseteq \beta_i$ . With (51) and (61),  $\text{Raw}^- :: \llbracket e_1 \rrbracket^\sharp :: \dots :: \llbracket e_j \rrbracket^\sharp \sqsubseteq M^\sharp(m')[\mathbf{args}]$  holds. Proofs of  $L^\sharp(m, i)[x \mapsto \text{NotNull}] \sqsubseteq L^\sharp(m, i + 1)$  and  $T^\sharp(m, i) \sqsubseteq T^\sharp(m, i + 1)$  are trivial from definitions (55) and (56) and equations (62), (64).

**Case 4: if `jmp`.** We have to prove

$$\begin{aligned} L^\sharp(m, i) &\sqsubseteq L^\sharp(m, i + 1) \\ L^\sharp(m, i) &\sqsubseteq L^\sharp(m, \text{jmp}) \\ T^\sharp(m, i) &\sqsubseteq T^\sharp(m, i + 1) \\ T^\sharp(m, i) &\sqsubseteq T^\sharp(m, \text{jmp}) \end{aligned}$$

By hypothesis,

$$F(m, i) \subseteq F(m, i + 1) \quad (65)$$

$$F(m, i) \subseteq F(m, \text{jmp}) \quad (66)$$

$$L(m, i) \sqsubseteq L(m, i + 1) \quad (67)$$

$$L(m, i) \sqsubseteq L(m, \text{jmp}) \quad (68)$$

$$(69)$$

Trivial with lemmas 23 and 24.

**Case 5: return** We have to prove

$$T^\sharp(m, i) \sqsubseteq M^\sharp(m)[\text{post}]$$

$$\text{name}(m) = \mathbf{init} \Rightarrow$$

$$\forall f \in \text{fields}(C). (T^\sharp(m, i)(f) = \text{UnDef} \Rightarrow \text{MaybeNull} \sqsubseteq H^\sharp(f))$$

By hypothesis,

$$\text{name}(m) = \mathbf{init} \Rightarrow$$

$$\forall f \in F(m, i). (\text{MaybeNull} \sqsubseteq \Gamma[f]) \quad (70)$$

$$\Gamma[m] = [\mathbf{inits } F]\alpha \rightarrow \mathbf{void} \Rightarrow$$

$$F \cap F(m, i) = \emptyset \quad (71)$$

$$\Gamma[m] \neq [\mathbf{inits } F]\alpha \rightarrow \mathbf{void} \Rightarrow$$

$$M^\sharp(m)[\text{post}] = \top_{\text{class}(m)} \quad (72)$$

$$\Gamma[m] = [\mathbf{inits } F]\alpha \rightarrow \mathbf{void} \Rightarrow$$

$$M^\sharp(m)[\text{post}] = \top_{\text{class}(m)} \left[ \bigcup_{f \in F} \{f \mapsto \text{Def}\} \right] \quad (73)$$

We first prove the first equation:  $T^\sharp(m, i) \sqsubseteq M^\sharp(m)[\text{post}]$ .

**Case 5.1:**  $\Gamma[m] \neq [\mathbf{inits } F]\alpha \rightarrow \mathbf{void}$ . From (72)  $T^\sharp(m, i) \sqsubseteq M^\sharp(m)[\text{post}]$  trivially holds.

**Case 5.2:**  $\Gamma[m] = [\mathbf{inits } F]\alpha \rightarrow \mathbf{void}$ . Equation (71) implies  $\text{fields}(\text{class}(m)) \cap F \subseteq \text{fields}(\text{class}(m)) \setminus F(m, i)$ , which implies, with (56) and (73), that  $T^\sharp(m, i) \sqsubseteq M^\sharp(m)[\text{post}]$ .

From (56), (54) and (70), we easily prove that, if  $\text{name}(m) = \mathbf{init}$ , then  $\forall f \in \text{fields}(C). (T^\sharp(m, i)(f) = \text{UnDef} \Rightarrow \text{MaybeNull} \sqsubseteq H^\sharp(f))$ .

**Case 6:**  $x.ms(e_1, \dots, e_j)$ . Let  $m' = \text{lookup}(ms)$  and  $C' = \text{class}(m')$ . We have to prove:

$$\left( \begin{array}{l} \text{if } (x = \mathbf{this} \wedge \text{class}(m) = C' \wedge M^\sharp(m')[\text{post}] \sqcap T^\sharp(m, i) \sqsubseteq \perp_{C'}) \vee \text{name}(m') = \mathbf{init} \\ \text{then } L^\sharp(m, i)[x \mapsto \text{Raw}(C') \sqcap \llbracket x \rrbracket^\sharp] \sqsubseteq L^\sharp(m, i+1) \\ \text{else } L^\sharp(m, i) \sqsubseteq L^\sharp(m, i+1) \end{array} \right) \\ \left( \begin{array}{l} \text{if } x = \mathbf{this} \wedge \text{class}(m) = C' \text{ then } M^\sharp(m')[\text{post}] \sqcap T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1) \\ \text{else } T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1) \end{array} \right) \\ \left( \begin{array}{l} \text{if } x = \mathbf{this} \wedge \text{class}(m) = C' \text{ then } T^\sharp(m, i) \sqsubseteq M^\sharp(m')[\text{this}] \\ \text{else } \rho(C', \llbracket x \rrbracket^\sharp) \sqsubseteq M^\sharp(m')[\text{this}] \\ \llbracket x \rrbracket^\sharp :: \llbracket e_1 \rrbracket^\sharp :: \dots :: \llbracket e_j \rrbracket^\sharp \sqsubseteq M^\sharp(m')[\text{args}] \end{array} \right)$$

Let  $\tau$  be such that  $\Gamma[m'] = [\tau]\beta \rightarrow \mathbf{void}$ .

**Case 6.1:**  $L^\sharp(m, i+1)$ . We consider three cases.

**Case 6.1.1: name(m') = init.** By hypothesis, we have  $L(m, i)[x \mapsto \tau \sqcap \text{Raw}(C')] \sqsubseteq L(m, i+1)$ . So with (55),  $\text{name}(m') = \mathbf{init} \Rightarrow L^\sharp(m, i)[x \mapsto \text{Raw}(C') \sqcap \llbracket x \rrbracket^\sharp] \sqsubseteq L^\sharp(m, i+1)$  holds.

**Case 6.1.2:  $\tau = \mathbf{inits} \ F \wedge x = \mathbf{this} \wedge \mathbf{class}(m) = C' \wedge \mathbf{fields}(C') \sqsubseteq F$ .** By hypothesis,  $\Gamma, L \vdash x : \text{Raw}(\text{super}(C'))$  so, by Lemma 22,  $\llbracket x \rrbracket^\sharp \sqsubseteq \text{Raw}(\text{super}(C'))$ , so  $x$  and  $\text{Raw}(C')$  are comparable, so  $\llbracket x \rrbracket^\sharp \sqcap \text{Raw}(C') \sqsubseteq \text{Raw}(C')$ , so  $L(m, i)[x \mapsto \llbracket x \rrbracket^\sharp \sqcap \text{Raw}(C')] \sqsubseteq L(m, i)[x \mapsto \text{Raw}(C')]$ . As, by hypothesis,  $L(m, i)[x \mapsto \tau] \sqsubseteq L(m, i+1)$ , from (55), we can deduce that  $L^\sharp(m, i)[x \mapsto \text{Raw}(C') \sqcap \llbracket x \rrbracket^\sharp] \sqsubseteq L^\sharp(m, i+1)$  holds.

**Case 6.1.3: otherwise.** By hypothesis we have  $\Gamma, L(m, i) \vdash x : \tau$  which implies that, with Lemma 22,  $\llbracket x \rrbracket^\sharp \sqsubseteq \tau$ . As  $L(m, i)[x \mapsto \tau] \sqsubseteq L(m, i+1)$ , by transitivity and with (55),  $L^\sharp(m, i) \sqsubseteq L^\sharp(m, i+1)$  holds.

**Case 6.2:**  $T^\sharp(m, i+1)$  and  $M^\sharp(m')[\text{this}]$ . We consider two cases.

**Case 6.2.1:  $\text{inits}_{m, m', x, F'}$ .** Let  $f_C = \text{fields}(C')$ . By hypothesis, we have  $F(m, i) \setminus F' \sqsubseteq F(m, i+1)$ , which implies  $(f_C \cap F(m, i)) \cap (f_C \setminus F') \sqsubseteq (f_C \cap F(m, i+1))$ . So

$$\perp_{C'}[\bigcup_{f \in f_C \cap F(m, i)} \{f \mapsto \text{UnDef}\}] \sqcap \perp_{C'}[\bigcup_{f \in f_C \setminus F'} \{f \mapsto \text{UnDef}\}] \sqsubseteq \perp_{C'}[\bigcup_{f \in f_C \cap F(m, i+1)} \{f \mapsto \text{UnDef}\}]$$

which, with (53) and (56), is equivalent to  $T^\sharp(m, i) \sqcap M^\sharp(m')[\text{post}] \sqsubseteq T^\sharp(m, i+1)$ .

Equation (50) trivially implies  $T^\sharp(m, i) \sqsubseteq M^\sharp(m')[\text{this}]$ .

**Case 6.2.2:  $\neg(\exists F', \text{inits}_{m, m', x, F'})$ .** By hypothesis we have  $F(m, i) \sqsubseteq F(m, i+1)$ , which implies  $T^\sharp(m, i) \sqsubseteq T^\sharp(m, i+1)$ .

Equation (50) trivially implies  $\rho(C', \llbracket x \rrbracket^\sharp) \sqsubseteq M^\sharp(m')[\text{this}]$ .

**Case 6.3:  $M^\sharp(m')[\text{args}]$ .** By hypothesis we have  $\forall i, \Gamma, L \vdash e_i : \beta_i$  which implies, with Lemma 22,  $\forall i, \llbracket e_i \rrbracket^\sharp \sqsubseteq \beta_i$ . If  $\tau \in \text{Val}^\sharp$ , by hypothesis and with Lemma 22, we also have  $\llbracket x \rrbracket^\sharp \sqsubseteq \tau$ , so with (51)  $\llbracket x \rrbracket^\sharp :: \llbracket e_1 \rrbracket^\sharp :: \dots :: \llbracket e_j \rrbracket^\sharp \sqsubseteq M^\sharp(m')[\text{args}]$  holds. If  $\tau \notin \text{Val}^\sharp$ , by hypothesis and with Lemma 22, we have  $\llbracket x \rrbracket^\sharp \sqsubseteq \text{Raw}(\text{super}(C'))$ , so with (51)  $\llbracket x \rrbracket^\sharp :: \llbracket e_1 \rrbracket^\sharp :: \dots :: \llbracket e_j \rrbracket^\sharp \sqsubseteq M^\sharp(m')[\text{args}]$  holds.  $\square$

**Lemma 26.**

$$\Gamma, L \vdash e : \tau \wedge (M^\sharp, H^\sharp, T^\sharp, L^\sharp) = \text{Conv}(\Gamma, L, F) \Rightarrow \forall C, \text{safe}_{\text{expr}}^\sharp(e)_{C, H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}$$

*Proof.* By induction on  $e$ .

- If  $e = x$ , then  $\text{safe}_{\text{expr}}^\sharp(e)_{C, H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}$  holds by definition.
- If  $e = e'.f$ , we suppose that  $\text{safe}_{\text{expr}}^\sharp(e')_{C, H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}$  holds. We distinguish two cases.
  - If  $e' = \text{MaybeNull}$ , then it contradicts  $\Gamma, L \vdash e : \tau$ . So  $e' \neq \text{MaybeNull}$ , so  $\text{safe}_{\text{expr}}^\sharp(e)_{C, H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}$  holds.

**Lemma 27.** *If  $P$  is LF-typable w.r.t.  $\Gamma$ , then there exists  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp)$  such that  $\text{safe}^\sharp(M^\sharp, H^\sharp, T^\sharp, L^\sharp)$  holds.*

*Proof.*  $P$  is LF-typable w.r.t.  $\Gamma$  means that there exists  $L$  and  $F$  such that  $\Gamma, L, F \vdash P$ . We choose  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) = \text{Conv}(\Gamma, L, F)$  and we must show that  $\text{safe}^\sharp(M^\sharp, H^\sharp, T^\sharp, L^\sharp)$  holds. We make a case analysis on  $P_m[i]$ .

- If  $P_m[i] = x \leftarrow e$ , we must show that  $\text{safe}_{\text{expr}}^\sharp(e)_{C, H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}$  holds. By hypothesis,  $\Gamma, L \vdash e : \tau$ , so from Lemma 26 the property holds.
- If  $P_m[i] = x.f \leftarrow e$ , we must show that  $L(m, i)(x) \neq \text{MaybeNull}$  and  $\text{safe}_{\text{expr}}^\sharp(e)_{C, H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}$  hold. By hypothesis,  $L(m, i)(x) \neq \text{MaybeNull}$  and  $\Gamma, L \vdash e : \tau$ , so, from (55) and Lemma 26, the property holds.
- If  $P_m[i] = x \leftarrow \mathbf{new} (C, \alpha)(e_1, \dots, e_j)$ , we must show that  $\forall j, \text{safe}_{\text{expr}}^\sharp(e_j)_{\text{class}(m), H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}$ . By hypothesis,  $(\forall i, \exists \beta_i, \Gamma, L \vdash e_i : \beta_i)$ , so, from Lemma 26, the property holds.
- If  $P_m[i] = x.mq(e_1, \dots, e_n)$ , we must show that  $L^\sharp(m, i)(x) \neq \text{MaybeNull}$  and *forall*  $j$ ,  $\text{safe}_{\text{expr}}^\sharp(e_j)_{\text{class}(m), H^\sharp, T^\sharp(m, i), L^\sharp(m, i)}$ . By hypothesis,  $(\forall i, \Gamma, L \vdash e_i : \beta_i)$  and if  $\Gamma[\text{lookup}(ms)] = [\tau]\beta \rightarrow \mathbf{void}$  then  $\Gamma, L \vdash x : \text{if } \tau \in \text{Val}^\sharp \text{ then } \tau \text{ else Raw}(\text{super}(C'))$  and  $\tau \neq \text{MaybeNull}$ . So, with Lemma 26, the property holds.

*QED.*

**Theorem 5.** *If  $P$  is LF-typable w.r.t.  $\Gamma, L, F$ , then there exists  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp)$  such that  $M^\sharp, H^\sharp, T^\sharp, L^\sharp \models P$  and  $\text{safe}^\sharp(M^\sharp, H^\sharp, T^\sharp, L^\sharp)$  holds.*

*Proof.* Let  $(M^\sharp, H^\sharp, T^\sharp, L^\sharp) = \text{Conv}(\Gamma, L, F)$ . For all  $m$  and  $m'$ , let  $\Gamma[m] = [\tau]\beta \rightarrow \text{void}$  and  $\Gamma[m'] = [\tau']\beta' \rightarrow \text{void}$ .

- By definition of a FL-typable program,  $\text{overrides}(m, m') \Rightarrow \beta' \sqsubseteq \beta$ , so with (51),  $M^\sharp(m')[\text{args}] \sqsubseteq M^\sharp(m)[\text{args}]$  holds.
- By definition of a FL-typable program,  $\text{overrides}(m, m') \Rightarrow \tau' \in \text{Val}^\sharp$ , so by (52)  $\top_{\text{class}(m)} \sqsubseteq M^\sharp(m)[\text{post}]$  holds.
- By (50),  $\top_{\text{class}(m')} \sqsubseteq M^\sharp(m')[\text{this}]$  holds.
- By (55), (51) and the definition of  $\Gamma, m \models P$ ,  $M^\sharp(m)[\text{args}] \sqsubseteq L^\sharp(m, 0)$  holds.
- We can conclude directly from Lemma 25 and Lemma 27.



---

Centre de recherche INRIA Rennes – Bretagne Atlantique  
IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex  
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier  
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq  
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex  
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex  
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex  
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

---

Éditeur  
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399