Normalization of Extended Abstract State Machines

Cegielski, Patrick
Université Paris Est – IUT, LACL EA 4219, Route forestière Hurtault, F-77300 Fontainebleau, France, e-mail: cegielski@univ-paris12.fr

Guessarian, Irène
LIAFA, UMR 7089 and Université Paris 6, 2 Place Jussieu, 75254 Paris Cedex 5, France; corresponding author e-mail: ig@liafa.jussieu.fr

ABSTRACT
We compare the control structure of Abstract State Machines (in short ASM) defined by Yuri Gurevich and AsmL, a language implementing it. AsmL is not an algorithmically complete language, as opposed to ASM, but it is closer to usual programming languages because it allows until and while iterations and sequential composition. We here give a formal definition of AsmL, its semantics, and we construct, for each AsmL program II, a normal form (which is an ASM program IIₙ) computing the same function as II. Keywords: Abstract State Machine, algorithm.

1. INTRODUCTION
Yuri Gurevich has given a schema of languages which is not only a Turing-complete language (a language allowing to express at least an algorithm for each computable function), but which also allows to express all algorithms for each computable function (it is an algorithmically complete language); this schema of languages was first called dynamic structures, then evolving algebras, and finally ASM (for Abstract State Machines) [4]. He proposed the Gurevich’s thesis (the notion of algorithm is entirely captured by the model) in [5].

There exist several partial implementations of ASMs as a programming language. These implementations are partial by nature for two reasons: (i) ASMs allow to program functions computable by Turing machines with oracles, but obviously not the implementations; (ii) in ASMs any (computable) first order structures may be defined, which is not true for the current implementations.

Yuri Gurevich has directed a group at Microsoft Laboratories (Redmond) which has implemented such a programming language, called AsmL (for ASM Language), written first in C++ then in C# as a language of the .NET framework of Microsoft. To invite programmers to use its language, this group extended the control structures used in ASMs. Pure ASM control structures are represented by normal forms of AsmL programs.

The aim of this paper is to give a formal definition of AsmL (more precisely of the part concerning control structures; we are not interested by constructions of first-order structures here), a formal definition of normal forms in AsmL and to show how to build a normal form from an AsmL program.

2. DEFINITIONS
2.1 Definition of ASMs
We first recall the definition of ASMs and then precise our point of view on ASMs because different definitions exist.

2.2 Syntax
Definition 1. An ASM vocabulary, or signature, is a first-order signature \( \mathcal{L} \) with a finite number of function symbols, two boolean constant symbols (true and false), a constant symbol (denoted by null or undef), logical connectives (not, and, or), and the equality predicate.

Terms are defined as usual in first order logic (see for instance [1]). Let us note that an ASM signature is a first-order signature whose only predicate symbol is equality. The fact that there are no other relation symbols simplifies some definitions and proofs but is not essential.

Definition 2. Boolean terms (or formulæ) are defined inductively by:

- if \( t \) and \( t' \) are terms, then \( t = t' \) is a formula;
- if \( p \) is a functional symbol onto \{true, false\} of arity \( n \) and \( t_1, \ldots, t_n \) are terms, then \( p(t_1, \ldots, t_n) \) is a formula;
- if \( F, F' \) are formulæ, then \( \neg F, F \lor F', F \lor F' \) are formulæ.

Definition 3. Let \( \mathcal{L} \) be an ASM signature. ASM rules are defined inductively as follows:

- An update rule is an expression of the form:

\[
f(t_1, \ldots, t_n) := t_0
\]

where \( f \) is a \( n \)-ary functional symbol and \( t_0, t_1, \ldots, t_n \) are closed terms of \( \mathcal{L} \).

- If \( R_1, \ldots, R_k \) are rules of signature \( \mathcal{L} \), then the following expression is also a rule of \( \mathcal{L} \), called block:

\[
\begin{array}{c}
\text{par} \\
R_1 \\
\vdots \\
R_k \\
\text{endpar}
\end{array}
\]

If \( k = 0 \), it is the empty instruction, denoted skip.

- Let \( \varphi \) be a boolean term, and let \( R_1 \) and \( R_2 \) be rules of signature \( \mathcal{L} \). The expression:

\[
\text{if } \varphi \text{ then } R_1 \text{ else } R_2
\]

is a rule of vocabulary \( \mathcal{L} \), called alternative rule.

The expression if \( \varphi \) then \( R_1 \) is a rule called test rule.
At the beginning of ASMs, par meant parallel, nowadays it stands for parenthesis.

Definition 4. Let \( \mathcal{L} \) be an ASM signature. A program on signature \( \mathcal{L} \), or \( \mathcal{L} \)-program is a synonym for a rule of that signature.

2.3 Semantics

Definition 5. If \( \mathcal{L} \) is an ASM signature, an ASM abstract state, or more precisely an \( \mathcal{L} \)-state, is a synonym for a first-order structure \( A \) of signature \( \mathcal{L} \) (an \( \mathcal{L} \)-structure).

The universe of \( A \) consists of the disjoint union of three sets: the basis set \( A \), the Boolean set \( \mathbb{B} = \{ \text{true}, \text{false} \} \), and a singleton set \( \{ \bot \} \). The values of the Boolean constant symbols \text{true} and \text{false} and of \text{null} (or \text{undef}) in \( A \) will be denoted by \text{true}, \text{false} and \text{null} (or \text{undef})

Definition 6. Let \( \mathcal{L} \) be an ASM signature and \( A \) a non empty set. A set of modifications (more precisely an \( (\mathcal{L}, A)\)-modification set) is any finite set of triples:

\[
(f, \pi, a),
\]
where \( f \) is a function symbol of \( \mathcal{L} \), \( \pi = (a_1, \ldots, a_n) \) is an \( n \)-tuple of \( A \) (where \( n \) is the arity of \( f \)), and \( a \) is an element of \( A \).

Definition 7. Let \( \mathcal{L} \) be an ASM signature, let \( A \) be an \( \mathcal{L} \)-state and let \( \Pi \) be an \( \mathcal{L} \)-program. Let \( \Delta_{\Pi}(A) \) denote the set defined by structural induction on \( \Pi \) as follows:

1. If \( \Pi \) is an update rule: \( f(t_1, \ldots, t_n) \mapsto t_0 \), then, denoting \( a_1 = t_1^A, \ldots, a_n = t_n^A \) and \( a = t_0^A \), the set \( \Delta_{\Pi}(A) \) is the singleton:

\[
\{(f, (a_1, \ldots, a_n), a)\}
\]
2. If \( \Pi \) is a block: \( \text{par} R_1 \ldots R_n \text{ endpar} \), then the set \( \Delta_{\Pi}(A) \) is the union:

\[
\Delta_{R_1}(A) \cup \ldots \cup \Delta_{R_n}(A)
\]
for \( n \neq 0 \) and the empty set otherwise.
3. If \( \Pi \) is a test: if \( \varphi \) then \( R \), we first have to evaluate the expression \( \varphi^A \). If it is false then the set \( \Delta_{\Pi}(A) \) is empty, otherwise it is equal to:

\[
\Delta_{\Pi}(A).
\]

The semantics of the alternative rule is similar.

We may check that \( \Delta_{\Pi}(A) \) is an \( (\mathcal{L}, A) \)-set of modifications.

Definition 8. A set of modifications is incoherent if it contains two elements \( (f, \pi, a) \) and \( (f, \pi, b) \) with \( a \neq b \). It is coherent otherwise.

Definition 9. Let \( \mathcal{L} \) be an ASM signature, \( \Pi \) an \( \mathcal{L} \)-program, and \( A \) an \( \mathcal{L} \)-state.

If \( \Delta_{\Pi}(A) \) is coherent, the transform \( \tau_{\Pi}(A) \) of \( A \) by \( \Pi \) is the \( \mathcal{L} \)-structure \( B \) defined by:

- the base set of \( B \) is the base set \( A \) of \( A \);
- for any element \( f \) of \( \mathcal{L} \) and any element \( \pi = (a_1, \ldots, a_n) \) of \( A^n \) (where \( n \) is the arity of \( f \)):
  - if \( (f, \pi, a) \in \Delta_{\Pi}(A) \) for an \( a \in A \), then:
    \[
    \tau_{\Pi}^f(\pi) = a;
    \]
  - otherwise:
    \[
    \tau_{\Pi}^f(\pi) = \tau_{\Pi}^{f_0}(\pi).
    \]
If \( \Delta_{\Pi}(A) \) is incoherent then \( \tau_{\Pi}(A) = A \) (hence the state is a fixed point).

Definition 10. Let \( \mathcal{L} \) be an ASM signature, \( \Pi \) an \( \mathcal{L} \)-program, and \( A \) an \( \mathcal{L} \)-state. The computation is the sequence \( (A_n)_{n \in \mathbb{N}} \) defined by:

- \( A_0 = A \) (called the initial algebra of the computation);
- \( A_{n+1} = \tau_{\Pi}(A_n) \) for \( n \in \mathbb{N} \).

A computation terminates if there exists a fixed point \( A_{n+1} = A_n \).

Definition 11. The semantics is the partial function which transforms \( \{[\Pi, A]\} \), where \( \Pi \) is an ASM program and \( A \) a state, in the result of iterating \( \tau_{\Pi} \) starting from \( \tau_{\Pi}(A) \) until a fixed point is reached (it will be denoted \( [[\Pi, A]] = \tau_{\Pi}(A) \)) if such a fixpoint exists, otherwise it is undefined.

3. DEFINITION OF EXTENDED ASMS

We call extended ASM the analog of ASM with additional rules to take into account the supplementary control structures of AsmL. No paper defines formally AsmL but [3] gives a precise informal description.

3.1 Syntax

Roughly speaking, AsmL considers more control structures than ASMs. We first define AsmL rules, then explain them.

Definition 12. Let \( \mathcal{L} \) be an ASM signature.

- An ASM rule is also an extended rule of \( \mathcal{L} \).
- If \( \varphi \) is a boolean term, and \( R_1 \) and \( R_2 \) are extended rules, then:
  \[
  \text{if } \varphi \text{ then } R_1 \text{ else } R_2 \quad \text{endif}
  \]
  is an extended rule, called alternative rule.
- If \( R_1, \ldots, R_k \) are extended rules of signature \( \mathcal{L} \), then the following expression is also an extended rule of \( \mathcal{L} \), called step rule:
  \[
  \text{par} \quad R_1, \ldots, R_k \quad \text{endpar}
  \]
If $R$ is an extended rule of signature $\mathcal{L}$, then the following expressions are also extended rules of $\mathcal{L}$, called iteration rules:
- step until $\varphi R$
- step while $\varphi R$
- step until fixpoint $R$

with $\varphi$ a boolean term.

We now explain the above rules. Following the semantics given above for ASMs, the meaning of the $\text{par}$ rule:

\[
\text{par}
R_1
\vdots
R_k
\text{endpar}
\]

is that rules $R_1, \ldots, R_k$ are running simultaneously, i.e. each rule is running (with a point of view sequential or parallel for the observer) independently of the other rules; incoherences might occur for some updates; if there is at least one incoherence, the program stops, otherwise updates are applied.

The paradigm of simultaneity is unusual for programmers who prefer sequentiality. Hence sequentiality was introduced using step in AsmL. The meaning of the extended rule

\[
\text{par} \text{ step}_1 \vdots \text{step}_k \text{ endpar}
\]

is as follows: rule $\text{step}_1$ is running first; then rule $\text{step}_2$ is running (with values of closed terms depending of the result of running rule $\text{step}_1$); and so on.

Finally classical iterations appear in AsmL via the step until and step while iteration rules, and the step until fixpoint rule of ASM is also present.

### 4. NORMALIZATION

**Definition 13.** Let $\mathcal{L}$ be an ASM vocabulary. If $R$ is an extended rule of vocabulary $\mathcal{L}$, then the expression step until fixpoint $R$ is called a running rule.

Running rules are implicit in ASMs and are made explicit in AsmL. If $R$ is an ASM rule (not extended), then step until fixpoint $R$ will have the same semantics as the ASM program $\Pi$ consisting of rule $R$.

**Definition 14.** Let $\mathcal{L}$ be an ASM vocabulary. If $R_1, \ldots, R_k$ are ASM rules of vocabulary $\mathcal{L}$ (not extended rules), then the extended rule:

\[
\text{step until fixpoint}
\text{par}
R_1
\vdots
R_k
\text{endpar}
\]

is a normal form of $\mathcal{L}$, where mode, $c$ are special nullary functional constants of the ASM language whose value for the initial algebra are 0.

The rule under step until fixpoint is called the core of the normal form.

In the sequel $\text{par}$ and $\text{endpar}$ will be omitted: when rules have the same indentation, they will be assumed to be in a $\text{par} \ldots \text{endpar}$ block.

To give a normal form for a given program is a classical issue of theoretical computer science.

**Example** Consider the following programming problem: to compute the average of an array of grades and to determine the number of grades whose value is greater than this average. A natural program in extended ASMs is:

\[
\begin{align*}
\text{step} & \quad \text{avg} := 0 \\
& \quad \text{i} := 0 \\
& \quad \text{step until (i = n)} \\
& \quad \text{avg} := \text{avg} + \text{grade}[i] \\
& \quad \text{i} := \text{i} + 1 \\
& \quad \text{step} & \quad \text{avg} := \text{avg}/n \\
& \quad \text{i} := 0 \\
& \quad \text{nb} := 0 \\
& \quad \text{step until (i = n)} \\
& \quad \text{if (grade}[i] > \text{avg}) & \quad \text{nb} := \text{nb} + 1 \\
& \quad \text{i} := \text{i} + 1
\end{align*}
\]

using the conventions of AsmL (a same indentation indicates a block).

The core of the normal form is:

\[
\begin{align*}
& \quad \text{if (mode = 0)} & \quad \text{avg} := 0 \\
& \quad & \quad \text{i} := 0 \\
& \quad & \quad \text{mode} := 1 \\
& \quad \text{if (mode = 1)} & \quad \text{avg} := \text{avg} + \text{grade}[i] \\
& \quad & \quad \text{i} := \text{i} + 1 \\
& \quad & \quad \text{if (i = n)} & \quad \text{mode} := 2 \\
& \quad \text{if (mode = 2)} & \quad \text{avg} := \text{avg}/n \\
& \quad & \quad \text{i} := 0 \\
& \quad & \quad \text{nb} := 0 \\
& \quad & \quad \text{mode} := 3 \\
& \quad \text{if (mode = 3)} & \quad \text{if (grade}[i] > \text{avg}) & \quad \text{nb} := \text{nb} + 1 \\
& \quad & \quad \text{i} := \text{i} + 1 \\
& \quad & \quad \text{if (i = n)} & \quad \text{mode} := 4 \\
& \quad \text{if (mode = 4)} & \quad \text{skip}
\end{align*}
\]

**Theorem 15.** For every extended ASM program $\Pi$, there exists a normal form ASM program $\Pi_n$ such that for every $\mathcal{L}$-state $A$ whose base set is infinite there exists an $\mathcal{L}_m$-state $A_m$ in which $\Pi_n$ computes the same function as $\Pi$.

**Proof.** We define the core of the normal form $\Pi_n$ of $\Pi$ by structural induction on the form of program $\Pi$. The proof that $\Pi_n$ and $\Pi$ compute the same function can be found in [2].
We will need to label some blocks. To allow this, we suppose the base set is infinite, and we use special nullary functions, called \textit{mode}, \textit{c} and whose corresponding constant symbols are added to the ASM language, i.e. \( \mathcal{L}_m = \mathcal{L} \cup \{ \text{mode}, \text{c} \} \). Without loss of generality, we suppose the set \( \mathbb{N} \) of natural integers is included in the base set.

- The core of the normal form \( R_m \) of an ASM rule \( R \) is:
  
  \[
  \text{if } (\text{mode} = 0) \text{ then } R_m \text{ \quad mode} := 1
  \]

- If \( R'_1, \ldots, R'_n \) are the respective cores of the extended rules \( R_1, \ldots, R_n \), we have to define the core of the extended rule \( \Pi \):
  
  \[
  \text{par}
  \begin{align*}
  &\text{step } R_1 \\
  &\ldots \\
  &\text{step } R_n
  \end{align*}
  \text{endpar}
  \]

  In the (abnormal) case \( k = 1 \), the core is simply \( R'_1 \).

  To explain the case \( k \geq 2 \), it is sufficient to suppose \( k = 2 \). In this case the core of \( \Pi_m \) is:
  
  \[
  R''_1 \\
  R''_2^{\text{fin}+1}
  \]

  with

  1. \( R''_1 = R'_1 \)
  2. \( R''_2^{\text{fin}+1} \) is defined as follows: Let \( \text{fin} \) be the greatest value used for \textit{mode} in \( R'_1 \). Rule \( R''_2^{\text{fin}+1} \) is \( R'_2 \) where \( \text{fin} \) is added to each constant occurring on the right side of the “\( = \)” (or “\( \geq \)” ) sign of an expression rule beginning by \textit{mode} (a boolean expression \( \textit{mode} = \text{constant} \) or an update rule \( \textit{mode} := \text{constant} \)).

- If \( \Pi \) is an iteration rule (\textit{step until} \( \varphi \) \( \mathcal{R} \)) and \( R' \) is the core of the normal form of \( R \) then the core \( \Pi' \) of the normal form \( \Pi_m \) of \( \Pi \) is
  
  \[
  \text{if } (\text{mode} = 0) \text{ then }
  \begin{align*}
  &\text{if } \neg \varphi \text{ then mode} := 1 \\
  &\text{if } \varphi \text{ then mode} := \text{fin} + 1
  \end{align*}
  \]

  where \( R'^{+1} \) is \( R' \) in which each constant occurring on the right side of the “\( = \)” (or “\( \geq \)” ) sign of an expression rule beginning by \textit{mode} (a boolean expression \( \textit{mode} = \text{constant} \) or an update rule \( \textit{mode} := \text{constant} \)) is incremented by 1 but for the greatest such value \( \text{fin} \), which is replaced by 0.

- If \( \Pi \) is an iteration rule (\textit{step until fixpoint} \( R \)), then the stopping condition of the iteration is defined as follows in [3] “In the case of until fixpoint, the stopping condition is met if no non-trivial updates have been made in the step. Updates that occur to variables declared in abstract state machines that are nested inside the fixed-point loop are not considered. An update is considered non-trivial if the new value is different from the old value.” We thus must check that all updates are trivial before stopping. Let \( f_i(t_1, \ldots, t_n) := t_0, i = 1, \ldots, k \), be all the (non nested) updates performed in \( R \). Let us introduce a new constant \( c \) (keeping track of non trivial updates), with value 0 in the initial algebra.

  Let \( R' \) be the core of the normal form of \( R \) and \( M \) the largest value of \textit{mode} occurring in \( R' \), then the core \( \Pi' \) of the normal form \( \Pi_m \) of \( \Pi \) is:
  
  \[
  \begin{align*}
  &\text{if } (c=1) \land (M \geq \text{mode} \geq 0) \text{ then } \\
  &\quad \text{mode} := 0 \\
  &\text{if } (c=0) \land (\text{mode} = 0) \text{ then } R''_1 \\
  &\quad R''_1 \text{ is } R' \text{ in which each update (other than updates on } c \text{ and } \text{mode} \text{) } f_i(t_1, \ldots, t_n) := t_0 \text{ has been replaced by }
  \\
  &\quad \text{if } f_i(t_1, \ldots, t_n) \neq t_0 \text{ then } \\
  &\quad \quad \text{if } \varphi \text{ then mode} := 1 \\
  &\quad \quad \text{if } \neg \varphi \text{ then mode} := \text{fin} + 1
  \end{align*}
  \]

  \[
  R'^{+1}
  \]

- If \( \Pi \) is an alternative rule, \( \Pi' \) is:
  
  \[
  \begin{align*}
  &\text{if } \varphi \text{ then } R_1 \\
  &\quad \text{else } R_2 \\
  &\quad \text{and } R'_1, R'_2 \text{ are the respective cores }
  \end{align*}
  \]

  of \( R_1, R_2 \), then the core of the normal form of \( \Pi \) is:

  \[
  \begin{align*}
  &\text{if } (\text{mode} = 0) \text{ then }
  \\
  &\quad \text{if } \varphi \text{ then mode} := 1 \text{ else mode} := \text{fin} + 1
  \end{align*}
  \]

  \[
  R'^{+1}_{\text{fin}+1}
  \]

  where \( R'^{+k} \) is \( R'_k \) where \( k \) is added to each constant occurring on the right side of the “\( = \)” (or “\( \geq \)” ) sign of an expression rule beginning by \textit{mode}.

\[ \square \]

REFERENCES


[2] **Cegielski, Patrick, GuesSarian, Irène**, Normal Forms for Extended Abstract State Machines, http://www.liafa.jussieu.fr/~ig/g70.ps

