This chapter presents some techniques for verifying programs with complex data structures. Typical examples of such data structures are arrays and lists. The main ingredient is to provide extensions to the language of expressions and formulas, under the form of an expressive specification language.

Section 4.1 presents generalities about specification languages, while other sections present an overview of the constructs for modeling and specifying as they are available in the Why3 tool [7]. Finally, Section 4.6 is dedicated to the modeling and the proof of programs on the array data structure.

4.1 Generalities on Specification Languages

There have been a lot of variants of specification languages proposed in the past. Various classes of such languages are characterized by the kind of logic that they are based on.

Algebraic Specifications  This class of languages is characterized by the use of multi-sorted first-order logic with equality, and usually the use of rewriting techniques [13], which permits to mix executable-style and declarative-style specifications. Famous languages and tools of this class are Larch [17], KIV [26], CASL (Common Algebraic Specification Language) [2], OBJ [16], Maude [10].

Set-theory based approaches  Another class of languages is based on classical Set theory. These languages are not typed, the typing properties being instead encoded by formulas on sets. Reasoning on such languages is also typically done by rewriting techniques. A precursor of this class is the Z notation [27]. It is used in tools like VDM [18], Atelier B [1], etc.

Higher-Order Logic  Higher-Order Logic is characterized by the use of mathematical functions as objects of the language. It is based on advanced lambda-calculi (e.g. System F). Reasoning on such specifications is typically done using interactive environment where proofs are constructed manually using tactics/strategies. Famous tools implementing some higher-order logics are PVS [24], Isabelle [22], HOL4 [23], Coq [5], etc.

Object-Oriented languages  These are based on the idea of using the object-oriented paradigm in a logic world. They are typically in use in a context of object-oriented-style development. The precursor is the language Eiffel [20], which is both a programming language and a specification language, and was designed to support the concept of design-by-contract. Other famous languages are OCL [28] (Object
Constraint Language) which allows to incorporate formal specifications into the UML diagrammatic design methodology, and the Java Modeling Language [9].

**Satisfiability Modulo Theories** In 1980, Nelson and Oppen [21] proposed a new technique for combining propositional reasoning and dedicated decision procedures for specific theories. This gave birth to a new class of automated theorem provers called SMT solvers. A precursor of such solvers is Simplify [14] which was used for program verification in the ESC/Java tool. The SMT-LIB format [25] is not really a specification language but a common language for describing verification conditions as they typically come from program verification. In this context, the intermediate languages for program verification Why [15] and Boogie [3] propose small specification languages that can be supported by SMT solvers. Nowadays, the successors of Why and Boogie, respectively Why3 [7] and Dafny [19], propose modeling languages that can be seen as a compromise between:

- Expressivity of the language, to be able to design specifications of complex programs;
- Support by automated provers, in particular modern SMT solvers like Alt-Ergo [6], Z3 [12], CVC3 [4] which support quantified first-order formulas.

The following sections present an overview of the constructs of the Why3 specification language.

### 4.2 Defined Functions and Predicates

Direct definitions of new functions and predicates can be given under the form

\[
\text{function } f(x_1 : \tau_1, \ldots, x_n : \tau_n) : \tau = e \\
\text{predicate } p(x_1 : \tau_1, \ldots, x_n : \tau_n) = e
\]

where \(\tau_i, \tau\) are not reference types.

In such definitions, no recursion is allowed, unless we are in the case of Section 4.3.3. Since expressions have no side-effects, these define total functions and predicates, which are *pure* in the sense that they have no side-effects.

**Example 4.2.1** Here are a few examples of definitions of functions and predicates.

\[
\text{function } \text{sqr}(x : \text{int}) : \text{int} = x \ast x \\
\text{function } \text{min}(x : \text{real}, y : \text{real}) : \text{real} = \text{if } x \leq y \text{ then } x \text{ else } y \\
\text{function } \text{abs}(x : \text{real}) : \text{real} = \text{if } x \geq 0.0 \text{ then } x \text{ else } -x \\
\text{predicate } \text{in}_{0..1}(x : \text{real}) = 0.0 \leq x \land x \leq 1.0 \\
\text{predicate } \text{prime}(x : \text{int}) = x \geq 2 \land \forall z : \text{int} . y \geq 0 \land z \geq 0 \land x = y \ast z \rightarrow y = 1 \lor z = 1
\]

### 4.3 Constructed Types

New logic data types can be defined by standard construction of compound types. Being in a logic language, these types must be pure in the sense of pure functional programming, that is, the values of these types cannot be modified “in-place”. 
4.3.1 Product Types

Types can be defined as being tuples of already defined types, for example

```plaintext
type intpair = (int, int)
function sumprod (n:int,m:int) : intpair = (n+m,n*m)
type vector = (real, real, real)
function diag(x:real) : vector = (x,x,x)
```

The construct `let x=e1 in e2` is extended so as to allow introducing a tuple, as in

```plaintext
function sum (p:intpair) =
  let (x,y) = p
  in
  x+y
```

The other kind of product type is the record type, defined classically by a sequence of named fields. The dot notation allows to access a given field. For example,

```plaintext
type vector = { x:real; y:real }
function norm(v:vector) : real = sqrt(v.x * v.x + v.y * v.y)
function rotate(v:vector,a:real) : vector =
  { x = cos(a) * v.x + sin(a) * v.y ;
    y = sin(a) * v.y - sin(a) * v.x }
```

Again, it is important to notice that these are pure types; the components of tuples and the fields of records are immutable. In other words, there is no way to assign a field.

4.3.2 Sum Types

We can define new sum types in a functional programming style, with the form of declaration

```plaintext
type t =
  | C_1(\tau_1,1,\ldots,\tau_1,n_1)
  | \ldots
  | C_k(\tau_k,1,\ldots,\tau_k,n_k)
```

where \( C_1,\ldots,C_k \) are constructors.

**Example 4.3.1** The type of cards in a standard 54-card game can be defined as

```plaintext
type suit = Spade | Heart | Diamond | Club
type value = Ace | King | Queen | Jack | Number(int)
type card = Normal(suit,value) | Joker
```

Associated to sum types is the definition by pattern-matching, with the form

```plaintext
match expression with
  | C_1(x_1,1,\ldots,x_1,n_1) \rightarrow expression_1
  | \ldots
  | C_k(x_k,1,\ldots,x_k,n_k) \rightarrow expression_k
end
```

where each \( x_{i,j} \) is a variable that is bound in \( expression_i \). For convenience, one may use extended pattern-matching [8]: inside patterns, an argument of a constructor may itself be a pattern. One may also use the wildcard character \_\_. As in OCaml or Why3, we distinguish between constructors and variables by enforcing a capital letter as the first character of a constructor.

**Example 4.3.2** Here is a function that returns the score of a given card in an imaginary game.
function score (c:card,trump:suit): int =
match c with
| Joker → 100
| Normal(Ace,s) → if s=trump then 50 else 20
| Normal(Number(n),s) → if s=trump then 50 else n
| Normal(v,_) →
  match v with
  | King | Queen → 40
  | Jack → 30
  | _ → 0
end
end

4.3.3 Recursive Sum Types

Sum type definitions can be recursive. Recursive definitions of functions or predicates are allowed if recursive calls are on structurally smaller arguments, which corresponds to the notion of primitive recursion on such a sum type.

Example 4.3.3 Here is a definition of binary trees holding integers at each node, and a function giving the sum of the elements of a tree.

type tree = Leaf | Node(tree,int,tree)
function sum (t:tree) : int =
  match t with
  | Leaf → 0
  | Node(l,x,r) → sum(l) + x + sum(r)
end

The recursion is allowed since arguments l and r are subtrees of the input.

4.3.4 Polymorphic Types

All the constructs for type definitions above can be parameterized, i.e. the type is given one or several type variables as parameters, as in

type vector α = (α,α,α)
type tagged α = { tag : int ; content : α }
type option α = None | Some(α)
type tree α β = Leaf(α) | Node(tree α β, β, tree α β)

4.3.5 Example of Lists

Lists, or finite sequences, are very common data structures. The type of lists can be modeled by a polymorphic, recursive sum type as

type list α = Nil | Cons α (list α)

Classical functions on lists can be defined by primitive recursion, such as

function append(l1:list α,l2:list α) : list α =
  match l1 with
  | Nil → l2
When proving programs operating on lists, and on recursive sum types in general like binary trees above, it is very common that some properties are needed, such as the properties

**lemma** append_nil: \( \forall l: \text{list } \alpha. \ append(l,\text{Nil}) = l \)

**lemma** append_assoc:
\[
\forall l1 \ l2 \ l3: \text{list } \alpha. \ append(append(l1,l2),l3) = append(l1,append(l2,l3))
\]

Such lemmas can be stated as this in the Why3 tool. They must be proved by induction, meaning that they are out of reach to SMT solvers. In the Why3 tool, an interactive theorem prover like Coq can be used to prove them. See Exercise 4.7.6 at end of this chapter.

### 4.4 Abstract Declarations, Axiomatizations

#### 4.4.1 Axiomatic Declarations of Functions and Predicates

Instead of defining new logic functions, it is sometimes handy to model abstract functions, in the sense that only their profile is declared, and their meaning is partially specified using axioms.

Such an abstract declarations has the form

**function** \( f(\tau_1, \ldots, \tau_n) : \tau \)

**predicate** \( p(\tau_1, \ldots, \tau_n) \)

and is accompanied with axioms

**axiom** id : formula

Such a declaration specifies that \( f \), resp. \( p \), is any symbol satisfying the axioms.

All functions in the logic are assumed total, so that there is no possible “runtime error” in the specifications themselves. Axiomatizations provide a way to specify partial functions, in the sense that the axioms do not specify their results outside their domains.

**Example 4.4.1** A typical example is the division which can be axiomatized as

**function** div(real, real) : real

**axiom** mul_div: \( \forall x,y: \text{real}. \ y \neq 0.0 \implies \text{div}(x,y) * y = x \)

An expression like \( \text{div}(1.0,0.0) \) is a valid expression, but its value is not specified.

**Example 4.4.2** The factorial function can be axiomatized as follows.
function fact(int):int
axiom fact0: fact(0) = 1
axiom factn: forall n:int. n ≥ 1 → fact(n) = n * fact(n-1)

The values of fact(n) for n negative are not specified.

Example 4.4.3 Taking respectively the head and the tail of a list is undefined when applied on an empty list. These two functions cannot be defined directly by pattern-matching, but they can be axiomatized as follows.

function head(l:list α):α
axiom head_cons: forall x:α, l:list α. head(Cons(x,l)) = x

function tail(l:list α):list α
axiom tail_cons: forall x:α, l:list α. tail(Cons(x,l)) = l

Again the values of head(Nil) and tail(Nil) are not specified.

4.4.2 A Warning about Consistency Issues

We repeat that the meaning of an axiomatization is that the abstract functions and predicates indeed denote any function or predicate satisfying the given axioms. This does not cause any problem of soundness but a potential problem of inconsistency: the axioms may be contradictory in the sense that no function could satisfy all axioms at the same time.

A toy example of inconsistent axiomatization is

function f(int):int
axiom f0: f(0) = 1
axiom f1: f(0) = 2

A less toy example is

function div(real, real):real
axiom mul_div: forall x, y. div(x, y) * y = x

from which we get div(1.0, 0.0) * 0.0 = 1.0 which contradicts forall x:real. x * 0.0 = 0.0.

Typically, in such an context, an automated theorem prover will exploit the inconsistency by proving any goal it is given, including false. In other words, special care must be taken when designing axiomatizations, to avoid introducing inconsistencies.

4.4.3 Axiomatic Type Declarations

Not only functions but also data types can be declared abstractly, by a declaration of the form

type t

For example, one may introduce a type of colors with

type color
function blue: color
function red: color
axiom distinct: red ≠ blue

Unlike sum types, such a type is open in the sense that there may exist values of this type other than the given constants. For instance, one could introduce later on

function green: color
function rgb(r:real, g:real, b:real): color
axiom green_rgb: green = rgb(0.0, 1.0, 0.0)
4.4.4 Polymorphic Types

Abstract types may also be declared as polymorphic, with a declaration of the form

\[ \text{type } t \, \alpha_1 \ldots \alpha_k \]

where \( \alpha_1 \ldots \alpha_k \) are type parameters

A classical example is the type of sets of elements of some type, which can axiomatized as follows.

\[ \text{type } \text{set } \alpha \]

\[ \text{function } \text{empty}: \text{set } \alpha \]

\[ \text{function } \text{single}(\alpha): \text{set } \alpha \]

\[ \text{function } \text{union}(\text{set } \alpha, \text{set } \alpha): \text{set } \alpha \]

\[ \text{axiom } \text{union_assoc}: \forall x \, y \, z: \text{set } \alpha. \text{union}(\text{union}(x, y), z) = \text{union}(x, \text{union}(y, z)) \]

\[ \text{axiom } \text{union_comm}: \forall x \, y: \text{set } \alpha. \text{union}(x, y) = \text{union}(y, x) \]

\[ \text{predicate } \text{mem}(\alpha, \text{set } \alpha) \]

\[ \text{axiom } \text{mem_empty}: \forall x: \alpha. \neg \text{mem}(x, \text{empty}) \]

\[ \text{axiom } \text{mem_single}: \forall x \, y: \alpha. \text{mem}(x, \text{single}(y)) \leftrightarrow x = y \]

\[ \text{axiom } \text{mem_union}: \forall x: \alpha, \, y \, z: \text{set } \alpha. \text{mem}(x, \text{union}(y, z)) \leftrightarrow \text{mem}(x, y) \lor \text{mem}(x, z) \]

4.5 Inductive Predicates

Defining a predicate \emph{inductively} is a notion that originally comes from the theory of automated reasoning based on resolution. A \emph{Horn clause} is a first order clause, i.e. a disjunction of literals, with at most one positive literal. For this form of clauses, there exist efficient reasoning methods, which gave birth to the Prolog programming language.

An inductive definition of a predicate is a kind of axiomatization where the axioms are Horn clauses. They appear in proof environments like Coq or PVS. It has the form

\[ \text{inductive } p(\tau_1, \ldots, \tau_n) = \]

\[ | \text{id}_1: \text{clause}_1 \]

\[ \ldots \]

\[ | \text{id}_k: \text{clause}_k \]

where each clause has the form

\[ \forall \bar{x}. \text{hyp} \to p(\bar{e}_1, \ldots, \bar{e}_n) \]

and \( p \) occurs only positively in \( \text{hyp} \) (thus it is a Horn clause). Another point of view on such definitions is that they define uniquely the predicate, satisfying the clauses, that is true the less often. Such a “smallest” predicate exists, it is the smallest fix-point of the set of clauses.

A typical case when an inductive definition comes naturally is the definition of the transitive closure of a relation.

\[ \text{predicate } r(x: t, y: t) = \ldots \]

\[ \text{inductive } r\_\text{star}(t, t) = \]

\[ | \text{empty: forall } x : t. \ r\_\text{star}(x, x) \]

\[ | \text{single: forall } x \, y : t. \ r(x, y) \to r\_\text{star}(x, y) \]

\[ | \text{trans: forall } x \, y \, z : t. \ r\_\text{star}(x, y) \land r\_\text{star}(y, z) \to r\_\text{star}(x, z) \]
The definition should informally read as “\( r_{\ast} \) is the smallest reflexive and transitive relation that includes \( r \).

4.6 Programs with Arrays

The most common data type appearing in programming languages is the type of arrays. Proving programs with arrays can be done by considering arrays as references to purely functional arrays, also called maps. Purely functional arrays are just one particular case of a logic datatype that can be axiomatized in first-order logic. The typical axiomatization is as follows.

\[
\begin{align*}
\text{type} & \quad \text{map} \quad \alpha \\
\text{function} & \quad \text{select} \ (\text{map} \ \alpha, \ \text{int}) : \alpha \\
\text{function} & \quad \text{store} \ (\text{map} \ \alpha, \ \text{int}, \ \alpha) : \text{map} \ \alpha \\
\text{axiom} & \quad \text{select\_store\_eq}:
\forall a: \text{map} \ \alpha, \ i: \text{int}, \ v: \alpha. \ \text{select} (\text{store}(a, i, v), i) = v \\
\text{axiom} & \quad \text{select\_store\_neq}:
\forall a: \text{map} \ \alpha, \ i, j: \text{int}, \ v: \alpha. \ i \neq j \rightarrow \ \text{select} (\text{store}(a, i, v), j) = \text{select}(a, j)
\end{align*}
\]

Type \( \text{map} \ \alpha \) denotes maps indexed by integers holding values of type \( \alpha \). In this axiomatization, the indexes are unbounded; the case of bounded arrays will be addressed later. The function \( \text{select} \) allows to access the \( i \)-th element of a map, that is, \( \text{select}(a, i) \) models the usual notation \( a[i] \). The function \( \text{store} \) denotes the functional update of a map, that is, \( \text{store}(a, i, v) \) denotes a new map whose values are identical to those of \( a \), except at index \( i \) where the value is the given \( v \). The two axioms above axiomatize this informal definition of \( \text{store} \).

It should be noticed that the theory of maps as above was one of the very first “built-in” theory considered into SMT solvers.

In a program, an array variable is seen as a reference to such a map. The standard assignment operation of the form \( a[i] := e \) is interpreted as \( a := \text{store} \ a \ i \ e \).

Remember that our IMP language does not allow references to references. Hence, with the approach of arrays above there is no way to build an array of references. In particular, this representation of arrays by references to maps cannot model arrays of arrays were some sharing of sub-arrays is allowed. Sharing sub-arrays raises the so-called aliasing issues and will be discussed in a next chapter.

A toy example of a program on an array is as follows.

\[
\begin{align*}
\text{val} & \quad a: \ \text{ref} \ (\text{map} \ \text{int}) \\
\text{procedure} & \quad \text{test}() \\
& \quad \text{writes} \ a \\
& \quad \text{ensures} \ \text{select}(a, 0) = 13 \\
\text{body} & \quad a := \text{store}(a, 0, 13); \quad (* \ a[0] := 13 *) \\
& \quad a := \text{store}(a, 1, 42) \quad (* \ a[1] := 42 *)
\end{align*}
\]
The proof using WP proceeds as follows:

\[
\begin{align*}
&\text{WP}((a := \text{store}(a, 0, 13); a := \text{store}(a, 1, 42)), \text{select}(a, 0) = 13) \\
&= \text{WP}(a := \text{store}(a, 0, 13), \text{WP}(a := \text{store}(a, 1, 42), \text{select}(a, 0) = 13)) \\
&= \text{WP}(a := \text{store}(a, 0, 13), \text{select}(\text{store}(a, 1, 42), 0) = 13) \\
&= \text{WP}(a := \text{store}(a, 0, 13), \text{select}(a, 0) = 13) \\
&\quad \text{(thanks to axiom select_store_neq)} \\
&= \text{select}(\text{store}(a, 0, 13), 0) = 13 \\
&= 13 = 13 \\
&\quad \text{(thanks to axiom select_store_eq)}
\end{align*}
\]

4.7 Exercises

Exercise 4.7.1 Find appropriate precondition, postcondition, loop invariant, and variant, for the following program, using the axiomatization of Example 4.4.2.

```plaintext
function fact_imp (x:int): int
requires ?
ensures ?
body
let ref y = 0 in
let ref res = 1 in
while y < x do
  y := y + 1;
  res := res * y
done;
res
```

Prove your program using WP.

Exercise 4.7.2 A procedure that permutes the contents of cells \(i\) and \(j\) in an array \(a\) is as follows.

```plaintext
val a: ref (map int)

procedure swap(i:int,j:int)
  writes a
  ensures select(a,i) = select(a@Old,j) \land select(a,j) = select(a@Old,i) \land
          \forall k:int. k \neq i \land k \neq j \rightarrow select(a,k) = select(a@Old,k)
body
  let tmp = select(a,i) in (* tmp := a[i] *)
  a := store(a,i,select(a,j)); (* a[i] := a[j] *)
  a := store(a,j,tmp) (* a[j] := tmp *)
```

- Prove swap using WP.
- Prove the procedure

```plaintext
procedure test ()
  requires select(a,0) = 13 \land select(a,1) = 42 \land select(a,2) = 64
  ensures select(a,0) = 64 \land select(a,1) = 42 \land select(a,2) = 13
body swap(0,2)
```
Exercise 4.7.3 Incrementing an array of reals.

1. Write a program over 3 variables: an array of reals \( a \), and two integers \( \text{min} \) and \( \text{max} \). The program should increment by 1 all the values of \( a \) between \( \text{min} \) and \( \text{max} \).

2. Specify a post-condition stating that the values between indexes \( \text{min} \) and \( \text{max} \) are incremented.

3. Specify an appropriate loop invariant.

Exercise 4.7.4 The goal is to specify and prove algorithms for searching a value in an array. Such a function is expected to fit the following canvas.

```plaintext
val a: ref(map real)

function search (n:int, v:real): int
  requires 0 ≤ n
  ensures ?
  body ?
```

1. Propose a suitable postcondition specifying that if \( v \) occurs in array \( a \), between indexes 0 and \( n - 1 \), then the result is an index of such an occurrence, otherwise \(-1\) is returned.

2. Implement and prove an algorithm of linear search, that follows the pseudo-code:

```plaintext
for each i from 0 to n - 1: if a[i] = v then return i
```

3. Specify and prove an algorithm of binary search, that assumes the array sorted in increasing order, and follows the pseudo-code:

```plaintext
low := 0; high := n - 1;
while low ≤ high
  let m be the middle of low and high
  if a[m] = v then return m
  if a[m] < v then continue search between m + 1 and high
  if a[m] > v then continue search between low and m - 1
```

Exercise 4.7.5 The goal is to specify and prove algorithms for sorting an array of reals in increasing order. Such a procedure is expected to fit this canvas:

```plaintext
val a: ref(map real)

procedure sort(n:int)
  requires 0 ≤ n
  writes a
  ensures ?
  body ?
```

1. Propose a suitable postcondition specifying that at the end of execution:

- the array is in increasing order between indexes 0 and \( n - 1 \),
- the array at exit is a permutation of the array at entrance.
It is suggested to use an inductive definition to formalize the permutation property.

2. Implement and prove an algorithm of selection sort, that follows the pseudo-code:

   for each \(i\) from 0 to \(n - 1\):
   
   finds the index \(idx\) of the minimal element between \(i\) and \(n - 1\);
   
   swap elements at indexes \(i\) and \(idx\)

The use of an auxiliary procedure is recommended.

3. Implement and prove an algorithm of insertion sort, that follows the pseudo-code:

   for each \(i\) from 1 to \(n - 1\):
   
   insert element at index \(i\) at the right place between indexes 0 and \(i - 1\)

Exercise 4.7.6 We consider a procedure that computes the reversal of a list, in an efficient way (i.e. in linear time in function of the size of the list).

```ocaml
val l : ref (list int)

procedure rev_append(r : list int)
  variant ?
  writes l
  ensures ?
  body
    match r with
    | Nil  → skip
    | Cons(x,r) →
        l := Cons(x,l);
        rev_append(r)
    end

procedure rev(r : list int)
  writes l
  ensures l = rev r
  body ?
```

1. Fill the ? with appropriate annotations or code.

2. Prove the program using WP. Which general lemmas on lists are needed?

Exercise 4.7.7 Resolution of a second degree equation.

- Write an IMP program that assumes 3 variables \(a, b, c\) of type real and stores into a variable of type (list real) the set of solutions of the equation \(ax^2 + bx + c\).
- Specify the expected behavior in the post-condition.
- Prove the program. You may pose lemmas to state some mathematical results.

Exercise 4.7.8 Sum of elements in a list.
• Using the standard datatype of polymorphic lists, define a logic function that gives the sum of elements of a list of real numbers.

• Assuming 2 global variables \( l \) of type \( \text{list} \ \text{real} \) and \( s \) of type \( \text{real} \), write a program that eventually stores in \( s \) the sum of elements of \( l \). You may add extra local variables if needed.

• Propose a post-condition that specifies the expected behavior of the program.

• Propose appropriate loop invariants and variants to allow the proof of the program using WP.

Example 4.7.9 Considering polymorphic binary trees defined as

\[
\text{type} \ \text{tree} \ \alpha = \text{Leaf} \mid \text{Node}(\text{tree} \ \alpha, \alpha, \text{tree} \ \alpha)
\]

specify, implement, and prove a procedure returning the maximum of a tree of integers.\(^1\)

Bibliography


---

\(^1\)Problem 2 of the FoVeOOS verification competition in 2011, http://foveoos2011.cost-ic0701.org/verification-competition


