Constrained Types for Object-Oriented Languages

Vijay Saraswat[1], Nathaniel Nystrom[1], Jens Palsberg[2], and Christian Grothoff[3]

[1] IBM T. J. Watson Research Center, P.O. Box 704, Yorktown Heights NY 10598 USA, vsaraswa@us.ibm.com
[2] UCLA Computer Science Department, Boelter Hall, Los Angeles CA 90095 USA, palsberg@cs.ucla.edu
[3] Department of Computer Science, University of Denver, 2360 S. Gaylord Street, John Green Hall, Room 214, Denver CO, 80208 USA, christian@grothoff.org

Abstract. X10 is a modern object-oriented language designed for productivity and performance in concurrent and distributed systems, such as (heterogeneous) multicores and clusters. In this context, dependent types arise naturally: objects may be located at one of many places, arrays may be multidimensional, activities may be associated with one or more clocks, variables may be marked as shared or private following an ownership discipline, etc. A framework for dependent types offers significant opportunities for detecting design errors statically, documenting design decisions, eliminating costly runtime checks (e.g., for array bounds, null values), and improving the quality of generated code.

We present the design and implementation of constraint-based dependent types in X10. The system is parametric on an underlying constraint system C: the compiler provides a framework that supports extension with new constraint systems using compiler plugins. Classes and interfaces are associated with *properties* (= final instance fields). A type C(:c) names a class or interface C and a *constraint* c on the properties of C and in-scope final variables. Constraints may also be associated with class definitions (representing class invariants) and with method and constructor definitions (representing preconditions). Dynamic casting is permitted.

In conclusion, we believe that constrained types offer a natural, simple, clean, and expressive extension to OO programming.

1 Introduction

X10 is a modern statically typed object-oriented language designed for high productivity in the high performance computing (HPC) domain [47]. Built essentially on the imperative sequential Java™ core, X10 introduces constructs for distribution and fine-grained concurrency (asynchrony, atomicity, ordering).

X10, like most object-oriented languages is designed around the notion of objects as instances of *classes*. However, X10 places equal emphasis on *arrays*, a central data-structure in high performance computing. In particular, X10 supports dense, distributed multi-dimensional arrays of value and reference types, built over index sets known as *regions*, and mappings from index sets to places, known as *distributions*. X10 supports a rich algebra of operations over regions, distributions and arrays.

A central design goal of X10 is to rule out large classes of error by design. For instance, the possibility of indexing a 2-d array with 3-d points should simply be ruled

out at compile-time. This means that one must permit the programmer to express types such as region(2), the type of all two-dimensional regions; int[5], the type of all arrays of int of length 5; int[region(2)], the type of all int arrays over two-dimensional regions; and Object!, the type of all Object located on the current node. For concurrent computations, one needs the ability to statically check that a method is being invoked by an activity that is registered with a given clock (i.e., dynamic barrier) [47].

In this paper we describe X10's support for *constrained types*. Constrained types are a form of *dependent type* [27, 53, 36, 6, 7, 3, 13]—types parametrized by *values*—defined on top of predicates over the *immutable* state of objects. Such types statically capture many common invariants that naturally arise in code. For instance, typically the shape of an array (the number of dimensions (the rank) and the size of each dimension) is determined at run time, but is fixed once the array is constructed. Thus, the shape of an array is part of its immutable state.

X10 provides a framework for specifying and checking constrained types that achieves certain desirable properties:

- **Ease of use.** The framework must be easy to use for practicing programmers. In particular, since X10 is an extension of Java, the syntax of types is a simple and natural extension of Java's types. Constrained types can interoperate smoothly with Java libraries.
- **Flexibility.** The framework permits the development of concrete, specific type systems tailored to the application area at hand. X10's compiler permits extension with different constraint systems via compiler plugins, enabling a kind of pluggable type system [9]. The framework is parametric in the kinds of expressions used in the type system, permitting the installed constraint system to interpret the constraints.
- **Modularity.** The rules for type-checking are specified once in a way that is independent of the particular vocabulary of operations used in the dependent type system. The type system supports separate compilation.
- **Static checking.** The framework permits mostly static type-checking. The user is able to escape the confines of static type-checking using dynamic casts, as is common for Java-like languages.

1.1 Constrained types

We permit the definition of a class C to specify a list of typed parameters, or *properties*, (T1 x1, ..., Tk xk) similar in syntactic structure to a method formal parameter list. Each property in this list is treated as a **public final** instance field. We also permit the specification of a *class invariant*, a *where clause* [14] in the class definition. A class invariant is a boolean expression on the properties of the class. The compiler ensures that all instances of the class created at run time satisfy the invariant. Syntactically, the class invariant is separated from the property list with a ":". For instance, we may specify a class List with an int length property as follows:

```
class List(int length: length >= 0) {...}
```

Given such a definition for a class C, types can be constructed by *constraining* the properties of C. In principle, *any* boolean expression over the properties specifies a

```
class List(int(:self >= 0) n) {
  Object head = null;
  List(n−1) tail = null;

  List(0)() { property(0); }
  List(1)(Object head) { this(head, new List());}
  List(tail.n+1)(Object head, List tail) {
    property(tail.n+1);
    this.head = head;
    this.tail = tail;
  }

  List(n+arg.n) append(final List arg) {
    return (n==0) ? arg : new List(head, tail.append(arg));
  }

  List(n) reverse() { return rev(new List()); }
  List(n+acc.n) rev(final List acc) {
    return (n==0) ? acc : tail.rev(new List(head, acc));
  }

  List(:self.n <= this.n) filter(Predicate f) {
    if (n==0) return this;
    List(:self.n <= this.n−1) l = tail.filter(f);
    return (f.isTrue(head)) ? new List(head,l) : l;
  }
}
```

Fig. 1. This program implements a mutable list of Objects. The size of a list does not change through its lifetime, even though at different points in time its head and tail might point to different structures.

type: the type of all instances of the class satisfying the boolean expression. Thus, List(:length == 3) is a permissible type, as are List(:length <= 41) and even List(:length * f() >= g()). In practice, e is restricted by the particular constraint system in use.

In general, a *constrained type* is of the form C(:e), the name of a class or interface[4] C, called the *base class*, followed by a *condition* e, a boolean expression on the properties of the base class and the **final** variables in scope at the type. Constraints specify (possibly) partial information about the variables of interest. Such a type represents a refinement of C: the set of all instances of C whose immutable state satisfies the condition e.

For brevity, we write C as a type as well; it corresponds to the (vacuously) constrained type C(:**true**). We also permit the syntax C(e1,. . . ,ek) for the type C(:x1==e1

[4] In X10, primitive types such as int are classes.

& ... & xk==ek) (assuming that the property list for C specifies the k properties x1, ..., xk, and each term ei is of the correct type).

Constrained types may occur wherever normal types occur. In particular, they may be used to specify the types of (possibly mutable) local variables, properties, (possibly mutable) fields, arguments to methods, return types of methods, in casts etc.

Using the definitions above, List(n), shown in Figure 1, is the type of all lists of length n. Intuitively, this definition states that a List has a int property n, which must be non-negative. The class has two fields that hold the head and tail of the list. The properties of the class are set through the invocation of **property**(...) (analogously to **super**(...)) in the constructors.

Our basic approach to introducing constrained types into X10 is to follow the spirit of generic types, but to use values instead of types.

Constructors have "return types" that can specify an invariant satisfied by the object being constructed. The compiler verifies that the constructor return type and the class invariant are implied by the **property** statement and any **super** calls in the constructor body. The List class has three constructors: the first constructor returns a empty list; the second returns a singleton list of length 1; the third returns a list of length m+1, where m is the length of the second argument.

In the third constructor, as well as the append and rev method, the return type depends on properties of the formal parameters. If an argument appears in a return type then the parameter must be declared **final**, ensuring the argument points to the same object throughout the evaluation of the constructor body. Parameters may also depend on other parameters in the argument list.

The use of constraints makes existential types very natural. Consider the return type of filter : it specifies that the list returned is of some unknown length. The only thing known about it is that its size is bounded by n. Thus, constrained types naturally subsume existential dependent types. Indeed, every base type C is an "existential" constrained type since it does not specify any constraint on its properties. Thus, code written with constrained types can interact seamlessly with legacy library code—using just base types wherever appropriate.

1.2 Constraint system plugins

The X10 compiler allows programmers to extend the semantics of the language with compiler plugins. Plugins may be used to support different constraint systems to be used in constrained types. Constraint systems provide code for checking consistency and entailment.

The condition of a constrained type is parsed and type-checked as a normal boolean expressions over properties and the **final** variables in scope at the type. Installed constraint systems translate the expression into an internal form, rejecting expressions that cannot be represented.

A given condition may be a conjunction of constraints from multiple constraint systems. A Nelson–Oppen procedure [30] is used to check consistency of the constraints.

The X10 compiler implements a simple equality-based constraint system. Constraint solver plugins have been implemented for Presburger constraints using the Omega

library [45] and the CVC3 solver [8] and a separate set-based constraint system built using CVC3. The implementation is discussed in Section 4.

1.3 Claims

The paper presents constrained types in the X10 programming language. We claim that the design is natural and easy to use and useful. Many example programs have been written using dependent types and are available at `x10.sf.net`.

As in staged languages [31, 51], the design distinguishes between compile-time and run-time evaluation. Constrained types are checked (mostly) at compile-time. The compiler uses a constraint solver to perform universal reasoning ("for all possible values of method parameters") for dependent type-checking. There is no run-time constraint-solving. However, run-time casts to dependent types are permitted; these casts involve arithmetic, not algebra—the values of all parameters are known.

The design supports separate compilation: a class needs to be recompiled only when it is modified or when the method and field signatures or invariants of classes on which it depends are modified.

We claim that the design is flexible. The language design is parametric on the constraint system being used. The compiler supports integration of different constraint solvers into the language. The design has been implemented in X10, available at `x10.sf.net`. Dependent clauses are also form the basis of a general user-definable annotation framework we have implemented separately [34].

Rest of this paper. **??** rewrite to agree with the rest of the paper

The next section reviews related work. Section 2 fleshes out the syntactic and semantic details of the proposal, and presents a formal semantics and a soundness theorem. Section 3 works through a number of examples using a variety of constraint systems. The implementation of constrained types in X10 is described in Section 4. Section 6 conclude the paper with a discussion of future work.

2 Constrained types

This section describes constrained types in X10.

2.1 Properties

A property is a **public final** instance field of the class that cannot be overridden by subclassing. Like any other field, a property is typed, and its type need not necessarily be primitive. Thus, properties capture the immutable public state of an object, initialized when the object is created, that can be classified by constrained types. Syntactically, properties are specified in a parameter list right after the name of the class in a class definition. The class body may contain specifications of other fields; these fields are considered mutable.

Properties may be of arbitrary type. For instance, the class region has an int property called rank. In turn, the class dist has a region property, called region, and also an

int property rank. The invariant for dist ensures that rank == region.rank. Similarly, an array has properties dist, region, and rank and appropriate constraints ensuring that the statically available information about them is consistent.[5] In this way, rich constraints on the immutable portion of the object reference graph, rooted at the current object and utilizing objects at user-defined types, may be specified.

2.2 Constraints

A constrained type is of the form C(:e), consisting of a *base class* C and a *condition* e, a boolean expresion on the properties of the base class and the **final** variables in scope at the type. Constraints specify (possibly) partial information about the variables of interest. Such a type represents a refinement of C—the set of all instances of C whose immutable state satisfies the condition c.

Constraints may use the special variable **self** to stand for the object whose type is being defined. Thus, int (: **self** >= 0) is the set of natural numbers, and point (:x*x + y*y <= 1.0) represents the interior of a circle (for a class point with two float properties x and y). The type C(: **self** != **null**) represents all instances of C. When there is no ambiguity, a property reference **self**.x may be abbreviated to x. The type int (: **self**==v) represents a "singleton" type, an int is of this type only if it has the same value as v.

To be clear, **self** is not the same as **this**. In the List example of Figure 1, a list a list with type List (: **self**.n <= **this**.n) is returned by the filter method: **self**.n is the length of the returned List; **this**.n is the length of the receiver of the call to filter.

Constraints are specified in terms of an underlying constraint system [48]—a predefined logical vocabulary of functions and predicates with algorithms for consistency and entailment. The X10 compiler permits different constraint systems to be installed using compiler plugins. [**?**]. Constraint system plugins define a language of constraints by symbolically interpreting the boolean expression specifying a type's condition; plugins may report an error if the condition cannot be interpreted.

In this framework, types may be constrained by any boolean expression over the properties. For practical reasons, restrictions need to be imposed to ensure constraint checking is decidable.

The condition of a constrained type must be a pure function only of the properties of the base class. Because properties are **final** instance fields of the object, this requirement ensures that whether or not an object belongs to a constrained type does not depend on the *mutable* state of the object. That is, the status of the predicate "this object belongs to this type" does not change over the lifetime of the object. Second, by insisting that each property be a *field* of the object, the question of whether an object is of a given type can be determined merely by examining the state of the object and evaluating a boolean expression. Of course, an implementation is free to not *explicitly* allocate memory in the object for such fields. For instance, it may use some scheme of tagged pointers to implicitly encode the values of these fields.

[5] All constraint languages used in constrained types permit object references, field selection and equality. Such constraint systems have been studied extensively under the name of "feature structures" [2].

Further, by requiring that the programmer distinguish certain **final** fields of a class as properties, we ensure that the programmer consciously controls *which* **final** fields should be available for constructing constrained types. A field that is "accidentally" **final** may not be used in the construction of a constrained type. It must be declared as a property.

2.3 Subtyping

Constrained types naturally come equipped with a subtype relation that combines the nominal subtyping relation of classes and interfaces with the logical entailment relation of the constraint system. Namely, a constraint $\mathsf{C}(:\mathsf{c})$ is a subtype of $\mathsf{D}(:\mathsf{d})$ if C is a subtype of D and every value in C that satisfies c also satisfies d.

This definition implies that $\mathsf{C}(:\mathsf{e1})$ is a subtype of $\mathsf{C}(:\mathsf{e2})$ if $\mathsf{e1}$ implies $\mathsf{e2}$. In particular, for all conditions e, $\mathsf{C}(:\mathsf{e})$ is a subtype of C. $\mathsf{C}(:\mathsf{e})$ is empty exactly when e conjoined with C's class invariant is inconsistent.

Two constrained types $\mathsf{C1}(:\mathsf{e1})$ and $\mathsf{C2}(:\mathsf{e2})$ are considered equivalent if $\mathsf{C1}$ and $\mathsf{C2}$ are the same base type and $\mathsf{e1}$ and $\mathsf{e2}$ are equivalent when considered as logical expressions. Thus, $\mathsf{C}(:\mathsf{x}{*}\mathsf{x}{==}4)$ and $\mathsf{C}(:\mathsf{x}{==}2 \;||\; \mathsf{x}{==}{-}2)$ are equivalent types.

2.4 Final variables

The use of **final** local variables, formal parameters, and fields in constrained types has proven to be particularly valuable in practice. The same variable that is being used in computation can also be used to specify types. There is no need to introduce separate, universally and existentially quantified "index" variables. During type-checking, **final** variables are turned into symbolic variables—some fixed but unknown value—of the same type. Computation is performed in a constraint-based fashion on such variables.

2.5 Method and constructor preconditions

Methods and constructors may specify preconditions on their parameters as where clauses. For an invocation of a method (or constructor) to be type-correct, the associated where clause must be statically known to be satisfied. Note that the where clause may contain constraints on the properties of **this**. Thus the where clause may be used to specify that a method is *conditionally* available on some objects of the class and not others.

The return type of a method may also contain expressions involving the arguments to the method. Any argument used in this way must be declared **final**, ensuring it is not mutated by the method body. For instance:

$$\mathsf{List}(\mathsf{arg.length}{-}1)\ \mathsf{tail}(\textbf{final}\ \mathsf{List}\ \mathsf{arg} : \mathsf{arg.length} > 0)\ \{\ldots\}$$

is a valid method declaration. It says that `tail` must be passed a non-empty list, and it returns a list whose length is one less than its argument.

2.6 Method overloading and overriding

The definitions of method overloading, overriding, hiding, shadowing and obscuring in X10 are the same as in Java [21], modulo the following considerations motivated by dependent types.

Our current implementation erases dependent type information when compiling to Java. Therefore it must be the case that a class does not have two different method definitions that conflict with each other when the constrained clauses in their types are erased.

A class C inherits from its direct superclass and superinterfaces all their methods that are visible according to the access modifiers and that are not hidden or overridden. A method m1 in a class C overrides a method m2 in a superclass D if m1 and m2 have signatures with equivalent (unerased) formal parameter types.

Dynamic method lookup does not take dependent type information into account, only the class hierarchy. This design decision ensures that serious errors such as method invocation errors are captured at compile-time. (Such errors can arise because multiple incomparable methods with the same name and acceptable argument lists might be available at the dynamic dependent type of the subject. Examples are not difficult to construct.)

2.7 Constructors for dependent classes

Constructors must ensure that the class invariants of the given class and its superclasses and superinterfaces hold. For instance, the nullary constructor for List ensures that the property length has the value 0:

```
public List(0)() { property(0); }
```

The **property** statement is used to set all the properties of the new object simultaneously. Capturing this assignment in a single statement simplifies checking that the constructor postcondition and class invariant are established. If a class has properties, every path through the constructor must contain exactly one **property** statement.

Java-like languages permit constructors to throw exceptions. This is necessary to deal with the situation in which the arguments to a constructor for a class C are such that no object can be constructed that satisfies the invariants for C. Dependent types make it possible to perform some of these checks at compile-time. The class invariant of a class explicitly captures conditions on the properties of the class that must be satisfied by any instance of the class. Constructor preconditions capture conditions on the constructor arguments. The compiler's static check for non-emptiness of the type of any variable captures these invariant violations at compile-time.

2.8 Extending dependent classes

A class may extend a constrained class, e.g., **class** C(...) **extends** D(:d). This documents the programmer's intention that every call to **super** in a constructor for C must ensure that the invariant d is established on the state of the class D. The expressions in the actual parameter list for the super class may involve only the properties of the class being defined.

2.9 Dependent interfaces

Java does not allow interfaces to specify instance fields. Rather all fields in an interface are final static fields (constants). However, since properties play a central role in the specification of refinements of a type, it makes sense to permit interfaces to specify properties. Similarly, an interface definition may specify an invariant on its properties. Methods in the body of an interface may have where clauses as well.

All classes implementing an interface must have a property with the same name and type (either declared in the class or inherited from the superclass) for each property in the interface. If a class implements multiple interfaces and more than one of them specify a property with the same name, then they must all agree on the type of the property. The class must have a single property with the given name and type.

The general form of a class declaration is now:

```
class C(T1 x1, ..., Tk xk)
      extends B(:c)
      implements I1(:c1), ..., In(:cn) {...}
```

For such a declaration to type-check, it must be that the class invariant of C implies *inv*(Ii) & ci, where *inv*(Ii) is the invariant associated with interface Ii . Again, a constrained class or interface I is taken as shorthand for I (: **true**). Further, every method specified in the interface must have a corresponding method in the class with the same signature whose precondition, if any, is implied by the precondition of the method in the interface.

2.10 Separation between compile-time and run-time computation

Our design distinguishes between compile-time execution (performed during type-checking) and run-time execution. At compile-time, the compiler processes the abstract syntax tree of the program generating queries to the constraint solver. The only computation engine running is the constraintnsolver, which operates on its own vocabulary of predicates and functions. Program variables (such as local variables) that occur in types are dealt with symbolically. They are replaced with logical variables—some fixed, but unknown value—of the same type. The constraint solver must know how to process pieces of partial information about these logical variables in order to determine whether some constraint is entailed. At run time, the same program variable will have a concrete value and will perform "arithmetic" (calculations) where the compiler performed "algebra" (symbolic analysis).

Constrained types may occur in a class cast (T) o. Code is generated to check at run time that o satisfies T.

2.11 Query evaluation

Because object-oriented languages permit arbitrary mutual recursion between classes: classes A and B may have fields of type B and A respectively—the type/property graph may have loops. The nodes in this graph are base types (class and interface names). There is an edge from node A to node B if A has a property whose base type is B.

Let us define the real-clause of a constrained type $\mathsf{C}(:\mathsf{c})$ to be the set of constraints that must be satisfied by any instance of $\mathsf{C}(:\mathsf{c})$. This includes c but also includes constraints that hold for all instances of C, as determined by the definition of C. Let us use the notation $rc(\mathsf{C})$ for the *real clause* of C. Since we consider only top-level classes, the only free variable in $rc(\mathsf{C})$ is **self**.

?? not clear, esp. what X is

What is $rc(\mathsf{C},\mathtt{X})$ (we have drawn out X as the formal variable)? Consider a general class definition:

```
class C(T1 x1, ..., Tk xk: c) extends D(:d) { ... }
```

Clearly, from this we get:

$$rc(\mathsf{C},\mathtt{X}) \iff (\mathsf{c}\wedge\mathsf{d})[\mathtt{X}/\mathtt{this}]\wedge rc(\mathtt{D},\mathtt{X})\wedge rc(\mathtt{P}_1,\mathtt{X}.\mathtt{x}_1)\wedge\cdots\wedge rc(\mathtt{P}_k,\mathtt{X}.\mathtt{x}_k)$$

That is, given a program P with classes $\mathsf{C}_1,\ldots,\mathsf{C}_k$, the set of real-clauses for $\mathsf{C}_1,\ldots,\mathsf{C}_k$ are defined in a mutually recursive fashion through the Clark completion of a Horn clause theory (over an underlying constraint system).

The central algorithmic question now becomes whether given a constrained clause d, does $rc(\mathsf{C},\mathtt{X})$ entail d?

From the above formulation the question is clearly semi-decidable. It is not clear however whether it is decidable. This is a direction for further work.

In practice, many data-structures have non-cyclic dependency graphs. For such programs the real-clause can be computed quickly and only a bounded number of questions to the constraint-solver are generated during type-checking.

2.12 Parametric consistency

Consider the set of final variables that are referenced in a type $\mathsf{T} = \mathsf{C}(:\mathsf{c})$. These are the *parameters* of the type. A type is said to be *parametrically consistent* if its where clause is solvable for each possible assignment of values to parameters. A parametrically consistent type has the property that its extension will never be empty.

Types are required to be parametrically consistent.

?? syntax of tail's type not explained

Consider a variation of List from Figure 1:

```
class List(int(:self >= 0) n) {
  Object head;
  List(:self.n == this.n−1 & self != null) tail;
  ...
}
```

The type of the field tail is not parametrically consistent. There exists a value for its parameter n, namely, 0 for which the real clause **self** != **null** & **self**.n == **this**.n−1 & **self**.n >= 0 is not satisfiable.

The compiler will throw a type error when it encounters the initializer for this field in a constructor since it will not be able to prove that the initial value is of the given type.

3 Examples

The following section presents example uses of constrained types using several different constraint systems. Many common object-oriented idioms and object-oriented type systems can be captured naturally using constrained types: specifically we discuss types for places, aliases, ownership, arrays and clocks. We have implemented the type system

?? Many of these constraint systems are more expressive than the constraint system implemented in the current X10 compiler and have not (yet) been implemented.

?? In the following we will use the shorthand $\mathtt{C}(\overline{\mathtt{t}} : \mathtt{c})$ for the type $\mathtt{C}(: \overline{\mathtt{f}} = \overline{\mathtt{t}}, \mathtt{c})$ where the declaration of the class C is $\mathtt{class\ C}(\overline{\mathtt{T}}\ \overline{\mathtt{f}} : \mathtt{c}) \ldots$ Also, we abbreviate $\mathtt{C}(\overline{\mathtt{t}} : \mathtt{true})$ as $\mathtt{C}(\overline{\mathtt{t}})$. Finally, we use the shorthand $\mathtt{T\ x} = \mathtt{t};\ \mathtt{c}$ for the constraint $\mathtt{T\ x};\ \mathtt{x} = \mathtt{t};\ \mathtt{c}$.

3.1 Presburger constraints: blocked LU factorization

3.2 Set constraints: region-based arrays

X10 takes another approach to ensuring array bounds violations do no occur. Following ZPL [10], arrays in X10 are defined over sets of n-dimensional *index points* called *regions* [22]. For instance, the region [0:200,1:100] specifies a collection of two-dimensional points (i,j) with i ranging from 0 to 200 and j ranging from 1 to 100.

Constrained types ensure array bounds violations do not occur: an array access type-checks if the index point can be statically determined to be in the region over which the array is defined.

Region constraints are subset constraints and have the following syntax:

(Constraint) c ::= r ⊆ r | ...
(Region) r ::= t | [$\mathtt{b}_1$:$\mathtt{d}_1$,...,$\mathtt{b}_k$:$\mathtt{d}_k$] | r|r | r&r | r-r | r+p
(Point) p ::= t | [$\mathtt{b}_1, \ldots, \mathtt{b}_k$]
(Integer) b,d ::= t | n

Regions used in constraints are either constraint terms t, region constants, unions (|), intersections (&), or differences (-), or regions where each point is offset by another point p.

For example, the code in Figure 2 performs a successive over-relaxation [44] of an $n \times n$ matrix G. The type-checker establishes that the `region` property of the point `ij` (line 17) is `inner & [i:i,d1min:d1max]`, and that this region is a subset of `outer`, the region of the array G.

3.3 AVL trees and red–black trees

AVL trees and red-black trees can be modeled so that the data structure invariant is enforced statically.

```
class AVLTree(int(:self >= 0) height) {...}
class Leaf(Object key) extends AVLTree(0) {...}
class Node(Object key, AVLTree l, AVLTree r
           : int d=l.height−r.height; −1 <= d, d <= 1)
    extends AVLTree(max(l.height,r.height)+1) {...}
```

```
point NORTH = point.factory.point(1,0);
point WEST = point.factory.point(0,1);
void sor(double omega, double[.] G, int iter) {
  region(:self==G.region) outer = G.region;
  region(:outer.contains(self)) inner =
    outer & (outer−NORTH) & (outer+NORTH)
          & (outer−WEST) & (outer+WEST);
  region d0 = inner.project(0);
  region d1 = inner.project(1);
  if (d1.size() == 0) return;
  int d1min = d1.min()[0];
  int d1max = d1.max()[0];
  for (point[off] : [1:iter∗2]) {
    int red_black = off % 2;
    finish foreach (point[i]: d0) {
      if (i % 2 == red_black) {
        for (point ij: inner & [i:i,d1min:d1max]) {
          G[ij] = omega / 4.
                ∗ (G[ij−NORTH] + G[ij+NORTH]
                  + G[ij−WEST] + G[ij+WEST])
                ∗ (1. − omega) ∗ G[ij];
        }
      }
    }
  }
}
```

Fig. 2. Successive over-relaxation with regions

Red–black trees may be modeled similarly. Such trees have the invariant that (a) all leaves are black, (b) each non-leaf node has the same number of black nodes on every path to a leaf (the black height), (c) the immediate children of every red node are black.

```
class RBTree(int blackHeight) {...}
class Leaf extends RBTree(0) { int value; ... }
class Node(boolean isBlack,
           RBTree(:this.isBlack || isBlack) l,
           RBTree(:this.isBlack || isBlack,
                  blackHeight=l.blackHeight) r)
    extends RBTree(l.blackHeight+1) { int value; ... }
```

3.4 Place types

This example is due to Satish Chandra. We wish to specify a balanced distributed tree with the property that its right child is always at the same place as its parent, and once the left child is at the same place then the entire subtree is at that place. In X10, every object has a field `location` of type `place` that specifies the location at which the object is located. The desired property may be specified thus:

```
class Tree(boolean localLeft) extends Object {
  Tree(:! this.localLeft || (location==here && self.localLeft)) left;
  Tree(:location==here) right);
  ...
}
```

The constraint on left states that if the localLeft property is true, then the location of the left child must be **here** and its localLeft property must be set. This ensures, recursively, that the entire left subtree will be located at the same place.

3.5 Ownership constraints

Figure 3 shows a fragment of a List class, demonstrating how ownership types [12] can be encoded as constrained types using a simple extension of X10's built-in equality constraint system.

Each Owned object has an owner property. Objects also have properties that are used as owner parameters. The List class has an owner property inherited from Owned and also declares a valOwner property that is instantiated with the owner of the values in the list, stored in the head field of each element. The tail of the list is owned by the list object itself.

To encode the "owners as dominators" property, the owner of the values valOwner must be contained within the owner of the list itself; that is, valOwner must be owner or valOwner's owner must be contained in owner. This is captured by the constraint **self**.owns(owner) on valOwner. Calls to the owns method in constraints are interpreted by the ownership constraint solver as the disjunction of conditions shown in the body of owns. The object world is the root of the ownership tree; all objects are transitively owned by world.

```
class Owned(Owned owner) {
    boolean owns(Owned o) {
        return this == world || this == o.owner || this.owns(o.owner);
    }

    Owned(:owner==o)(final Owned o) { property(o); }

    static final Owned(null) world = new Owned(null);
}

class List(Owned(:owns(owner)) valOwner)
    extends Owned
{
    Owned(:owner==valOwner) head;
    List(:owner==this & valOwner==this.valOwner) tail;

    List(:owner==o & valOwner==v)(Owned o, Owned v: o.owns(v)) {
        super(o);
        property(v);
    }

    List(:owner==this & valOwner==this.valOwner) tail() {
        return tail;
    }

    ...
}
```

Fig. 3. Ownership types

For example, the type List (:owner==world & valOwner == **this**) is invalid, because its constraint is interpreted as owner == world & valOwner == **this** & **this**.owns(world), which is satisfiable only when **this** == world (which it is not).

An additional check is needed to ensure objects owned by **this** are encapsulated. The tail () method for instance, incorrecty leaks the list's tail field. To eliminate this case, the ownership constraint system must additionally check that owner parameters are bound only to **this**, world, or an owner property of **this**. These conditions ensure that tail () can be called only on **this** because its return type is otherwise not valid. For instance, in the following code, the type of ys is not valid because the owner property is bound to xs:

```
final Owned o = ...;
final List(:owner==o & valOwner==o) xs;
List(:owner==xs & valOwner==o) ys = xs.tail();
```

3.6 Disequalities: non-null types

A constraint system that supports disequalities can be used to enforce a non-null invariant on reference types. A non-null type C can be written simply as C(:**self** != **null**).

4 Implementation

The X10 compiler provides a framework for writing and checking constrained types. The X10 language, constraints in X10 are conjunctions of equalities over immutable side-effect-free expressions. Compiler plugins may be installed that support other constraint languages, Presburger constraints (linear inequalities) [**?**].

The X10 compiler is implemented as an extension of Java using the Polyglot compiler framework [32]. Expressions used in constrained types are type-checked as normal non-dependent X10 expressions; no constraint solving is performed on these expressions. During type-checking, constraints are generated and solved using the built-in constraint solver or using solvers provided by plugins. The system allows types to constrained by conjunctions of constraints in different constraint languages. If constraints cannot be solved, an error is reported.

4.1 Constraint checking

After type-checking a constraint as a boolean expression e, the abstract syntax tree for the boolean expression is transformed into a list of conjuncts. e1 & ... & ek. Each conjunct ei is given to the installed constraint system plugins, which symbolically evaluate the expression to create an internal representation of the conjunct. If no constraint system can handle the conjunct, an error is reported.

To interoperate, the constraint solvers must share a common vocabulary: constraint terms t range over the properties of the base type, the final variables in scope at the type (including **this**), the special variable **self** representing the a value of the type, and field selections t.f. All constraint systems are required to support the trivial constraint **true**, conjunction, existential quantification and equality on constraint terms.

In this form, the constraint is represented as a conjunction of constraints from different theories. Constraints are checked for satisfiability using a Nelson–Oppen procedure [30]. After constructing a constraint-system specific representation of a conjunct, each plugin computes the set of term equalities entailed by the conjunct. These equalities are propagated to the other conjuncts, which are again checked for satisfiability and any new equalities generated are propagated. If a conjunct is found to be unsatisfiable, an error is reported.

During type-checking, the type checker needs to determine if the type C(:c) is a subtype of D(:d). This is true if the base type C is a subtype of D and if the constraint c entails d.

To check entailment, each constraint solver is asked if a given conjunct of d is entailed by c. If any report false, the entailment does not hold and the subtyping check fails.

4.2 Translation

After constraint-checking, the X10 code is translated to Java in a straightforward manner. Each dependent class is translated into a single class of the same name without dependent types. The explicit properties of the dependent class are translated into **public final** (instance) fields of the target class. A **property** statement in a constructor is translated to a sequence of assignments to initialize the property fields.

For each property, a getter method is also generated in the target Java class. Properties declared in interfaces are translated into getter method signatures. Subclasses implementing these interfaces thus provide the required properties by implementing the generated interfaces.

Usually, constrained types are simply translated to non-constrained types by erasure; constraints are checked statically and need no run-time representation. However, dependent types may be used in casts and **instanceof** expressions. Because the constraint syntax in X10 is a subset of the X10 expression syntax, run-time tests of constrained types are translated to Java by evaluating the constraint with **self** bound to the expression being tested. For examples, casts are translated as:

```
⟦(C(:c) e⟧ =
  new Object() {
    C cast(C self) {
      if (⟦c⟧) return self;
      throw new ClassCastException(); }
  }.cast((C) ⟦e⟧)
```

Wrapping the evaluation of c in an anonymous class ensures the expression e is evaluated only once.

To support separate compilation, abstract syntax trees for constraints are embedded into the generated Java code, and from there into the generated class file. The compiler reconstructs dependent types in referenced class files from their ASTs.

5 Related work

Constraint-based type systems enjoy a long history. Mitchell [29] and Reynolds [46] developed the use of constraints for type inference and subtyping. Trifonov and Smith [52] proposed a type system where types are refined by subtyping constraints. Dependent types are not supported. Pottier [40, 42] presents a constraint-based type system for an ML-like language with subtyping.

HM(X) [50, 41, 43] is a constraint-based framework for Hindley–Milner style type systems. The framework is parameterized on the specific constraint system X; instantiating X yields extensions of the HM type system. Constraints in HM(X) are over types, not values.

Several systems have been proposed that refine types in a base type system through constraints. *Refinement types* [20] extend the Hindley–Milner type system with intersection, union, and constructor types, enabling specification and inference of more precise type information. *Conditional types* [1] extend refinement types to encode control-flow information in the types. Jones introduced *qualified types*, which permit types to be constrained by a finite set of predicates [25]. *Sized types* [23] annotate types with the sizes of recursive data structures. Sizes are linear functions of size variables. Size inference is performed using a constraint solver for Presburger arithmetic [45].

Our work is most closely related to DML, the extension of ML with dependent types. We discuss this in detail in the next section.

With hybrid type-checking [16, 17], types can be constrained by arbitrary boolean expressions. While typing is undecidable, dynamic checks are inserted into the program when necessary if the type-checker cannot determine type safety statically. In X10 dynamic type checks, including tests of dependent clauses, are inserted only at explicit casts.

Singleton types [5, 49] are dependent types containing only one value. In Stone's formulation [49], $S(e : \tau)$ is the type of all values of type τ that are equal to e. Term equivalence is constructed so that type-checking is decidable. The singleton $S(e : \tau)$ can be encoded in CFJ as $\tau(:\texttt{self}{=}\lfloor e \rfloor)$, where $\lfloor e \rfloor$ lifts e to a constraint term.

Several languages—gbeta [15], Scala [35, 38], J& [33] and others [37, 36]—provide *path-dependent types*. For a final access path `p`, `p.type` in Scala is the singleton type containing the object `p`. In J& `p.class` is a type containing all objects whose runtime class is the same as `p`'s. Scala's `p.type` can be encoded in CFJ using an equality constraint `C(:self == p)`, where `C` is a supertype of `p`'s static type.

Cayenne [7] is a Haskell-like language with fully dependent types. There is no distinction between static and dynamic types. Type-checking is undecidable. There is no notion of datatype refinement as in DML.

Epigram [28, 3] is a dependently typed functional programming language based on a type theory with inductive families. Epigram does not have a phase distinction between values and types.

ESC/Java [18] allow programmers to write object invariants and pre- and post-conditions that are enforced statically by the compiler using an automated theorem prover. Static checking is undecidable and, in the presence of loops, is unsound (but still useful) unless the programmer supplies loop invariants. ESC/Java can enforce invariants on mutable state.

Pluggable and optional type systems were proposed by Bracha [9] and provide another means of specifying refinement types. Type annotations, implemented in compiler plugins, serve only to reject programs statically that might otherwise have dynamic type errors. CQual [19] extends C with user-defined type qualifiers. These qualifiers may be flow-sensitive and may be inferred. CQual supports only a fixed set of typing rules for all qualifiers. In contrast, the *semantic type qualifiers* of Chin, Markstrum, and Millstein [11] allow programmers to define typing rules for qualifiers in a meta language that allows type-checking rules to be specified declaratively. JavaCOP [4] is a pluggable type system framework for Java. Annotations are defined in a meta language that allows type-checking rules to be specified declaratively. JSR 308 [26] is a proposal for adding user-defined type qualifiers to Java.

DML Our work is most closely related to DML, [53], an extension of ML with dependent types. DML is also built parametrically on a constraint solver. Types are refinement types; they do not affect the operational semantics and erasing the constraints yields a legal ML program.

At a conceptual level the intuitions behind the development of DML and constrained types are similar. Both are intended for practical programming by mainstream programmers, both introduce a strict separation between compile-time and run-time processing, are parametric on a constraint solver, and deal with mutually recursive data-structures, mutable state, and higher-order functions (encoded as objects in the case of constrained types). Both support existential types.

The most obvious distinction between the two lies in the target domain: DML is designed for functional programming, specifically ML, whereas constrained types are designed for imperative, concurrent OO languages. Hence technically our development of constrained types takes the route of an extension to FJ. But there are several other crucial differences as well.

First, DML achieves its separation by not permitting program variables to be used in types. Instead, a parallel set of (universally or existentially quantified) "index" variables are introduced. For instance the typing of the `append` operation on lists is provided by:

```
fun('a)
  append(nil, ys) = ys
| append(cons(x, xs), ys) = cons(x, append(xs,ys))
where append <| {m:nat}{n:nat}
            'a list(m) * 'a list(n) -> 'a list(m+n)
```

in contrast with the direct embedded expression with constrained types:

```
class List(int(:self >= 0) n) {
   Object item;
   List(n−1) tail;
   List(n+a.n) append(final List a) {
     return n==0 ? a : new List(item, tail.app(a)); }
   ...
}
```

Second, DML permits only variables of basic index sorts known to the constraint solver (e.g., `bool`, `int`, `nat`) to occur in types. In contrast, constrained types permit pro-

gram variables at any type to occur in constrained types. As with DML, only operations specified by the constraint system are permitted in types. However, these operations always include field selection and equality on object references. (As we have seen permitting arbitrary type/property graphs may lead to undecidability.) Note that DML style constraints are easily encoded in constrained types.

Third, DML does not permit any runtime checking of constraints (dynamic casts).

6 Conclusion and Future work

We have presented a simple design for constrained types in Java-like languages. The design considerably enriches the space of statically checkable types expressible in the language. This is particularly important for data-structures such as lists and arrays. We have formalized constrained types in a sound extension of FJ, Constrained FJ. Several examples of constrained types were presented. Constrained types have been implemented in X10 and used for place types, clocked types, and array types.

In future work, we plan to investigate optimizations (such as array bounds check elimination) enabled by constrained types. We also plan to explore type inference for constrained types and to pursue more expressive constraint systems and extensions of constrained types for handling mutable state, control flow, and effects.

References

1. Alexander Aiken, Edward L. Wimmers, and T. K. Lakshman. Soft typing with conditional types. In *Proceedings of the 21st Annual ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL)*, pages 163–173, January 1994.
2. Hassan Ait-Kaci. *A lattice theoretic approach to computation based on a calculus of partially ordered type structures (property inheritance, semantic nets, graph unification)*. PhD thesis, University of Pennsylvania, 1984.
3. Thorsten Altenkirch, Conor McBride, and James McKinna. Why dependent types matter. `http://www.e-pig.org/downloads/ydtm.pdf`, April 2005.
4. Chris Andreae, James Noble, Shane Markstrum, and Todd Millstein. A framework for implementing pluggable type systems. In *Proceedings of the 2006 ACM Conference on Object Oriented Programming Systems, Languages, and Applications (OOPSLA)*, October 2006.
5. David Aspinall. Subtyping with singleton types. In *CSL '94: Selected Papers from the 8th International Workshop on Computer Science Logic*, volume 933 of *LNCS*, pages 1–15, London, UK, 1995. Springer-Verlag.
6. David Aspinall and Martin Hofmann. *Dependent Types*, chapter 2. In Pierce [39], 2004.
7. Lennart Augustsson. Cayenne: a language with dependent types. In *Proceedings of the ACM SIGPLAN International Conference on Functional Programming (ICFP '98)*, pages 239–250, 1998.
8. Clark Barrett and Sergey Berezin. CVC Lite: A new implementation of the cooperating validity checker. In Rajeev Alur and Doron A. Peled, editors, *Proceedings of the* 16^{th} *International Conference on Computer Aided Verification (CAV '04)*, volume 3114 of *Lecture Notes in Computer Science*, pages 515–518. Springer-Verlag, July 2004. Boston, Massachusetts.
9. Gilad Bracha. Pluggable type systems. In *OOPSLA'04 Workshop on Revival of Dynamic Languages*, October 2004.

10. Bradford L. Chamberlain, Sung-Eun Choi, Steven J. Deitz, and Lawrence Snyder. The high-level parallel language ZPL improves productivity and performance. In *Proceedings of the IEEE International Workshop on Productivity and Performance in High-End Computing*, 2004.
11. Brian Chin, Shane Markstrum, and Todd Millstein. Semantic type qualifiers. In *Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*, pages 85–95, 2005.
12. D. Clarke, J. Noble, and J. Potter. Simple ownership types for object containment. In *ECOOP*, 2001.
13. Thierry Coquand and Gerard Huet. The Calculus of Constructions. *Information and Computation*, 76, 1988.
14. Mark Day, Robert Gruber, Barbara Liskov, and Andrew C. Myers. Subtypes vs. where clauses: Constraining parametric polymorphism. In *Proceedings of the 1995 ACM Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA)*, pages 156–168, Austin, TX, October 1995.
15. Erik Ernst. *gbeta: A Language with Virtual Attributes, Block Structure, and Propagating, Dynamic Inheritance*. PhD thesis, Department of Computer Science, University of Aarhus, Århus, Denmark, 1999.
16. Cormac Flanagan. Hybrid type checking. In *Proceedings of the 33rd Annual Symposium on Principles of Programming Languages (POPL'06)*, pages 245–256, 2006.
17. Cormac Flanagan, Stephen N. Freund, and Aaron Tomb. Hybrid types, invariants, and refinements for imperative objects. In *International Workshop on Foundations of Object-Oriented Programming (FOOL)*, 2006.
18. Cormac Flanagan, K. Rustan M. Leino, Mark Lillibridge, Greg Nelson, James B. Saxe, and Raymie Stata. Extended static checking for Java. In *Proceedings of ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*, June 2002.
19. Jeffrey S. Foster, Tachio Terauchi, and Alex Aiken. Flow-sensitive type qualifiers. In *Proceedings of the 29th ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL)*, pages 1–12. ACM Press, June 2002.
20. Tim Freeman and Frank Pfenning. Refinement types for ML. In *Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*, pages 268–277, June 1991.
21. J. Gosling, W. Joy, G. Steele, and G. Bracha. *The Java Language Specification, Third Edition*. Addison Wesley, 2006.
22. Christian Grothoff, Jens Palsberg, and Vijay Saraswat. Safe arrays via regions and dependent types. Technical Report RC23911, IBM T.J. Watson Research Center, 2006.
23. John Hughes, Lars Pareto, and Amr Sabry. Proving the correctness of reactive systems using sized types. In *Proceedings of the 23rd ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL)*, pages 410–423, 1996.
24. A. Igarashi, B. Pierce, and P. Wadler. Featherweight Java: A minimal core calculus for Java and GJ. In *ACM Symposium on Object-Oriented Programming: Systems, Languages and Applications*, 1999.
25. Mark P. Jones. *Qualified Types: Theory and Practice*. Cambridge University Press, 1994.
26. JSR 308: Annotations on Java types. `http://jcp.org/en/jsr/detail?id=308`.
27. Per Martin-Löf. *A Theory of Types*. 1971.
28. Conor McBride and James McKinna. The view from the left. *Journal of Functional Programming*, 14(1):69–111, 2004.
29. John C. Mitchell. Coercion and type inference. In *Proceedings of the 11th Annual ACM Symposium on Principles of Programming Languages (POPL'84)*, pages 174–185, 1984.
30. G. Nelson and D. Oppen. Simplification by cooperating decision procedures. *ACM Transactions on Programming Languages and Systems*, 1(2), October 1979.

31. Flemming Nielson and Hanne Riis Nielson. *Two-level functional languages*. Cambridge University Press, 1992.
32. Nathaniel Nystrom, Michael R. Clarkson, and Andrew C. Myers. Polyglot: An extensible compiler framework for Java. In Görel Hedin, editor, *Compiler Construction, 12th International Conference, CC 2003*, number 2622 in LNCS, pages 138–152. Springer-Verlag, April 2003.
33. Nathaniel Nystrom, Xin Qi, and Andrew C. Myers. J&: Nested intersection for scalable software extension. In *Proceedings of the 2006 ACM Conference on Object Oriented Programming Systems, Languages, and Applications (OOPSLA)*, pages 21–36, Portland, OR, October 2006.
34. Nathaniel Nystrom and Vijay Saraswat. An annotation and compiler plugin system for X10. Technical Report RC24198, IBM T.J. Watson Research Center, 2007.
35. Martin Odersky, Philippe Altherr, Vincent Cremet, Burak Emir, Sebastian Maneth, Stéphane Micheloud, Nikolay Mihaylov, Michel Schinz, Erik Stenman, and Matthias Zenger. An overview of the Scala programming language. Technical report, École Polytechnique Fédérale de Lausanne, June 2004. `http://scala.epfl.ch/docu/files/ScalaOverview.pdf`.
36. Martin Odersky, Vincent Cremet, Christine Röckl, and Matthias Zenger. A nominal theory of objects with dependent types. In *Proceedings of 17th European Conference on Object-Oriented Programming (ECOOP 2003)*, volume 2743 of *LNCS*, pages 201–224. Springer-Verlag, July 2003.
37. Martin Odersky and Christoph Zenger. Nested types. In *8th Workshop on Foundations of Object-Oriented Languages (FOOL)*, 2001.
38. Martin Odersky and Matthias Zenger. Scalable component abstractions. In *OOPSLA05*, pages 41–57, San Diego, CA, USA, October 2005.
39. Benjamin C. Pierce, editor. *Advanced Topics in Types and Programming Languages*. MIT Press, 2004.
40. François Pottier. Simplifying subtyping constraints. In *Proceedings of the ACM SIGPLAN International Conference on Functional Programming (ICFP '96)*, pages 122–133, 1996.
41. François Pottier. A semi-syntactic soundness proof for HM(X). Technical Report RR 4150, INRIA, March 2001.
42. François Pottier. Simplifying subtyping constraints, a theory. *Information and Computation*, 170(2):153–183, November 2001.
43. François Pottier and Didier Rémy. *The Essence of ML Type Inference*, chapter 10. In Pierce [39], 2004.
44. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, pages 866–869. Cambridge University Press, 1992. Successive overrelaxation (SOR).
45. William Pugh. The omega test: A fast and practical integer programming algorithm for dependence analysis. In *Supercomputing '91: Proceedings of the 1991 ACM/IEEE Conference on Supercomputing*, pages 4–13, 1991.
46. John C. Reynolds. Three approaches to type structure. In *Proceedings of TAPSOFT/CAAP 1985*, volume 185 of *LNCS*, pages 97–138. Springer-Verlag, 1985.
47. V. Saraswat et al. Report on the programming language X10. Technical report, IBM T.J. Watson Research Center, 2006.
48. Vijay Saraswat. The category of constraint systems is Cartesian closed. In *LICS '92*, pages 341–345, 1992.
49. Christopher A. Stone. *Singleton Types and Singleton Kinds*. PhD thesis, Carnegie–Mellon University, August 2000. Also availalble as CMU technical report CMU-CS-00-153.

50. Martin Sulzmann, Martin Odersky, and Martin Wehr. Type inference with constrained types. In *Fourth International Workshop on Foundations of Object-Oriented Programming (FOOL 4)*, 1997.
51. Walid Taha and Tim Sheard. Multi-stage programming with explicit annotations. In *ACM/SIGPLAN Workshop on Partial Evaluation and Semantics-Based Program Manipulation*, pages 203–217, 1997.
52. Valery Trifonov and Scott Smith. Subtyping constrained types. In *Third International Static Analysis Symposium (SAS)*, number 1145 in LNCS, pages 349–365, 1996.
53. Hongwei Xi and Frank Pfenning. Dependent types in practical programming. In *Proceedings of the 26th Annual ACM Symposium on Principles of Programming Languages (POPL'99)*, pages 214–227, San Antonio, TX, January 1999.

A Formal semantics

In this section we formalize a small fragment of X10 to illustrate the basic concepts behind constrained type-checking. In fact a very tiny language is chosen—a small extension of FJ with constrained types.

The language is functional in that assignment is not admitted. However, it is not difficult to introduce the notion of mutable fields, and assignment to such fields. Since constrained types may onl refer to immutable state, the validity of these types is not compromised by the introduction of state.

Further, we do not formalize overloading of methods. Rather, with FJ, we simply require that the input program be such that the class name `C` and method name `m` uniquely select the associated method on the class.

We do model properties, constrained clauses, class invariants, where clauses in methods and constructors, and dependent type casts.

A.1 Constraint system

Constraints are assumed to be drawn from a fixed constraint system, $\mathcal{C}$, with inference relation $\vdash_{\mathcal{C}}$ [48]. All constraint systems are required to support the trivial constraint `true`, conjunction, existential quantification and equality on constraint terms. Constraint terms include (final) variables, the special variable `self` (which may occur only in constraints `c` which occur in a constrained type `C(:c)`), and field selections `t.f`.

We summarize here properties of constraint systems described in [48] that are needed for the proofs: constraint systems may be thought of as presented via an intuitionistic Gentzen proof system supporting identity; affine and exchange on the left; existential quantification and conjunction on the left and right; and closure under substitution of terms. We denote the application of the substitution $\theta = [\overline{\mathtt{t}}/\overline{\mathtt{x}}]$ to a constraint `c` by $\mathtt{c}[\overline{\mathtt{t}}/\overline{\mathtt{x}}]$.

$$
\begin{array}{rl}
\text{(C Term)} & \mathtt{t} ::= \mathtt{x} \mid \mathtt{self} \mid \mathtt{this} \mid \mathtt{t.f} \\
& \quad \mid \mathtt{new}\ \mathtt{C}(\overline{\mathtt{t}}) \mid \mathtt{g}(\overline{\mathtt{t}}) \\
\text{(Constraint)} & \mathtt{c},\mathtt{d} ::= \mathtt{true} \mid \mathtt{p}(\overline{\mathtt{t}}) \\
& \quad \mid \mathtt{t} = \mathtt{t} \mid \mathtt{c},\mathtt{c} \mid \mathtt{T}\,\mathtt{x};\mathtt{c}
\end{array}
$$

All constraint systems are required to satisfy: $\texttt{new C}(\overline{\texttt{t}}).\texttt{f}_i = \texttt{t}_i$ provided that $\mathit{fields}(\texttt{C}) = \overline{\texttt{T}}\ \overline{\texttt{f}}$ (for some sequence of types $\overline{\texttt{T}}$).

Above, "," binds tighter than ";". We use the syntax `T x; c` for the constraint obtained by existentially quantifying the variable `x` of type `T` in `c`. `p` ranges over the collection of predicates supplied by the underlying constraint system, and `g` over the collection of functions.

A.2 Syntax

The syntax for the language is specified in Figure 4. The definitions are based on those in Featherweight Java [24].

A type is taken to be of the form `C(:c)` where `C` is the name of a class or interface and `c` is a constraint; we say that `C` is the *base* of the type `C(:c)`.

The *base type* of a type `C(:c)` (read as `C` *with* `c`) is `C`. We use the following shorthand for types: For a type `T` equal to `C(:c)`, we will write `S x; T` for `C(:S x; c)`, and `d,T` for `C(:d,c)`. Application of substitutions is extended to types by: $\texttt{C}(:\texttt{c})\theta = \texttt{C}(:\texttt{c}\theta)$.

A type assertion `C(:c) x` constrains the variable `x` to contain references to only those objects `o` that are instances of (subclasses of) `C` and for which the constraint `c` is true provided that occurrences of `self` in `c` are replaced by `o`. Thus in the constraint `c` of a constrained type `C(:c)`, `self` may be used to reference the object whose type is being specified. Note that `self` is distinct from `this`—`this` is permitted to occur in the clause of a type `T` only if `T` occurs in an instance field declaration or instance method declaration of a class; as usual, `this` is considered bound to the instance of the class to which the field or method declaration applies.

A *class declaration* $\texttt{class C}(\overline{\texttt{T}}\ \overline{\texttt{f}} : \texttt{c})\ \texttt{extends D}(:\texttt{d})\ \{\overline{\texttt{M}}\}$ is thought of as declaring a class `C` with the fields $\overline{\texttt{f}}$ (of type $\overline{\texttt{T}}$), a *declared class invariant* `c`, a *super-class invariant* `d` and a collection of methods $\overline{\texttt{M}}$. The constraints `c` and `d` are true for all instances of the class `C` (this is verified in the rule for type-checking constructors, T-NEW). In these constraints, `this` may be used to reference the current object; `self` does not have any meaning and must not be used.

A *method declaration* $\texttt{T}_0\ \texttt{m}(\overline{\texttt{T}}\ \overline{\texttt{x}} : \texttt{c})\{\ldots\}$ specifies the type of the arguments and the result, as usual. The method arguments $\overline{\texttt{x}}$ may occur in the argument types $\overline{\texttt{T}}$ and the return type $\texttt{T}_0$. The constraint `c` specifies additional constraints on the arguments $\overline{\texttt{x}}$ and `this` that must hold for a method invocation to be legal. Note that `self` does not make sense in `c` (no type is being defined), and must not occur in `c`.

(Class)	`L`	::= $\texttt{class C}(\overline{\texttt{T}}\ \overline{\texttt{f}} : \texttt{c})\ \texttt{extends T}\ \{\overline{\texttt{M}}\}$
(Method)	`M`	::= $\texttt{T m}(\overline{\texttt{T}}\ \overline{\texttt{x}} : \texttt{c})\{\texttt{return e;}\}$
(Expr)	`e`	::= `x` | `e.f` | $\texttt{e.m}(\overline{\texttt{e}})$ | $\texttt{new C}(\overline{\texttt{e}})$ | `(T)e`
(Type)	`S,T,U,Z`	::= $\texttt{C}(:\texttt{d})$

Fig. 4. CFJ Syntax

A.3 Static semantics

Type judgments. Typing judgments are of the form $\Gamma \vdash \mathtt{T}\ \mathtt{e}$ where Γ is a multiset of type assertions $\mathtt{T}\ \mathtt{x}$ and constraints $\mathtt{c}$.

T-VAR extends the identity rule ($\Gamma, x : C \vdash x : C$) of FJ to take into account the constraint entailment relation.

T-CAST encapsulates the three inference rules of FJ: T-UCAST, T-DCAST and T-SCAST for upwards cast, downwards cast, and "stupid" cast respectively.

In T-FIELD, we postulate the existence of a receiver object $\mathtt{o}$ of the given static type ($\mathtt{T}_0$). *fields*$(\mathtt{T}_0, \mathtt{o})$ is the set of typed fields for $\mathtt{T}_0$ with all occurrences of `this` replaced by $\mathtt{o}$. We record in the resulting constraint that $\mathtt{o.f}_i = \mathtt{self}$.[6] This permits transfer of information that may have been recorded in $\mathtt{T}_0$ about the field $\mathtt{f}_i$.

Similarly, in T-INVK we postulate the existence of a receiver object $\mathtt{o}$ of the given static type. For any type T, object $\mathtt{o}$ of type T and method name $\mathtt{m}$, let *mtype*$(\mathtt{T}, \mathtt{m}, \mathtt{o})$ be a copy of the signature of the method with `this` replaced by $\mathtt{o}$. We establish (under the assumption that the formals ($\overline{\mathtt{z}}$) have the static type of the actuals)[7] that actual types are subtypes of the formal types, and the method constraint is satisfied. This permits us to record the constraint $\mathtt{d}$ on the return type, with the formal variables $\overline{\mathtt{z}}$ existentially quantified.[8]

In T-NEW, similarly, we establish that the static types of the actual arguments to the constructor are subtypes of the declared types of the field, and contain enough information to satisfy the class invariant, $\mathtt{c}$. The declared types (and $\mathtt{c}$) contain references to $\mathtt{this}.\overline{\mathtt{f}}$; these must be replaced by the formals $\overline{\mathtt{f}}$, which carry information about the static type of the actuals. Note that the object $\mathtt{o}$ we hypothesized in an analogous situation in T-INVK does not exist; it will exist on successful invocation of the constructor. The constrained clause of the `new` expression contains all the information that can be gleaned from the static types of the actuals by assigning them to the corresponding fields of the object being created.

Subtyping. We add a single rule to the rules of FJ:

$$\mathtt{C} \sqsubseteq \mathtt{C} \quad \frac{\mathtt{class\ C}(\ldots)\ \mathtt{extends\ D}(\ldots)\{\ldots\}}{\mathtt{C} \sqsubseteq \mathtt{D}} \quad \frac{\mathtt{C} \sqsubseteq \mathtt{D} \quad \mathtt{D} \sqsubseteq \mathtt{E}}{\mathtt{C} \sqsubseteq \mathtt{E}} \quad \frac{\mathtt{C} \sqsubseteq \mathtt{D}\ \sigma(\Gamma, \mathtt{C}(:\mathtt{c})\ \mathtt{x}) \vdash_{\mathcal{C}} \mathtt{d}[\mathtt{x}/\mathtt{self}] \quad (\mathtt{x}\ \text{fresh})}{\Gamma \vdash \mathtt{C}(:\mathtt{c}) \sqsubseteq \mathtt{D}(:\mathtt{d})}$$

(Whenever we state an assumption of the form "$\mathtt{x}$ is fresh" in a rule we mean it is not free in the consequent of the rule.)

Typing judgment. We let Γ stand for multisets of type assertions, of the form $\mathtt{T}\ \mathtt{x}$,[9] and constraints. Typing judgments are of the form $\Gamma \vdash \mathtt{S}\ \mathtt{t}$ When Γ is empty, it is omitted.

[6] A new name $\mathtt{o}$ is necessary to name this object since $\mathtt{e}$ cannot be used. Arbitrary term expressions $\mathtt{e}$ are not permitted in constraints; the functions used in $\mathtt{e}$ may not be known to the constraint system, and $\mathtt{e}$ may have side-effects.

[7] This is stronger than assuming $\overline{\mathtt{Z}}$.

[8] Recall that the $\overline{\mathtt{z}}$ may occur in $\mathtt{d}$ but must not occur in a type in the calling environment; hence they must be existentially quantified in the resulting constraint.

[9] We use the non-standard notation $\mathtt{T}\ \mathtt{x}$ rather than the more familiar $\mathtt{x} : \mathtt{T}$ since `:` is used in the syntax of a type.

Let $\mathtt{C}$ be a class declared as $\mathtt{class\ C}(\overline{\mathtt{T}}\ \overline{\mathtt{f}} : \mathtt{c})\ \mathtt{extends\ D}(:\mathtt{d})\{\overline{\mathtt{M}}\}$. Let θ be a substitution and the type $\mathtt{T}$ be based on $\mathtt{C}$. We define $inv(T,\theta)$ as the conjunction $\mathtt{c}\theta, \mathtt{d}\theta$ and (recursively) $inv(\mathtt{D},\theta)$. We bottom out with $inv(\mathtt{Object},\theta) = \mathtt{true}$. For a variable $\mathtt{x}$, we use the shorthand $inv(\mathtt{C},\mathtt{x})$ to mean $inv(\mathtt{C},[\mathtt{x}/\mathtt{self}])$.

The definition of *mtype*(C,m) (the signature of a method named m in class C), *mbody*(C,m), (the body associated with method m in type C) and *fields*(C) (the sequence of fields and their types inherited or defined at C) is essentially as specified in FJ [24] with the difference that the method of a signature is taken to be of the form $\overline{\mathtt{S}}\ \overline{\mathtt{x}} : \mathtt{c} \to \mathtt{T}$. The variables $\mathtt{x}$ are permitted to occur in the types $\overline{\mathtt{S}}, \mathtt{T}$, and are considered bound, and subject to alpha-renaming. The definitions of *mtype*, *mbody*, *fields* are extended to apply to constrained types by ignoring the constraint. For a substitution θ we define $mtype(\mathtt{T},\mathtt{m},\theta)$ as the signature obtained by applying θ to $mtype(\mathtt{T},\mathtt{m})$, renaming bound variables as necessary. Similarly, for a substitution θ we define $fields(\mathtt{T},\theta)$ to be $\overline{S}\theta\ \overline{f}$, if the sequence of inherited and defined fields of the class underlying the type $\mathtt{T}$ is $\overline{S}\ \overline{f}$. We let $fields(\mathtt{T},\mathtt{x})$ stand for $fields(\mathtt{T},[\mathtt{x}/\mathtt{self}])$.

We define $\sigma(\Gamma)$ to be the set of constraints obtained from Γ by replacing each type assertion $\mathtt{C}(:\mathtt{d})\ \mathtt{x}$ in Γ with $\mathtt{d}[\mathtt{x}/\mathtt{self}], inv(C,x)$ and retaining any constraint in Γ.

$$\frac{\sigma(\Gamma, \mathtt{C}(:\mathtt{c})\ \mathtt{x}) \vdash_{\mathcal{C}} \mathtt{d}[\mathtt{x}/\mathtt{self}]}{\Gamma, \mathtt{C}(:\mathtt{c})\ \mathtt{x} \vdash \mathtt{C}(:\mathtt{d})\ \mathtt{x}}\ (\text{T-VAR}) \qquad \frac{\Gamma \vdash \mathtt{T}_0\ \mathtt{e} \quad fields(\mathtt{T}_0, \mathtt{z}_0) = \overline{\mathtt{U}}\ \overline{\mathtt{f}}_i \quad (\mathtt{z}_0 \text{ fresh})}{\Gamma \vdash (\mathtt{T}_0\ \mathtt{z}_0; \mathtt{z}_0.\mathtt{f}_i = \mathtt{self}; \mathtt{U}_i)\ \mathtt{e}.\mathtt{f}_i}\ (\text{T-FIELD}) \qquad \frac{\Gamma \vdash \mathtt{S}\ \mathtt{e}}{\Gamma \vdash \mathtt{T}\ (\mathtt{T})\mathtt{e}}\ (\text{T-CAST})$$

$$\frac{\begin{array}{l}\Gamma \vdash \mathtt{T}_{0:n}\ \mathtt{e}_{0:n} \qquad\qquad mtype(\mathtt{T}_0, \mathtt{m}, \mathtt{z}_0) = \mathtt{Z}_{1:n}\ \mathtt{z}_{1:n} : \mathtt{c} \to \mathtt{S} \\ \Gamma, \mathtt{T}_{0:n}\ \mathtt{z}_{0:n} \vdash \mathtt{T}_{1:n} \sqsubseteq \mathtt{Z}_{1:n} \quad \sigma(\Gamma, \mathtt{T}_{0:n}\ \mathtt{z}_{0:n}) \vdash_{\mathcal{C}} \mathtt{c} \quad (\mathtt{z}_{0:n} \text{ fresh})\end{array}}{\Gamma \vdash (\mathtt{T}_{0:n}\ \mathtt{z}_{0:n}; S)\ \mathtt{e}_0.\mathtt{m}(\mathtt{e}_{1:n})}\ (\text{T-INVK}) \qquad \frac{\begin{array}{l}\Gamma \vdash \overline{\mathtt{T}}\ \overline{\mathtt{e}} \quad \theta = [\overline{\mathtt{f}}/\mathtt{this}.\overline{\mathtt{f}}]\ fields(C,\theta) = \overline{\mathtt{Z}}\ \overline{\mathtt{f}} \\ \Gamma, \overline{\mathtt{T}}\ \overline{\mathtt{f}} \vdash \overline{\mathtt{T}} \sqsubseteq \overline{\mathtt{Z}} \qquad\qquad \sigma(\Gamma, \overline{\mathtt{T}}\ \overline{\mathtt{f}}) \vdash_{\mathcal{C}} inv(\mathtt{C},\theta)\end{array}}{\Gamma \vdash \mathtt{C}(:\overline{\mathtt{T}}\ \overline{\mathtt{f}}; \mathtt{self}.\overline{\mathtt{f}} = \overline{\mathtt{f}})\ \mathtt{new\ C}(\overline{\mathtt{e}})}\ (\text{T-NEW}$$

Method and class typing.

$$\frac{\overline{\mathtt{T}}\ \overline{\mathtt{x}}, \mathtt{C\ this}, \mathtt{c} \vdash \mathtt{S\ e}, \mathtt{S} \sqsubseteq \mathtt{T}}{\mathtt{T\ m}(\overline{\mathtt{T}}\ \overline{\mathtt{x}} : \mathtt{c})\{\mathtt{return\ e};\}\ \text{OK in C}} \qquad \frac{\overline{M}\ \text{OK in } C}{\mathtt{class\ C}(\overline{\mathtt{T}}\ \overline{\mathtt{f}} : \mathtt{c})\ \mathtt{extends\ D}(:\mathtt{d})\ \{\overline{\mathtt{M}}\}\ \text{OK}}$$

Fig. 5. Typing rules

A.4 Soundness

Theorem 1 (Subject Reduction).
If $\Gamma \vdash T\ e$ *and* $e \longrightarrow e'$*, then for some type S,* $\Gamma \vdash S\ e'$ *and* $\Gamma \vdash S \sqsubseteq T$.

Let the normal form of expressions be given by *values*, i.e. expressions:

(Values) $\mathtt{v} ::= \mathtt{new\ C}(\overline{\mathtt{v}})$

Theorem 2 (Progress). *If* $\vdash \mathtt{T\ e}$*, then one of the following conditions holds:*

1. $\mathtt{e}$ *is a value* $\mathtt{v}$*,*

Computation.

$$\frac{\mathit{fields}(C) = \overline{C}\ \overline{f}}{(\texttt{new C}(\overline{\texttt{e}})).\texttt{f}_i \longrightarrow \texttt{e}_i}\ (\text{R-FIELD}) \quad \frac{\mathit{mbody}(\texttt{m},\texttt{C}) = \overline{x}.\texttt{e}_0}{(\texttt{new C}(\overline{\texttt{e}})).\texttt{m}(\overline{\texttt{d}}) \longrightarrow [\overline{d}/\overline{x},\texttt{new}\ C(\overline{e})/\texttt{this}]\texttt{e}_0}\ (\text{R-INVK}) \quad \frac{\vdash C \sqsubseteq T[\texttt{new}\ C(\overline{\texttt{d}})/\texttt{self}]}{(\texttt{T})(\texttt{new C}(\overline{\texttt{d}})) \longrightarrow \texttt{new}\ C(\overline{\texttt{d}})}\ (\text{R-CAST})$$

Congruence.

$$\frac{\texttt{e}_0 \longrightarrow \texttt{e}_0{}'}{\texttt{e}_0.\texttt{f} \longrightarrow \texttt{e}_0{}'.\texttt{f}}\ (\text{RC-FIELD}) \quad \frac{\texttt{e}_0 \longrightarrow \texttt{e}_0{}'}{\texttt{e}_0.\texttt{m}(\overline{\texttt{e}}) \longrightarrow \texttt{e}_0{}'.\texttt{m}(\overline{\texttt{e}})}\ (\text{RC-INVK-RECV}) \quad \frac{\texttt{e}_i \longrightarrow \texttt{e}_i{}'}{\texttt{e}_0.\texttt{m}(\ldots,\texttt{e}_i,\ldots) \longrightarrow \texttt{e}_0.\texttt{m}(\ldots,\texttt{e}_i',\ldots)}\ (\text{RC-INVK-ARG})$$

$$\frac{\texttt{e}_i \longrightarrow \texttt{e}_i{}'}{\texttt{new C}(\ldots,\texttt{e}_i,\ldots) \longrightarrow \texttt{new C}(\ldots,\texttt{e}_i',\ldots)}\ (\text{RC-NEW-ARG}) \quad \frac{\texttt{e}_0 \longrightarrow \texttt{e}_0{}'}{(\texttt{C})\texttt{e}_0 \longrightarrow (\texttt{C})\texttt{e}_0{}'}\ (\text{RC-CAST})$$

Fig. 6. Reduction rules

2. $\texttt{e}$ *contains a subexpression* $(\texttt{T})\texttt{new C}(\overline{\texttt{v}})$ *such that* $\not\vdash \texttt{C} \sqsubseteq \texttt{T}[\texttt{new C}(\overline{\texttt{v}})/\texttt{self}]$,
3. *there exists* $\texttt{e}'$ *s.t.* $\texttt{e} \longrightarrow \texttt{e}'$.

Theorem 3 (Type Soundness).
If $\vdash \texttt{T e}$ *and* $\texttt{e} \longrightarrow^* \texttt{e}'$, *with* $\texttt{e}'$ *in normal form, then* $\texttt{e}'$ *is either (1) a value* $\texttt{v}$ *with* $\vdash \texttt{S v}$ *and* $\vdash \texttt{S} \sqsubseteq \texttt{T}$, *for some type* $\texttt{S}$, *or, (2) an expression containing a subexpression* $(\texttt{T})\texttt{new C}(\overline{\texttt{v}})$ *where* $\not\vdash \texttt{C} \sqsubseteq \texttt{T}[\texttt{new C}(\overline{\texttt{v}})/\texttt{self}]$.

Lemma 1 (Substitution Lemma). *Assume* $\Gamma \vdash \overline{\texttt{A}}\ \overline{\texttt{d}}$, $\Gamma \vdash \overline{\texttt{A}} \sqsubseteq \overline{\texttt{B}}$, *and* $\Gamma, \overline{\texttt{B}}\ \overline{\texttt{x}} \vdash \texttt{T e}$. *Then for some type* $\texttt{S}$ *s.t.* $\Gamma \vdash \texttt{S} \sqsubseteq \overline{\texttt{A}}\ \overline{\texttt{x}}; \texttt{T}$ *it is the case that* $\Gamma \vdash \texttt{S}\ \texttt{e}[\overline{\texttt{d}}/\overline{\texttt{x}}]$.

Lemma 2 (Weakening). *If* $\Gamma \vdash \texttt{T e}$, *then* $\Gamma, \texttt{S x} \vdash \texttt{T e}$.

Lemma 3 (Body type). *If* $\mathrm{mtype}(\texttt{T}_0, \texttt{m}) = \overline{\texttt{T}}\ \overline{\texttt{x}} : \texttt{c} \to \texttt{S}$, *and* $\mathit{mbody}(\texttt{m}, \texttt{T}_0) = \overline{\texttt{x}}.\texttt{e}$, *then for some* $\texttt{U}_0$ *with* $\texttt{T}_0 \sqsubseteq \texttt{U}_0$, *there exists* $\texttt{V} \sqsubseteq \texttt{S}$ *such that* $\overline{\texttt{T}}\ \overline{\texttt{x}}, \texttt{U}_0\ \texttt{this} \vdash \texttt{V e}$

A.5 Erasure

Constrained types in CFJ are a form of *refinement type* [20]. If constraints are erased from a well-typed program, the resulting program will behave identically to the original unerased program except that the original program might be unable to take a step on a cast.

Let $[\![\texttt{e}]\!]$ be the erasure of $\texttt{e}$ defined as follows:

$$\begin{aligned}
[\![\texttt{x}]\!] &= \texttt{x} \\
[\![\texttt{e.f}]\!] &= [\![\texttt{e}]\!].\texttt{f} \\
[\![\texttt{e.m}(\overline{\texttt{e}})]\!] &= [\![\texttt{e}]\!].\texttt{m}(\overline{[\![\texttt{e}]\!]}) \\
[\![\texttt{new C}(\overline{\texttt{e}})]\!] &= \texttt{new C}(\overline{[\![\texttt{e}]\!]}) \\
[\![(\texttt{C}(:\texttt{c}))\ \texttt{e}]\!] &= (\texttt{C})\ \overline{[\![\texttt{e}]\!]}
\end{aligned}$$

Theorem 4 (Erasure).
If $\vdash \texttt{C}(:\texttt{c})\ \texttt{e}$ *and* $\texttt{e} \longrightarrow^* \texttt{v}$, *then* $\vdash \texttt{C}\ [\![\texttt{e}]\!]$ *and* $[\![\texttt{e}]\!] \longrightarrow^* [\![\texttt{v}]\!]$.