

# An Analysis of Arithmetic Constraints on Integer Intervals

Krzysztof R. Apt

CWI, P.O. Box 94079, 1090 GB Amsterdam, the Netherlands and

University of Amsterdam, the Netherlands

Peter Zoetewij\*

Faculty of Electrical Engineering, Mathematics and Computer Science,

Delft University of Technology, P.O. Box 5031, 2600 GA Delft, the Netherlands

## Abstract

Arithmetic constraints on integer intervals are supported in many constraint programming systems. We study here a number of approaches to implement constraint propagation for these constraints. To describe them we introduce integer interval arithmetic. Each approach is explained using appropriate proof rules that reduce the variable domains. We compare these approaches using a set of benchmarks. For the most promising approach we provide results that characterize the effect of constraint propagation.

## 1 Introduction

### 1.1 Motivation

The subject of arithmetic constraints on reals has attracted a great deal of attention in the literature. In contrast, arithmetic constraints on integer intervals have not been studied even though they are supported in a number of constraint programming systems. In fact, constraint propagation for them is present in ECL<sup>i</sup>PS<sup>e</sup>, SICStus Prolog, GNU Prolog, ILOG Solver and undoubtedly most of the systems that support constraint propagation for linear constraints on integer intervals. Yet, in contrast to the case of linear constraints — see notably [14] — we did not encounter in the literature any analysis of this form of constraint propagation.

In this paper we study these constraints in a systematic way. It turns out that in contrast to linear constraints on integer intervals there are a number of

---

\*The work of the second author was supported by NWO, The Netherlands Organization for Scientific Research, under project number 612.069.003.

natural approaches to constraint propagation for these constraints. They differ in the extent to which the constraints are decomposed.

Even though arithmetic constraints on integer intervals need not be decomposed into atomic arithmetic constraints, as is common practice for constraints on reals, we found that it is beneficial to do so: it allows for efficient scheduling of the reduction rules and for reuse of auxiliary variables for common subterms between constraints.

It could be argued that since integer arithmetic is a special case of real arithmetic, specialized constraint propagation methods for integer arithmetic constraints are not needed. Indeed, a constraint satisfaction problem (CSP) involving arithmetic constraints on integer variables can be solved using any known method for constraints on reals, with additional constraints ensuring that the variables assume only integer values. This was suggested in [6] and is implemented, for example, in RealPaver [13]. However, a dedicated study and implementation of the integer case is beneficial for a number of reasons.

- In some cases the knowledge that we are dealing with integers yields a stronger constraint propagation than the approach through the constraint propagation for arithmetic constraints on reals. This can be also beneficial when we are dealing with hybrid problems that involve arithmetic constraints on both integer and real variables.
- The ‘indirect’ approach through the reals is based on floating-point numbers, which are of limited precision. This implies that no exact representation exists for integers outside certain bounds. We believe that it should be possible to deal with large integers precisely, and that we should not revert to a floating-point representation when other options exist. Using a library like GNU MP [11] we can use arbitrary length integers (called *multiple precision* integers in GNU MP), whose size is limited only by the available memory.
- Since arithmetic constraints on integer intervals are supported in a number of constraint programming systems, it is natural to investigate in a systematic way various approaches to their implementation. The approaches based on the integers are amenable to a clear theoretical analysis. In particular, in Section 9 and Subsection 10.1 we provide the characterization results that clarify the effect of constraint propagation for the approach that emerged in our studies as the fastest.

An example that supports the first argument is the constraint  $x \cdot y = z$ , where  $-3 \leq x \leq 3$ ,  $-1 \leq y \leq 1$ , and  $1 \leq z \leq 2$ . When all variables are integers, there are no solutions having  $x = 3$  or  $x = -3$ , and the constraint propagation methods that we consider here will actually remove these values from the domain of  $x$ . However, if these variables are considered to be reals, these values may not be removed, and solving the integer problem through constraint propagation methods for constraints on reals may lead to a larger search space.

As an indication that integer representation is not entirely a theoretical issue, consider the following benchmark from [6]. Find  $n$  integers  $x_1, \dots, x_n$ ,

$n$	solutions	CPU time (sec.)		
		double	long long	mpz
13	22	0.44	0.41	1.69
14	60	1.37	1.35	5.27
15	159	4.45	4.50	17.44
16	377	14.54	15.04	57.31
17	377	32.66	33.54	128.26
18	1007	106.77	110.98	419.74

Table 1: Comparison of timing results for various representations

$1 \leq x_i \leq n$ , verifying the conditions

$$\sum_{i=1}^n x_i = \sum_{i=1}^n i, \quad \prod_{i=1}^n x_i = \prod_{i=1}^n i, \quad x_1 \leq x_2 \leq \dots \leq x_n.$$

For  $n = 10$  the initial maximum value of the left-hand side expression of the second constraint equals  $10^{10}$ , which exceeds  $2^{32}$ , the number of values that can be represented as 32-bit integers. For  $n = 16$ , there is already no signed integer representation of this bound in 64 bits.

To show that arbitrary length integers can be affordable, Table 1 shows timing results for three small C++ programs that solve the above benchmark via a basic branch-and-propagate search process. These programs differ only in the representation of the bounds of the variables, and in the signature of the arithmetic operations applied to these bounds: 64-bit floating point numbers (`double`), 64-bit integers (`long long`), and arbitrary length integers (using the `mpz` data type of the GNU MP library). The programs were compiled using the same optimization flags as the default build of the GNU MP library, and the reported CPU times are user time in seconds, measured by the `time` command on a 1200 MHz AMD Athlon CPU.

The results for 64-bit integers and  $n \geq 16$  could be computed by initializing the upper bound of the auxiliary variable equated to the product of all problem variables to  $n!$ , which works for  $n < 20$ . These results indicate that on our hardware, the 64-bit integer and floating-point implementations are equally efficient, while for these specific problem instances, the cost of using arbitrary length integers is roughly a factor four. Note that in a full-fledged constraint solver, this overhead would be far less prominent, because compared to these small C++ programs, a large part of the execution time is spent in the framework that coordinates the computation (cf. the results in Subsection 11.2).

## 1.2 Plan of the Paper

In the next section we provide the relevant background material on CSPs and arithmetic constraints. The unifying tool in our analysis is integer interval arithmetic that is modeled after the real interval arithmetic (see, e.g., [15]). There are, however, essential differences since we deal with integers instead of

reals. For example, the product of two integer intervals does not need to be an integer interval. In Section 3 we introduce integer interval arithmetic and establish the basic results. Then in Section 4 we show that using integer interval arithmetic we can define succinctly the well-known constraint propagation for linear constraints on integer intervals.

The next three sections, 5, 6 and 7, form the main part of the paper. We introduce there three approaches to constraint propagation for arithmetic constraints on integer intervals. They differ in the way the constraints are treated: either they are left intact, or the multiple occurrences of variables are eliminated, or the constraints are decomposed into a set of atomic constraints. In Section 8 we discuss how these three approaches relate to various methods used to solve arithmetic constraints on reals.

Then in Section 9 we characterize the effect of constraint propagation for the atomic constraints. In Section 10 we discuss in detail our implementation of the alternative approaches, and in Section 11 we describe the experiments that were performed to compare them. They indicate that decomposition of the constraints, combined with a scheduling of the reduction rules that respects the hierarchical dependencies between the atomic constraints is superior to the other approaches. Finally, in Section 12 we provide the conclusions.

The preliminary results of this work were reported in [2] and [3].

## 2 Preliminaries

### 2.1 Constraint Satisfaction Problems

We now review the standard concepts of a constraint and of a constraint satisfaction problem. Consider a sequence of variables  $X := x_1, \dots, x_n$  where  $n \geq 0$ , with respective domains  $D_1, \dots, D_n$  associated with them. So each variable  $x_i$  ranges over the domain  $D_i$ . By a **constraint**  $C$  on  $X$  we mean a subset of  $D_1 \times \dots \times D_n$ . Given an element  $d := d_1, \dots, d_n$  of  $D_1 \times \dots \times D_n$  and a subsequence  $Y := x_{i_1}, \dots, x_{i_l}$  of  $X$  we denote by  $d[Y]$  the sequence  $d_{i_1}, \dots, d_{i_l}$ . In particular, for a variable  $x_i$  from  $X$ ,  $d[x_i]$  denotes  $d_i$ .

A **constraint satisfaction problem**, in short CSP, consists of a finite sequence of variables  $X$  with respective domains  $\mathcal{D}$ , together with a finite set  $\mathcal{C}$  of constraints, each on a subsequence of  $X$ . We write it as  $\langle \mathcal{C} ; x_1 \in D_1, \dots, x_n \in D_n \rangle$ , where  $X := x_1, \dots, x_n$  and  $\mathcal{D} := D_1, \dots, D_n$ .

By a **solution** to  $\langle \mathcal{C} ; x_1 \in D_1, \dots, x_n \in D_n \rangle$  we mean an element  $d \in D_1 \times \dots \times D_n$  such that for each constraint  $C \in \mathcal{C}$  on a sequence of variables  $X$  we have  $d[X] \in C$ . We call a CSP **consistent** if it has a solution and **inconsistent** if it does not. Two CSPs with the same sequence of variables are called **equivalent** if they have the same set of solutions. In what follows we consider CSPs whose constraints are defined in a simple language and when reasoning about them we identify the syntactic description of a constraint with its meaning being the set of tuples that satisfy it.

We view **constraint propagation** as a process of transforming CSPs that

maintains their equivalence. In what follows we define this process by means of proof rules that operate on CSPs and preserve equivalence. An interested reader can consult [1] or [2] for a precise explanation of this approach to describing constraint propagation.

## 2.2 Arithmetic Constraints

To define the arithmetic constraints we use the alphabet that comprises

- variables,
- two constants, 0 and 1,
- the unary minus function symbol ‘-’,
- three binary function symbols, ‘+’, ‘-’ and ‘·’, all written in the infix notation.

By an *arithmetic expression* we mean a term formed in this alphabet and by an *arithmetic constraint* a formula of the form

$$s \text{ op } t,$$

where  $s$  and  $t$  are arithmetic expressions and  $op \in \{<, \leq, =, \neq, \geq, >\}$ . For example

$$x^5 \cdot y^2 \cdot z^4 + 3x \cdot y^3 \cdot z^5 \leq 10 + 4x^4 \cdot y^6 \cdot z^2 - y^2 \cdot x^5 \cdot z^4 \quad (1)$$

is an arithmetic constraint. Here  $x^5$  is an abbreviation for  $x \cdot x \cdot x \cdot x \cdot x$ , while  $3x \cdot y^3 \cdot z^5$  is an abbreviation for  $x \cdot y^3 \cdot z^5 + x \cdot y^3 \cdot z^5 + x \cdot y^3 \cdot z^5$ , and similarly with the other expressions. If ‘·’ is not used in an arithmetic constraint, we call it a *linear constraint*.

By an *extended arithmetic expression* we mean a term formed in the above alphabet extended by the unary function symbols ‘ $\cdot^n$ ’ and ‘ $\sqrt[n]{\cdot}$ ’ for each  $n \geq 1$  and the binary function symbol ‘/’ written in the infix notation. For example

$$\sqrt[3]{(y^2 \cdot z^4)/(x^2 \cdot u^5)} \quad (2)$$

is an extended arithmetic expression. Here, unlike in (1),  $x^5$  is a term obtained by applying the function symbol ‘ $\cdot^5$ ’ to the variable  $x$ . The extended arithmetic expressions will be used only to define constraint propagation for the arithmetic constraints.

Fix now some arbitrary linear ordering  $\prec$  on the variables of the language. By a *monomial* we mean an integer or a term of the form

$$a \cdot x_1^{n_1} \cdot \dots \cdot x_k^{n_k}$$

where  $k > 0$ ,  $x_1, \dots, x_k$  are different variables ordered w.r.t.  $\prec$ , and  $a$  is a non-zero integer and  $n_1, \dots, n_k$  are positive integers. We call then  $x_1^{n_1} \cdot \dots \cdot x_k^{n_k}$  the *power product* of this monomial.

Next, by a **polynomial** we mean a term of the form

$$\sum_{i=1}^n m_i,$$

where  $n > 0$ , at most one monomial  $m_i$  is an integer, and the power products of the monomials  $m_1, \dots, m_n$  are pairwise different. Finally, by a **polynomial constraint** we mean an arithmetic constraint of the form  $s \text{ op } b$ , where  $s$  is a polynomial with no monomial being an integer,  $\text{op} \in \{<, \leq, =, \neq, \geq, >\}$ , and  $b$  is an integer. It is clear that by means of appropriate transformation rules we can transform each arithmetic constraint to a polynomial constraint. For example, assuming the ordering  $x \prec y \prec z$  on the variables, the arithmetic constraint (1) can be transformed to the polynomial constraint

$$2x^5 \cdot y^2 \cdot z^4 - 4x^4 \cdot y^6 \cdot z^2 + 3x \cdot y^3 \cdot z^5 \leq 10$$

So, without loss of generality, from now on we shall limit our attention to polynomial constraints.

Next, let us discuss the domains over which we interpret the arithmetic constraints. By an **integer interval**, or an **interval** in short, we mean an expression of the form

$$[a..b]$$

where  $a$  and  $b$  are integers;  $[a..b]$  denotes the set of all integers between  $a$  and  $b$ , including  $a$  and  $b$ . If  $a > b$ , we call  $[a..b]$  the **empty interval** and denote it by  $\emptyset$ . By a **range** we mean an expression of the form

$$x \in I$$

where  $x$  is a variable and  $I$  is an interval. Sets of the form  $\{x \in \mathcal{Z} \mid x \geq a\}$  and  $\{x \in \mathcal{Z} \mid x \leq b\}$  are called **extended intervals**.

We link the arithmetic constraints with the notion of a constraint defined in the previous section by associating in the standard way with each arithmetic constraint its **interpretation**. For an arithmetic constraint on variables  $x_1, \dots, x_n$  with respective integer interval domains  $D_1, \dots, D_n$  this is a subset of  $D_1 \times \dots \times D_n$ .

### 3 Integer Set Arithmetic

To reason about the arithmetic constraints we employ a generalization of the arithmetic operations to the sets of integers. Here and elsewhere  $\mathcal{Z}$ ,  $\mathcal{N}$ , and  $\mathcal{R}$  denote the sets of all integers, natural numbers, and reals, respectively.

#### 3.1 Definitions

For  $X, Y$  sets of integers we define the following operations:

- addition:

$$X + Y := \{x + y \mid x \in X, y \in Y\},$$

- subtraction:

$$X - Y := \{x - y \mid x \in X, y \in Y\},$$

- multiplication:

$$X \cdot Y := \{x \cdot y \mid x \in X, y \in Y\},$$

- division:

$$X/Y := \{u \in \mathcal{Z} \mid \exists x \in X \exists y \in Y u \cdot y = x\},$$

- exponentiation:

$$X^n := \{x^n \mid x \in X\},$$

for each natural number  $n > 0$ ,

- root extraction:

$$\sqrt[n]{X} := \{x \in \mathcal{Z} \mid x^n \in X\},$$

for each natural number  $n > 0$ .

All the operations except division and root extraction are defined in the expected way. We shall return to the division operation in Section 7. At the moment it suffices to note the division operation is defined for all sets of integers, including  $Y = \emptyset$  and  $Y = \{0\}$ . This division operation corresponds to the following division operation on the sets of reals introduced in [18]:

$$X \oslash Y := \{u \in \mathcal{R} \mid \exists x \in X \exists y \in Y u \cdot y = x\}.$$

For an integer or real number  $a$  and  $op \in \{+, -, \cdot, /, \oslash\}$  we identify  $a op X$  with  $\{a\} op X$  and  $X op a$  with  $X op \{a\}$ .

To present the rules we are interested in we shall also use the addition and division operations on the sets of reals. Addition is defined in the same way as for the sets of integers, and for division we use the  $\oslash$  operator defined above. In [15] it is explained how to implement these operations on, possibly unbounded, real intervals.

Further, given a set  $A$  of integers or reals, we define

$$\leq A := \{x \in \mathcal{Z} \mid \exists a \in A x \leq a\},$$

$$\geq A := \{x \in \mathcal{Z} \mid \exists a \in A x \geq a\},$$

so for example  $\leq \mathcal{N} = \mathcal{Z}$ , and  $\geq \{-1, 1\}$  and  $\geq (-2, 2)$  both denote the extended interval of all integers greater than or equal to  $-1$ , where  $(-2, 2)$  denotes an open interval of real numbers.

When limiting our attention to intervals of integers the following simple observation is of importance.

**Note 3.1** For  $X, Y$  integer intervals and  $a$  an integer the following holds:

- $X \cap Y, X + Y, X - Y$  are integer intervals.

- $X/\{a\}$  is an integer interval.
- $X \cdot Y$  does not have to be an integer interval, even if  $X = \{a\}$  or  $Y = \{a\}$ .
- $X/Y$  does not have to be an integer interval.
- For each  $n > 1$ ,  $X^n$  does not have to be an integer interval.
- For odd  $n > 1$ ,  $\sqrt[n]{X}$  is an integer interval.
- For even  $n > 1$ ,  $\sqrt[n]{X}$  is an integer interval or a disjoint union of two integer intervals.  $\square$

For example in the following cases we get intervals as outcomes:

$$\begin{aligned} [2..4] + [3..8] &= [5..12], \\ [3..7] - [1..8] &= [-5..6], \\ \sqrt[3]{[-30..100]} &= [-3..4], \\ \sqrt[2]{[-100..9]} &= [-3..3], \end{aligned}$$

while in the following ones not:

$$\begin{aligned} [3..3] \cdot [1..2] &= \{3, 6\}, \\ [3..5]/[-1..2] &= \{-5, -4, -3, 2, 3, 4, 5\}, \\ [-3..5]/[-1..2] &= \mathcal{Z}, \\ [1..2]^2 &= \{1, 4\}, \\ \sqrt[2]{[1..9]} &= [-3..-1] \cup [1..3]. \end{aligned}$$

To deal with the problem that non-interval domains can be produced by some of the operations we introduce the following operation on the sets of integers:

$$\text{int}(X) := \begin{cases} \text{smallest integer interval containing } X & \text{if } X \text{ is finite,} \\ \mathcal{Z} & \text{otherwise.} \end{cases}$$

For example  $\text{int}([3..5]/[-1..2]) = [-5..5]$  and  $\text{int}([-3..5]/[-1..2]) = \mathcal{Z}$ .

## 3.2 Implementation

To define constraint propagation for the arithmetic constraints on integer intervals we shall use the integer set arithmetic, mainly limited to the integer intervals. This brings us to the discussion of how to implement the introduced operations on the integer intervals. Since we are only interested in maintaining the property that the sets remain integer intervals or the set of integers  $\mathcal{Z}$  we shall clarify how to implement the intersection, addition, subtraction and root extraction operations of the integer intervals and the  $\text{int}(\cdot)$  closure of the multiplication, division and exponentiation operations on the integer intervals. The case when one of the intervals is empty is easy to deal with. So we assume that we deal with non-empty intervals  $[a..b]$  and  $[c..d]$ , that is  $a \leq b$  and  $c \leq d$ .

**Intersection, addition and subtraction** It is easy to see that

$$\begin{aligned} [a..b] \cap [c..d] &= [\max(a, c).. \min(b, d)], \\ [a..b] + [c..d] &= [a + c .. b + d], \\ [a..b] - [c..d] &= [a - d .. b - c]. \end{aligned}$$

So the interval intersection, addition, and subtraction are straightforward to implement.

**Root extraction** The outcome of the root extraction operator applied to an integer interval will be an integer interval or a disjoint union of two integer intervals. We shall explain in Section 5 why it is advantageous not to apply  $\text{int}(\cdot)$  to the outcome. This operator can be implemented by means of the following case analysis.

*Case 1.* Suppose  $n$  is odd. Then

$$\sqrt[n]{[a..b]} = [\lceil \sqrt[n]{a} \rceil .. \lfloor \sqrt[n]{b} \rfloor].$$

*Case 2.* Suppose  $n$  is even and  $b < 0$ . Then

$$\sqrt[n]{[a..b]} = \emptyset.$$

*Case 3.* Suppose  $n$  is even and  $b \geq 0$ . Then

$$\sqrt[n]{[a..b]} = [-\lfloor \sqrt[n]{|b|} \rfloor .. -\lceil \sqrt[n]{|a_+|} \rceil] \cup [\lceil \sqrt[n]{|a_+|} \rceil .. \lfloor \sqrt[n]{|b|} \rfloor]$$

where  $a_+ := \max(0, a)$ .

**Multiplication** For the remaining operations we only need to explain how to implement the  $\text{int}(\cdot)$  closure of the outcome. First note that

$$\text{int}([a..b] \cdot [c..d]) = [\min(A).. \max(A)],$$

where  $A = \{a \cdot c, a \cdot d, b \cdot c, b \cdot d\}$ .

Using an appropriate case analysis we can actually compute the bounds of  $\text{int}([a..b] \cdot [c..d])$  directly in terms of the bounds of the constituent intervals.

**Division** In contrast, the  $\text{int}(\cdot)$  closure of the interval division is not so straightforward to compute. The reason is that, as we shall see in a moment, we cannot express the result in terms of some simple operations on the interval bounds.

Consider non-empty integer intervals  $[a..b]$  and  $[c..d]$ . In analyzing the outcome of  $\text{int}([a..b]/[c..d])$  we distinguish the following cases.

*Case 1.* Suppose  $0 \in [a..b]$  and  $0 \in [c..d]$ .

Then by definition  $\text{int}([a..b]/[c..d]) = \mathcal{Z}$ . For example,

$$\text{int}([-1..100]/[-2..8]) = \mathcal{Z}.$$

*Case 2.* Suppose  $0 \notin [a..b]$  and  $c = d = 0$ .

Then by definition  $\text{int}([a..b]/[c..d]) = \emptyset$ . For example,

$$\text{int}([10..100]/[0..0]) = \emptyset.$$

*Case 3.* Suppose  $0 \notin [a..b]$  and  $c < 0$  and  $0 < d$ .

It is easy to see that then

$$\text{int}([a..b]/[c..d]) = [-e..e],$$

where  $e = \max(|a|, |b|)$ . For example,

$$\text{int}([-100.. -10]/[-2..5]) = [-100..100].$$

*Case 4.* Suppose  $0 \notin [a..b]$  and either  $c = 0$  and  $d \neq 0$  or  $c \neq 0$  and  $d = 0$ .

Then  $\text{int}([a..b]/[c..d]) = \text{int}([a..b]/([c..d] - \{0\}))$ . For example

$$\text{int}([1..100]/[-7..0]) = \text{int}([1..100]/[-7.. -1]).$$

This allows us to reduce this case to Case 5 below.

*Case 5.* Suppose  $0 \notin [c..d]$ .

This is the only case when we need to compute  $\text{int}([a..b]/[c..d])$  indirectly.

First, observe that we have

$$\text{int}([a..b]/[c..d]) \subseteq [\lceil \min(A) \rceil .. \lfloor \max(A) \rfloor],$$

where  $A = \{a/c, a/d, b/c, b/d\}$ .

However, the equality does not need to hold here. Indeed, note for example that  $\text{int}([155..161]/[9..11]) = [16..16]$ , whereas for  $A = \{155/9, 155/11, 161/9, 161/11\}$  we have  $\lceil \min(A) \rceil = 15$  and  $\lfloor \max(A) \rfloor = 17$ . The problem is that the value 16 is obtained by dividing 160 by 10 and none of these two values is an interval bound.

This complication can be solved by preprocessing the interval  $[c..d]$  so that its bounds are actual divisors of an element of  $[a..b]$ . First, we look for the least  $c' \in [c..d]$  such that  $\exists x \in [a..b] \exists u \in \mathcal{Z} u \cdot c' = x$ . Using a case analysis, the latter property can be established without search. Suppose for example that  $a > 0$  and  $c > 0$ . In this case, if  $c' \cdot \lfloor b/c' \rfloor \geq a$ , then  $c'$  has the required property. Similarly, we look for the largest  $d' \in [c..d]$  for which an analogous condition holds. Now  $\text{int}([a..b]/[c..d]) = [\lceil \min(A) \rceil .. \lfloor \max(A) \rfloor]$ , where  $A = \{a/c', a/d', b/c', b/d'\}$ .

In view of this auxiliary computation (in case when  $0 \notin [c..d]$ ) we shall introduce in Section 10 a modified division operation with a more direct implementation.

**Exponentiation** The  $\text{int}(\cdot)$  closure of the interval exponentiation is straightforward to implement by distinguishing the following cases.

*Case 1.* Suppose  $n$  is odd. Then

$$\text{int}([a..b]^n) = [a^n.. b^n].$$

Case 2. Suppose  $n$  is even and  $0 \leq a$ . Then

$$\text{int}([a..b]^n) = [a^n..b^n].$$

Case 3. Suppose  $n$  is even and  $b \leq 0$ . Then

$$\text{int}([a..b]^n) = [b^n..a^n].$$

Case 4. Suppose  $n$  is even and  $a < 0$  and  $0 < b$ . Then

$$\text{int}([a..b]^n) = [0..\max(a^n, b^n)].$$

### 3.3 Correctness Lemma

Given now an extended arithmetic expression  $s$  each variable of which ranges over an integer interval, we define  $\text{int}(s)$  as the integer interval or the set  $\mathcal{Z}$  obtained by systematically replacing each function symbol by the application of the  $\text{int}(\cdot)$  operation to the corresponding integer set operation. For example, for the extended arithmetic expression  $s := \sqrt[3]{(y^2 \cdot z^4)/(x^2 \cdot u^5)}$  of (2) we have

$$\text{int}(s) = \text{int}(\sqrt[3]{\text{int}(\text{int}(Y^2) \cdot \text{int}(Z^4)) / \text{int}(\text{int}(X^2) \cdot \text{int}(U^5))}),$$

where we assume that  $x$  ranges over  $X$ , etc.

The discussion in the previous subsection shows how to compute  $\text{int}(s)$  given an extended arithmetic expression  $s$  and the integer interval domains of its variables.

The following lemma is crucial for our considerations. It is a counterpart of the so-called ‘Fundamental Theorem of Interval Arithmetic’ established in [17]. Because we deal here with the integer domains an additional assumption is needed to establish the desired conclusion.

**Lemma 3.2 (Correctness)** *Let  $s$  be an extended arithmetic expression with the variables  $x_1, \dots, x_n$ . Assume that each variable  $x_i$  of  $s$  ranges over an integer interval  $X_i$ . Choose  $a_i \in X_i$  for  $i \in [1..n]$  and denote by  $s(a_1, \dots, a_n)$  the result of replacing in  $s$  each occurrence of a variable  $x_i$  by  $a_i$ .*

*Suppose that each subexpression of  $s(a_1, \dots, a_n)$  evaluates to an integer. Then the result of evaluating  $s(a_1, \dots, a_n)$  is an element of  $\text{int}(s)$ .*

**Proof.** The proof follows by a straightforward induction on the structure of  $s$ . □

## 4 An Intermezzo: Constraint Propagation for Linear Constraints

Even though we focus here on arithmetic constraints on integer intervals, it is helpful to realize that the integer interval arithmetic is also useful to define in a succinct way the well-known rules for constraint propagation for linear

constraints (studied in detail in [14]). To this end consider first a constraint  $\sum_{i=1}^n a_i \cdot x_i = b$ , where  $n \geq 0$ ,  $a_1, \dots, a_n$  are non-zero integers,  $x_1, \dots, x_n$  are different variables, and  $b$  is an integer. We rewrite this constraint  $n$  times, each time isolating one variable, to obtain an extended arithmetic expression for each variable  $x_j$ . To each of these extended arithmetic expressions we apply then the int operation of Subsection 3.3, which yields an update for the domain of the corresponding variable  $x_j$ . To reason about this procedure we can use the following rule parametrized by  $j \in [1..n]$ :

*LINEAR EQUALITY*

$$\frac{\langle \sum_{i=1}^n a_i \cdot x_i = b ; x_1 \in D_1, \dots, x_j \in D_j, \dots, x_n \in D_n \rangle}{\langle \sum_{i=1}^n a_i \cdot x_i = b ; x_1 \in D_1, \dots, x_j \in D'_j, \dots, x_n \in D_n \rangle}$$

where  $D'_j := D_j \cap \text{int} \left( (b - \sum_{i \in [1..n] - \{j\}} a_i \cdot x_i) / a_j \right)$ .

Note that by virtue of Note 3.1

$$D'_j = D_j \cap (b - \sum_{i \in [1..n] - \{j\}} \text{int}(a_i \cdot D_i)) / a_j.$$

To see that this rule preserves equivalence, first note that all our reduction rules compute the domain updates via intersection with the original domain, preventing that domains are extended by their application. Further, suppose that for some  $d_1 \in D_1, \dots, d_n \in D_n$  we have  $\sum_{i=1}^n a_i \cdot d_i = b$ . Then for  $j \in [1..n]$  we have

$$d_j = (b - \sum_{i \in [1..n] - \{j\}} a_i \cdot d_i) / a_j$$

which by the Correctness Lemma 3.2 implies that

$$d_j \in \text{int} \left( (b - \sum_{i \in [1..n] - \{j\}} a_i \cdot x_i) / a_j \right),$$

i.e.,  $d_j \in D'_j$ .

Next, consider a constraint  $\sum_{i=1}^n a_i \cdot x_i \leq b$ , where  $a_1, \dots, a_n, x_1, \dots, x_n$  and  $b$  are as above. To reason about it we can use the following rule parametrized by  $j \in [1..n]$ :

*LINEAR INEQUALITY*

$$\frac{\langle \sum_{i=1}^n a_i \cdot x_i \leq b ; x_1 \in D_1, \dots, x_j \in D_j, \dots, x_n \in D_n \rangle}{\langle \sum_{i=1}^n a_i \cdot x_i \leq b ; x_1 \in D_1, \dots, x_j \in D'_j, \dots, x_n \in D_n \rangle}$$

where  $D'_j := D_j \cap (\leq \text{int}(b - \sum_{i \in [1..n] - \{j\}} a_i \cdot x_i) / a_j)$ .

To see that this rule preserves equivalence, suppose that for some  $d_1 \in D_1, \dots, d_n \in D_n$  we have  $\sum_{i=1}^n a_i \cdot d_i \leq b$ . Then  $a_j \cdot d_j \leq b - \sum_{i \in [1..n] - \{j\}} a_i \cdot d_i$ . By the Correctness Lemma 3.2

$$b - \sum_{i \in [1..n] - \{j\}} a_i \cdot d_i \in \text{int}(b - \sum_{i \in [1..n] - \{j\}} a_i \cdot x_i),$$

so by definition

$$a_j \cdot d_j \in \leq \text{int}(b - \sum_{i \in [1..n] - \{j\}} a_i \cdot x_i)$$

and consequently

$$d_j \in^{\leq} \text{int}(b - \sum_{i \in [1..n] - \{j\}} a_i \cdot x_i) / a_j.$$

This implies that  $d_j \in D'_j$ .

As an alternative to evaluating  $\text{int}(\sum_{i \in [1..n] - \{j\}} a_i \cdot x_i)$  on every application of the *LINEAR EQUALITY* and *LINEAR INEQUALITY* rules, we could maintain the interval  $\text{int}(\sum_1^n a_i \cdot x_i)$  in an auxiliary variable, and subtract  $\text{int}(a_j \cdot x_j)$  from it. This corresponds to the two-step propagation described in [14]. If changes to  $D_j$  are propagated back to the auxiliary variable, this does not affect the reduction achieved by the subsequent applications of the rules, while the number of interval arithmetic operations involved in the application of a rule becomes constant, instead of linear in the number  $n$  of variables.

## 5 Constraint Propagation: Direct Approach

As a first approach to constraint propagation for arithmetic constraints on integer intervals, we propose to use the constraints directly, in their original form. This is an extension of the approach of Section 4 from linear constraints to general arithmetic constraints, and entails that these constraints are rewritten to isolate all occurrences of each variable. The resulting extended arithmetic expressions are then evaluated to obtain updates for the isolated variables.

The following example illustrates this approach. Consider the constraint

$$x^3 \cdot y - x \leq 40$$

and the ranges  $x \in [1..100]$  and  $y \in [1..100]$ . We can rewrite it as

$$x \leq \left\lfloor \sqrt[3]{(40 + x)/y} \right\rfloor \tag{3}$$

since  $x$  assumes integer values. The maximum value the expression on the right-hand side can take is  $\lfloor \sqrt[3]{140} \rfloor$ , so we conclude  $x \leq 5$ . By reusing (3), now with the information that  $x \in [1..5]$ , we conclude that the maximum value the expression on the right-hand side of (3) can take is actually  $\lfloor \sqrt[3]{45} \rfloor$ , from which it follows that  $x \leq 3$ .

In the case of  $y$  we can isolate it by rewriting the original constraint as  $y \leq 40/x^3 + 1/x^2$  from which it follows that  $y \leq 41$ , since by assumption  $x \geq 1$ . So we could reduce the domain of  $x$  to  $[1..3]$  and the domain of  $y$  to  $[1..41]$ . This interval reduction is optimal, since  $x = 1, y = 41$  and  $x = 3, y = 1$  are both solutions to the original constraint  $x^3 \cdot y - x \leq 40$ . So rewriting the constraint as  $x \geq x^3 \cdot y - 40$  does not yield a new lower bound for  $x$ .

More formally, consider a polynomial constraint  $\sum_{i=1}^m m_i = b$  where  $m > 0$ , no monomial  $m_i$  is an integer, the power products of the monomials are pairwise different, and  $b$  is an integer. Suppose that  $x_1, \dots, x_n$  are its variables ordered w.r.t.  $\prec$ .

Select a non-integer monomial  $m_l$  and assume it is of the form

$$a \cdot y_1^{n_1} \cdot \dots \cdot y_k^{n_k},$$

where  $k > 0$ ,  $y_1, \dots, y_k$  are different variables ordered w.r.t.  $\prec$ ,  $a$  is a non-zero integer and  $n_1, \dots, n_k$  are positive integers. So each  $y_i$  variable equals to some variable in  $\{x_1, \dots, x_n\}$ . Suppose that  $y_p$  equals to  $x_j$ . We introduce the following proof rule:

*POLYNOMIAL EQUALITY*

$$\frac{\langle \sum_{i=1}^n m_i = b ; x_1 \in D_1, \dots, x_j \in D_j, \dots, x_n \in D_n \rangle}{\langle \sum_{i=1}^n m_i = b ; x_1 \in D_1, \dots, x_j \in D'_j, \dots, x_n \in D_n \rangle}$$

where  $D'_j := \text{int} \left( D_j \cap \sqrt[n_p]{\text{int} \left( (b - \sum_{i \in [1..m] - \{l\}} m_i) / s \right)} \right)$ , with  $s := a \cdot y_1^{n_1} \cdot \dots \cdot y_{p-1}^{n_{p-1}} \cdot y_{p+1}^{n_{p+1}} \cdot \dots \cdot y_k^{n_k}$ .

To see that this rule preserves equivalence, choose some  $d_1 \in D_1, \dots, d_n \in D_n$ . To simplify the notation, given an extended arithmetic expression  $t$  denote by  $t'$  the result of evaluating  $t$  after each occurrence of a variable  $x_i$  is replaced by  $d_i$ .

Suppose that  $\sum_{i=1}^m m'_i = b$ . Then

$$d_j^{n_p} \cdot s' = b - \sum_{i \in [1..m] - \{l\}} m'_i,$$

so by the Correctness Lemma 3.2 applied to  $b - \sum_{i \in [1..m] - \{l\}} m'_i$  and to  $s$

$$d_j^{n_p} \in \text{int}(b - \sum_{i \in [1..m] - \{l\}} m_i) / \text{int}(s).$$

Hence

$$d_j \in \sqrt[n_p]{\text{int}(b - \sum_{i \in [1..m] - \{l\}} m_i) / \text{int}(s)}$$

and consequently

$$d_j \in \text{int} \left( D_j \cap \sqrt[n_p]{\text{int} \left( (b - \sum_{i \in [1..m] - \{l\}} m_i) / s \right)} \right)$$

i.e.,  $d_j \in D'_j$ .

Note that we do not apply  $\text{int}(\cdot)$  to the outcome of the root extraction operation. For even  $n_p$  this means that the second operand of the intersection can be a union of two intervals, instead of a single interval. To see why this is desirable, consider the constraint  $x^2 - y = 0$  in the presence of ranges  $x \in [0..10]$ ,  $y \in [25..100]$ . Using the  $\text{int}(\cdot)$  closure of the root extraction we would not be able to update the lower bound of  $x$  to 5.

Next, consider a polynomial constraint  $\sum_{i=1}^m m_i \leq b$ . Below we adopt the assumptions and notation used when defining the *POLYNOMIAL EQUALITY*

rule. To formulate the appropriate rule we stipulate that for the extended arithmetic expressions  $s$  and  $t$

$$\text{int}((\leq s)/t) := \geq Q \cap \leq Q,$$

with  $Q = (\leq \text{int}(s))/\text{int}(t)$ .

To reason about this constraint we use the following rule:

*POLYNOMIAL INEQUALITY*

$$\frac{\langle \sum_{i=1}^n m_i \leq b ; x_1 \in D_1, \dots, x_j \in D_j, \dots, x_n \in D_n \rangle}{\langle \sum_{i=1}^n m_i \leq b ; x_1 \in D_1, \dots, x_j \in D'_j, \dots, x_n \in D_n \rangle}$$

where  $D'_j := \text{int} \left( D_j \cap \sqrt[n_p]{\text{int}(\leq (b - \sum_{i \in [1..m] - \{l\}} m_i)/s)} \right)$ , with  $s := a \cdot y_1^{n_1} \cdot \dots \cdot y_{p-1}^{n_{p-1}} \cdot y_{p+1}^{n_{p+1}} \cdot \dots \cdot y_k^{n_k}$ .

To prove that this rule preserves equivalence, choose some  $d_1 \in D_1, \dots, d_n \in D_n$ . As above given an extended arithmetic expression  $t$  we denote by  $t'$  the result of evaluating  $t$  when each occurrence of a variable  $x_i$  in  $t$  is replaced by  $d_i$ .

Suppose that  $\sum_{i=1}^m m'_i \leq b$ . Then

$$d_j^{n_p} \cdot s' \leq b - \sum_{i \in [1..m] - \{l\}} m'_i.$$

By the Correctness Lemma 3.2

$$b - \sum_{i \in [1..m] - \{l\}} m'_i \in \text{int}(b - \sum_{i \in [1..m] - \{l\}} m_i),$$

so by definition

$$d_j^{n_p} \cdot s' \in \leq \text{int}(b - \sum_{i \in [1..m] - \{l\}} m_i).$$

Hence by the definition of the division operation on the sets of integers

$$d_j^{n_p} \in \leq \text{int}(b - \sum_{i \in [1..m] - \{l\}} m_i) / \text{int}(s)$$

Consequently

$$d_j \in \sqrt[n_p]{\leq \text{int}(b - \sum_{i \in [1..m] - \{l\}} m_i) / \text{int}(s)}.$$

This implies that  $d_j \in D'_j$ .

Note that the set  $\leq \text{int}(b - \sum_{i \in [1..m] - \{l\}} m_i)$ , which occurs when the expression for  $D'_j$  is expanded according to the above definition of  $\text{int}((\leq s)/t)$ , is not an interval. So to properly implement this rule we need to extend the implementation of the division operation discussed in Subsection 3.2 to the case when the numerator is an extended interval.

If the sum of the intervals associated with each of the monomials in a polynomial constraint is maintained in an auxiliary variable, as we discussed at the end of Section 4 for linear constraints, then the rules can be applied using a constant number of interval additions. However, interval division is not the inverse operation of interval multiplication, so the same technique cannot be applied

to monomials, and the number of multiplications and exponentiations will be linear in the size of the monomial  $s$ .

In an optimized version of the direct approach we simplify the fractions of two polynomials by splitting the division over addition and subtraction and by dividing out common powers of variables and greatest common divisors of the constant factors. Subsequently, fractions whose denominators have identical power products are added. We used this optimization in the initial example by simplifying  $(40 + x)/x^3$  to  $40/x^3 + 1/x^2$ . The reader may check that without this simplification step we can only deduce that  $y \leq 43$ .

To provide details of this optimization, given two monomials  $s$  and  $t$ , we denote by

$$[s/t]$$

the result of performing this simplification operation on  $s$  and  $t$ . For example,  $[(2 \cdot x^3 \cdot y)/(4 \cdot x^2)]$  equals  $(x \cdot y)/2$ , whereas  $[(4 \cdot x^3 \cdot y)/(2 \cdot y^2)]$  equals  $(2 \cdot x^3)/y$ .

Because the validity of the simplification depends on the sign of the denominator, we assume that the domains of the variables  $y_1, \dots, y_{p-1}, y_{p+1}, \dots, y_n$  of  $m_i$  do not contain 0. For a monomial  $s$  involving variables ranging over the integer intervals that do not contain 0, the set  $\text{int}(s)$  either contains only positive numbers or only negative numbers. In the first case we write  $\text{sign}(s) = +$  and in the second case we write  $\text{sign}(s) = -$ .

The new domain of the variable  $x_j$  in the *POLYNOMIAL INEQUALITY* rule is defined using two sequences  $m'_0 \dots m'_n$  and  $s'_0 \dots s'_n$  of extended arithmetic expressions such that

$$m'_0/s'_0 = [b/s] \text{ and } m'_i/s'_i = -[m_i/s] \text{ for } i \in [1..m].$$

Let  $S := \{s'_i \mid i \in [0..m] - \{l\}\}$  and for an extended arithmetic expression  $t \in S$  let  $I_t := \{i \in [0..m] - \{l\} \mid s'_i = t\}$ . We denote then by  $p_t$  the polynomial  $\sum_{i \in I_t} m'_i$ . The new domains are then defined by

$$D'_j := \text{int} \left( D_j \cap \sqrt[n_p]{\leq \text{int}(\sum_{t \in S} p_t \oslash t)} \right)$$

if  $\text{sign}(s) = +$ , and by

$$D'_j := \text{int} \left( D_j \cap \sqrt[n_p]{\geq \text{int}(\sum_{t \in S} p_t \oslash t)} \right)$$

if  $\text{sign}(s) = -$ . Here the  $\text{int}(s)$  notation used in the Correctness Lemma 3.2 is extended to expressions involving the division operator  $\oslash$  on real intervals in the obvious way. We define the  $\text{int}(\cdot)$  operator applied to a bounded set of reals, as produced by the division and addition operators in the above two expressions for  $D'_j$ , to denote the smallest interval of reals containing that set.

Returning again to the discussion of the two-step propagation technique of [14], which we started at the end of Section 4, note that in this case, the  $\text{int}(\cdot)$  operation is applied after removing the common power products. For this reason, there is no straightforward way to calculate  $\text{int}(\sum_{t \in S} p_t \oslash t)$  from the sum of all intervals associated with the monomials of a polynomial constraint.

## 6 Constraint Propagation: Partial Decomposition

As a second approach to constraint propagation for arithmetic constraints on integer intervals, we limit our attention to a special type of polynomial constraints, namely the ones of the form  $s \text{ op } b$ , where  $s$  is a polynomial in which each variable occurs *at most once* and where  $b$  is an integer. We call such a constraint a *simple polynomial constraint*. By introducing auxiliary variables that are equated with appropriate monomials we can decompose any polynomial constraint into a sequence of simple polynomial constraints. This allows us also to compute the integer interval domains of the auxiliary variable from the integer interval domains of the original variables. We apply then to the simple polynomial constraints the rules introduced in the previous section.

To see that the restriction to simple polynomial constraints can make a difference consider the constraint

$$100x \cdot y - 10y \cdot z = 212,$$

and ranges  $x, y, z \in [1..9]$ . We rewrite it into the sequence

$$u = x \cdot y, \quad v = y \cdot z, \quad 100u - 10v = 212,$$

where  $u, v$  are auxiliary variables, each with the domain  $[1..81]$ .

It is easy to check that the *POLYNOMIAL EQUALITY* rule introduced in the previous section does not yield any domain reduction when applied to the original constraint  $100x \cdot y - 10y \cdot z = 212$ . In the presence of the discussed optimization the domain of  $x$  gets reduced to  $[1..3]$ .

However, if we repeatedly apply the *POLYNOMIAL EQUALITY* rule to the simple polynomial constraint  $100u - 10v = 212$ , we eventually reduce the domain of  $u$  to the empty set (since this constraint has no integer solution in the ranges  $u, v \in [1..81]$ ) and consequently can conclude that the original constraint  $100x \cdot y - 10y \cdot z = 212$  has no solution in the ranges  $x, y, z \in [1..9]$ , without performing any search. Note that this effect still occurs if we replace one occurrence of  $y$  by a fresh variable with the same domain.

As noted in [8], decomposing constraints also prevents the evaluation of subexpressions whose domains did not change, which may reduce the number of interval arithmetic operations performed during constraint propagation. In our case duplicate occurrences of variables are removed, so the reduction rules additionally become idempotent. However, this can be seen as a side-effect: rules still update variables that they depend on, only now this update is indirect, through other variables.

Consider for example the constraint  $x^3 \cdot y - x \leq 40$  of Section 5. If we rewrite this constraint as  $u - x \leq 40$ , with  $u = x^3 \cdot y$  and  $x, y \in [1..100]$ , then via  $u \leq x + 40$  we can set the upper bound for  $u$  to 140. Via  $x = \sqrt[3]{u/y}$  we can then set the upper for  $x$  to 5. This allows us to set the upper bound for  $u$  to 45 via  $u \leq x + 40$ , etc. From this point of view the auxiliary variables, and the idempotence that they entail, can be seen as an optimization that prevents the evaluation of expressions that will not lead to further domain updates.

## 7 Constraint Propagation: Full Decomposition

In this third approach we focus on a small set of ‘atomic’ arithmetic constraints. We call an arithmetic constraint *atomic* if it is in one of the following two forms:

- a linear constraint,
- $x \cdot y = z$ .

Using appropriate transformation rules involving auxiliary variables we can decompose any arithmetic constraint to a sequence of atomic arithmetic constraints, similar to the decomposition of linear constraints into constraints on groups of three variables in `clp(FD)` [8]. In this transformation, as with partial decomposition, the auxiliary variables are equated with monomials, so we can easily compute their domains.

We explained already in Section 4 how to reason about linear constraints. For a treatment of disequalities see, e.g., [14, 20]. Next, we focus on the reasoning for the multiplication constraint  $x \cdot y = z$  in the presence of the non-empty ranges  $x \in D_x$ ,  $y \in D_y$  and  $z \in D_z$ . To this end we introduce the following three domain reduction rules:

### MULTIPLICATION 1

$$\frac{\langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle}{\langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D'_z \rangle}$$

### MULTIPLICATION 2

$$\frac{\langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle}{\langle x \cdot y = z ; x \in D'_x, y \in D_y, z \in D_z \rangle}$$

### MULTIPLICATION 3

$$\frac{\langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle}{\langle x \cdot y = z ; x \in D_x, y \in D'_y, z \in D_z \rangle}$$

where  $D'_z := D_z \cap \text{int}(D_x \cdot D_y)$ ,  $D'_x := D_x \cap \text{int}(D_z/D_y)$ , and  $D'_y := D_y \cap \text{int}(D_z/D_x)$ .

The way we defined the multiplication and the division of the integer intervals ensures that the *MULTIPLICATION* rules 1, 2, and 3 are equivalence preserving. Consider for example the *MULTIPLICATION 2* rule. Take some  $a \in D_x$ ,  $b \in D_y$  and  $c \in D_z$  such that  $a \cdot b = c$ . Then  $a \in \{x \in \mathcal{Z} \mid \exists z \in D_z \exists y \in D_y x \cdot y = z\}$ , so  $a \in D_z/D_y$  and a fortiori  $a \in \text{int}(D_z/D_y)$ . Consequently  $a \in D_x \cap \text{int}(D_z/D_y)$ . Because we also have  $D_x \cap \text{int}(D_z/D_y) \subseteq D_x$ , this shows that the *MULTIPLICATION 2* rule is equivalence preserving.

The following example shows an interaction between all three *MULTIPLICATION* rules.

**Example 7.1** Consider the CSP

$$\langle x \cdot y = z ; x \in [1..20], y \in [9..11], z \in [155..161] \rangle. \quad (4)$$

To facilitate the reading we underline the modified domains. An application of the *MULTIPLICATION 2* rule yields

$$\langle x \cdot y = z ; x \in [\underline{16..16}], y \in [9..11], z \in [155..161] \rangle$$

since, as already noted in Subsection 3.2,  $[155..161]/[9..11] = [16..16]$ , and  $[1..20] \cap \text{int}([16..16]) = [16..16]$ . Applying now the *MULTIPLICATION 3* rule we obtain

$$\langle x \cdot y = z ; x \in [16..16], y \in [\underline{10..10}], z \in [155..161] \rangle$$

since  $[155..161]/[16..16] = [10..10]$  and  $[9..11] \cap \text{int}([10..10]) = [10..10]$ . Next, by the application of the *MULTIPLICATION 1* rule we obtain

$$\langle x \cdot y = z ; x \in [16..16], y \in [10..10], z \in [\underline{160..160}] \rangle$$

since  $[16..16] \cdot [10..10] = [160..160]$  and  $[155..161] \cap \text{int}([160..160]) = [160..160]$ .

So using all three multiplication rules we could solve the CSP (4).  $\square$

Now let us clarify why we did not define the division of the sets of integers  $Z$  and  $Y$  by

$$Z/Y := \{z/y \in Z \mid y \in Y, z \in Z, y \neq 0\}.$$

The reason is that in that case for any set of integers  $Z$  we would have  $Z/\{0\} = \emptyset$ . Consequently, if we adopted this definition of the division of the integer intervals, the resulting *MULTIPLICATION 2* and *3* rules would not be anymore equivalence preserving. Indeed, consider the CSP

$$\langle x \cdot y = z ; x \in [-2..1], y \in [0..0], z \in [-8..10] \rangle.$$

Then we would have  $[-8..10]/[0..0] = \emptyset$  and consequently by the *MULTIPLICATION 2* rule we could conclude

$$\langle x \cdot y = z ; x \in \emptyset, y \in [0..0], z \in [-8..10] \rangle.$$

So we reached an inconsistent CSP while the original CSP is consistent.

The transformation to atomic constraints can strengthen the reduction. Consider for example the simple constraint

$$w \cdot x \cdot y \cdot z = 24$$

with  $w = 4$  and  $x, y, z \in [1..4]$ . Application of the *POLYNOMIAL EQUALITY* rule does not reduce any of the domains, but if we replace the constraint with

$$u \cdot v = t, \quad w \cdot x = u, \quad y \cdot z = v$$

with additional ranges  $t = 24$ ,  $u \in [4..16]$ , and  $v \in [1..16]$ , then by application of the *MULTIPLICATION 3* rule to  $u \cdot v = t$  we reduce the domain of  $v$  to [2..6]. Next, by application of the *MULTIPLICATION 2* rule to the same constraint we reduce the domain of  $u$  to [4..12], and finally by application of the *MULTIPLICATION 3* rule to  $w \cdot x = u$  we reduce the domain of  $x$  from [1..4] to [1..3]. Note, however, that this effect depends on the decomposition. If we had decomposed the constraint as

$$z \cdot (y \cdot (x \cdot w)) = 24$$

with an auxiliary variable introduced for each pair of matching brackets, then we would not have been able to reduce any of the domains of  $x$ ,  $y$ , and  $z$ .

In the remainder of the paper we will also consider variants of the full decomposition approach where we allow squaring and exponentiation as atomic constraints. For this purpose we explain the reasoning for the constraint  $x = y^n$  in the presence of the non-empty ranges  $x \in D_x$  and  $y \in D_y$ , and for  $n > 1$ . To this end we introduce the following two rules:

*EXPONENTIATION*

$$\frac{\langle x = y^n ; x \in D_x, y \in D_y \rangle}{\langle x = y^n ; x \in D'_x, y \in D_y \rangle}$$

*ROOT EXTRACTION*

$$\frac{\langle x = y^n ; x \in D_x, y \in D_y \rangle}{\langle x = y^n ; x \in D_x, y \in D'_y \rangle}$$

where  $D'_x := D_x \cap \text{int}(D_y^n)$ , and  $D'_y := \text{int}(D_y \cap \sqrt[n]{D_x})$ .

To prove that these rules are equivalence preserving suppose that for some  $a \in D_x$  and  $b \in D_y$  we have  $a = b^n$ . Then  $a \in D_y^n$ , so  $a \in \text{int}(D_y^n)$  and consequently  $a \in D_x \cap \text{int}(D_y^n)$ . Also  $b \in \sqrt[n]{D_x}$ , so  $b \in D_y \cap \sqrt[n]{D_x}$ , and consequently  $b \in \text{int}(D_y \cap \sqrt[n]{D_x})$ .

With exponentiation as an atomic constraint, full decomposition leads to idempotent rules, and the discussion at the end of Section 6 applies.

## 8 Relation to Hull and Box Consistency

In this section we relate the three approaches introduced above to the well-known methods for constraint propagation of arithmetic constraints on real variables, whose domains are represented by floating-point intervals. An overview of these methods is provided in [9]. *Floating-point intervals* are intervals of reals, with bounds from a finite set  $\mathcal{F} \subseteq \mathcal{R} \cup \{-\infty, \infty\}$  of floating-point numbers that contains representations  $-\infty$  and  $\infty$  for plus and minus infinity. For floating-point intervals, the counterpart of the  $\text{int}(\cdot)$  operation is the *hull* of a set of real numbers defined as the smallest floating-point interval containing

the set. Ideally, for an arithmetic constraint  $c$  on the variables  $x_1, \dots, x_n$  with respective floating-point interval domains  $D_1, \dots, D_n$  we would like to enforce **hull consistency**, which entails that for all  $i \in [1..n]$

$$D_i = \text{hull}(x_i \in \mathcal{R} \mid \exists x_1 \in D_1, \dots, x_{i-1} \in D_{i-1}, x_{i+1} \in D_{i+1}, \dots, x_n \in D_n \\ (x_1, \dots, x_n) \in C).$$

However, no efficient procedure exists for enforcing hull consistency on arbitrary arithmetic constraints. Therefore, the natural approach is to first decompose constraints into atomic constraints, each containing a single arithmetic operation. Maintaining hull consistency for the decomposed constraints can be done efficiently, using proof rules similar to the ones that we introduced, but hull consistency for the resulting decomposed CSP is a weaker notion of consistency than hull consistency for the original CSP.

Our full decomposition approach can be seen as the integer interval equivalent of the method for computing hull consistency for a decomposition that we just described, with the exception that linear constraints are not decomposed further. In the floating-point case, because of the accumulation of the rounding errors, the characterization of the resulting form of constraint propagation in terms of hull consistency is possible only if all constraints, including linear constraints, are decomposed into single-operator constraints.

To illustrate this consider the constraint  $x + y + z = w$  with the variables ranging over the floating-point intervals  $D_x, D_y, D_z$  and  $D_w$ . When we evaluate  $D_x + D_y + D_z$  using the floating-point interval arithmetic to compute an update for  $D_w$ , we have three options which two intervals to add first. Because the floating-point addition is non-associative, we actually compute the hull of a decomposition that has an extra variable added for either  $x + y$ ,  $x + z$  or  $y + z$ , and the resulting interval is potentially a proper superset of  $\text{hull}(D_x + D_y + D_z)$ . Moreover, different rewritings of the constraint correspond to different decompositions, and although this need not be a problem in practice, the resulting form of local consistency is no longer clearly defined.

In contrast, for integer intervals, we do not need to deal with the accumulation of the rounding errors and the linear constraints can be left intact. Our other two approaches can be seen as variants of the full decomposition approach that exploit this property further: for partial decomposition we allow more than one multiplication per proof rule, and in the direct approach the decomposition is not made explicit at all. Apart from these variations, all three approaches are the same in one important aspect: multiple occurrences of the same variable are treated as different variables. To illustrate this, consider the constraint  $x^3 + x = 0$ , with  $x \in [-1..1]$ . While  $x = 0$  is the unique solution, none of our three approaches will be able to reduce the domain of  $x$ . The reason is that the two occurrences of  $x$  are essentially treated as different variables in the reduction rules. This problem is known as the **dependency problem** of interval arithmetic.

In the context of constraints on reals [5] proposed to deal with the dependency problem using the notion of box consistency. It is a weaker notion of local

consistency than hull consistency, but is potentially stronger than hull consistency for the decomposition of a constraint into atomic constraints (see, e.g., [9]). Enforcing box consistency, as described in [21], consists of fixing the domains of all variables except one, and then narrowing the domain of this variable by iteratively instantiating it with subdomains at the boundary of the original domain, each time verifying consistency of the constraint in the presence of the domains of the other variables, and subtracting the subdomain from the original domain if the instantiation leads to a failure.

The second step of the ‘trial-and-prune’ procedure for enforcing box consistency that we just sketched can be implemented by enforcing hull consistency on a decomposition of the original constraint. So the procedure for enforcing box consistency can be seen as consisting of a number of procedures including the one that enforces hull consistency. One could apply the same technique to the arithmetic constraints on integer intervals, replacing the enforcement of hull consistency by one of our approaches to constraint propagation. This would lead to an integer equivalent of the box consistency. The efficiency of the resulting procedure depends on the choice of the underlying approach to constraint propagation, which provides another argument for the efficiency analysis of the approaches here considered.

## 9 A Characterization of the *MULTIPLICATION* Rules

It is useful to reflect on the effect of the proof rules used to achieve constraint propagation. In this section, by way of example, we focus on the *MULTIPLICATION* rules and characterize their effect using the notion of bounds consistency as defined in [16], limited to integer intervals. Let us recall first the definition that we adopt here to the multiplication constraint. Given an integer interval  $[l..h]$  we denote by  $[l, h]$  the corresponding real interval.

**Definition 9.1** The CSP  $\langle x \cdot y = z ; x \in [l_x..h_x], y \in [l_y..h_y], z \in [l_z..h_z] \rangle$  is called *bounds consistent* if

- $\forall a \in \{l_x, h_x\} \exists b \in [l_y, h_y] \exists c \in [l_z, h_z] a \cdot b = c,$
- $\forall b \in \{l_y, h_y\} \exists a \in [l_x, h_x] \exists c \in [l_z, h_z] a \cdot b = c,$
- $\forall c \in \{l_z, h_z\} \exists a \in [l_x, h_x] \exists b \in [l_y, h_y] a \cdot b = c.$  □

The following result entails that the *MULTIPLICATION* rules will not reduce a CSP beyond bounds consistency.

**Theorem 9.2 (Bounds consistency)** *Suppose a CSP  $\langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle$  with the integer interval domains is bounds consistent. Then it is closed under the applications of the *MULTIPLICATION* 1,2 and 3 rules.*

**Proof.** See the Appendix.  $\square$

This theorem shows that the *MULTIPLICATION* rules entail a notion of local consistency, say *M-consistency*, that is implied by bounds consistency. However, *M-consistency* does not imply bounds consistency. Here is an example. Consider the CSP

$$\langle x \cdot y = z ; x \in [-2..1], y \in [-3..10], z \in [8..10] \rangle.$$

It is not bounds consistent, since for  $y = -3$  no real values  $a \in [-2, 1]$  and  $c \in [8, 10]$  exist such that  $a \cdot (-3) = c$ . Indeed, it is easy to check that

$$\{y \in \mathcal{R} \mid \exists x \in [-2, 1] \exists z \in [8, 10] x \cdot y = z\} = (-\infty, -4] \cup [8, \infty).$$

However, this CSP is closed under the applications of the *MULTIPLICATION 1, 2* and *3* rules since

- $[8..10] \subseteq \text{int}([-2..1] \cdot [-3..10])$ , as  $\text{int}([-2..1] \cdot [-3..10]) = [-20..10]$ ,
- $[-2..1] \subseteq \text{int}([8..10]/[-3..10])$  as  $\text{int}([8..10]/[-3..10]) = [-10..10]$ , and
- $[-3..10] \subseteq \text{int}([8..10]/[-2..1])$  as  $\text{int}([8..10]/[-2..1]) = [-10..10]$ .

The following result clarifies that this example identifies the only cause of discrepancy between *M-consistency* and bounds consistency. Here, given an integer interval  $D := [l..h]$  we define  $\langle D \rangle := \{x \in \mathcal{Z} \mid l < x < h\}$ .

**Theorem 9.3 (Bounds consistency 2)** *Consider a CSP  $\phi := \langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle$  with non-empty integer interval domains and such that*

$$0 \in \langle D_x \rangle \cap \langle D_y \rangle \text{ implies } 0 \in D_z. \quad (5)$$

*Suppose  $\phi$  is closed under the applications of the *MULTIPLICATION 1, 2* and *3* rules. Then it is bounds consistent.*

**Proof.** See the Appendix.  $\square$

Consequently the *MULTIPLICATION* rules only fail to enforce bounds consistency for the constraint  $x \cdot y = z$  in case the domains of  $x$  and  $y$  are both of the form  $[l..h]$ , with  $l < 0$  and  $h > 0$  while  $z$  can assume either only positive numbers, or only negative numbers. Because the zeroes in the domains of  $x$  and  $y$  do not contribute to any solution, we can remedy this effect by temporarily splitting these domains in a positive interval and a negative interval. Bounds consistency for the constraint  $x \cdot y = z$  is then achieved by applying the *MULTIPLICATION* rules to the resulting subproblems, and updating the domain of each variable with the  $\text{int}(\cdot)$  closure of the union of its domain in these subproblems.

In [20] similar rules to our *MULTIPLICATION* rules are defined that apply this technique directly. They were defined without the use of interval arithmetic. It is also shown there that the *LINEAR EQUALITY* and *LINEAR INEQUALITY* rules enforce bounds consistency.

## 10 Implementation Details

### 10.1 Weak Division

We already mentioned in Section 3 that the division operation on the intervals does not admit an efficient implementation. The reason is that the  $\text{int}(\cdot)$  closure of the interval division  $[a..b]/[c..d]$  requires an auxiliary computation in case when  $0 \notin [c..d]$ . The preprocessing of  $[c..d]$  becomes impractical for small intervals  $[a..b]$ , and large  $[c..d]$ , occurring for example for the constraint  $\prod_{i=1}^n x_i = \prod_{i=1}^n i$ , of the benchmark problem mentioned in Subsection 1.1. This can be remedied by using the following variant of the division operation. We call it *weak division* since it yields a larger set (and so is ‘weaker’).

$$[a..b] : [c..d] := \begin{cases} [[\min(A)] .. [\max(A)]] & \text{if } 0 \notin [c..d], \text{ or} \\ & 0 \notin [a..b] \text{ and } 0 \in \{c, d\} \text{ and } c < d, \\ [a..b]/[c..d] & \text{otherwise} \end{cases}$$

where  $A = \{a/c', a/d', b/c', b/d'\}$ , and  $[c'..d'] = [c..d] - \{0\}$ .

Then  $\text{int}([a..b] : [c..d])$  can be computed by a straightforward case analysis already used for  $\text{int}([a..b]/[c..d])$  but now without any auxiliary computation. The weak division operator gives rise to the following versions of the *MULTIPLICATION* rules 2 and 3:

*MULTIPLICATION 2w*

$$\frac{\langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle}{\langle x \cdot y = z ; x \in D'_x, y \in D_y, z \in D_z \rangle}$$

*MULTIPLICATION 3w*

$$\frac{\langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle}{\langle x \cdot y = z ; x \in D_x, y \in D'_y, z \in D_z \rangle}$$

where  $D'_x := D_x \cap \text{int}(D_z : D_y)$ , and  $D'_y := D_y \cap \text{int}(D_z : D_x)$ .

In the assumed framework based on constraint propagation and tree search, all domains become eventually singletons or empty sets. It can easily be verified that both division operations are then equal, i.e.,  $[a..b] : [c..d] = [a..b]/[c..d]$ , for  $a \geq b$  and  $c \geq d$ . For this reason, we can safely replace any of the reduction rules introduced in this paper, notably *POLYNOMIAL EQUALITY*, *POLYNOMIAL INEQUALITY*, and *MULTIPLICATION* 2 and 3, by their counterparts based on the weak division. For the *MULTIPLICATION* rules specifically, the following theorem states that both sets of rules actually achieve the same constraint propagation.

**Theorem 10.1 (*MULTIPLICATION*)** *A CSP  $\langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle$  with the integer interval domains is closed under the applications of the *MULTIPLICATION* 1, 2 and 3 rules iff it is closed under the applications of the *MULTIPLICATION* 1, 2w and 3w rules.*

**Proof.** See the Appendix. □

Let us clarify now the relation between the *MULTIPLICATION* rules and the corresponding rules based on real interval arithmetic combined with the rounding of the resulting real intervals inwards to the largest integer intervals. The CSP  $\langle x \cdot y = z ; x \in [-3..3], y \in [-1..1], z \in [1..2] \rangle$ , which we already discussed in the introduction, shows that these approaches yield different results. Indeed, using the *MULTIPLICATION* rule 2 we can reduce the domain of  $x$  to  $[-2..2]$ , while the latter approach yields no reduction.

On the other hand, the applications of the *MULTIPLICATION* rules 2w and 3w to  $\langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle$  such that  $\text{int}(D_z : D_x) \neq \text{int}(D_z/D_x)$  and  $\text{int}(D_z : D_y) \neq \text{int}(D_z/D_y)$  (so in cases when the use of the weak interval division differs from the use of the interval division) do coincide with the just discussed approach based on real interval arithmetic and inward rounding. This is a consequence of the way the multiplication and division of real intervals are defined, see [15]. However, as we already stated in the introduction, we believe that the limited precision of floating-point interval arithmetic, and the modest overhead of arbitrary length integers justify a separate implementation of these rules for arithmetic constraints on integer intervals.

## 10.2 Implementation

### Platform

Our experiments were performed using OpenSolver [23], an experimental constraint solver based on constraint propagation and tree search. OpenSolver can be configured by software plug-ins in a number of predefined categories, corresponding to different aspects of constraint propagation and tree search, which makes it particularly well-suited for carrying out comparative studies of implementations of constraint solvers. The categories of plug-ins that are relevant for the experiments reported here are:

- variable domain types, which implement the domains of variables,
- domain reduction functions (DRFs), which correspond to the reduction rules,
- schedulers of DRFs, which determine the order in which the DRFs are applied,
- branching strategies, which split the search tree after constraint propagation has terminated, and
- several categories corresponding to different aspects of a search strategy that determine how to traverse a search tree.

All experiments were performed using the `IntegerInterval` variable domain type plug-in. Domains of this type consist of an indication of the type of the interval (bounded, unbounded, left/right-bounded, or empty), and a pair of

arbitrary length integer bounds. This plug-in, and the DRFs operating on it are built using the already mentioned GNU MP library, which provides arbitrary length integers and arithmetic operations on them, including operations for rounding the outcome of divisions and root extractions in the desired direction.

The branching strategy that we used selects variables using a chronological ordering in which the auxiliary variables come last. The domain of the selected variable is split into two subdomains using bisection, so the resulting search trees are binary trees. In all experiments we searched for all solutions, traversing the entire search tree by means of depth-first leftmost-first chronological backtracking.

For the experiments in this paper a DRF plug-in has been developed that implements the domain reduction rules discussed in the previous sections. Every constraint of a CSP is enforced by a number of instantiations of this DRF: one for each variable occurrence.

The scheduler plug-in that we used in the experiments maintains a flag per DRF, indicating whether the DRF is pending application or not. Initially, all DRFs are pending application. If the application of a DRF (or the branching strategy) modifies the domains of one or more variables, all DRFs whose output depends on these variables become pending application. Since in general —as illustrated by the example at the beginning of Section 5— the DRFs are non-idempotent, this may include the DRF that has just been applied. By default, the scheduler plug-in keeps cycling through the set of DRFs for a given CSP in a specified order, applying those DRFs that are pending application. The cycling stops when no DRF is pending application, or when the domain of a variable becomes empty.

### Scheduling of Reduction Rules

It was already shown in [22] that controlling the order in which variables are updated can improve the efficiency of constraint propagation algorithms, and for this purpose, our scheduler plug-in can be supplied with a *schedule*. Such a schedule is a sequence of indices into the set of DRFs that describes the order in which the scheduler will visit them, as an alternative to cycling. This is used in combination with full and partial decomposition, where we distinguish *user constraints* from the constraints that are introduced to define the values of auxiliary variables. Before considering for execution a DRF  $f$  that is part of the implementation of a user constraint, we make sure that all auxiliary variables that  $f$  relies on are updated. For this purpose, the indices of the DRFs that update these variables precede the index of  $f$  in the schedule. If  $f$  can change the value of an auxiliary variable, its index is followed by the indices of the DRFs that propagate back these changes to the variables that define the value of this auxiliary variable.

For example, rewriting  $x^3 \cdot y - x \leq 40$  to simple constraints introduces an auxiliary variable  $u$ , which is equated with  $x^3 \cdot y$ . This leads to five reduction rules: one for each occurrence of a variable after the rewriting step. We number these reduction rules as follows, where we underline in the constraint the variable

that is updated by the rule:

$$\begin{array}{lll}
 1. \underline{u} = x^3 \cdot y & 2. u = \underline{x}^3 \cdot y & 3. u = x^3 \cdot \underline{y} \\
 4. \underline{u} - x \leq 40 & & 5. u - \underline{x} \leq 40
 \end{array}$$

The fragment of the generated schedule that corresponds to enforcing the constraint  $x^3 \cdot y - x \leq 40$  is then 4,2,3,1,5. Rules 4 and 5 correspond to the original inequality, but rule 4 potentially modifies  $u$ , so in the schedule, rule 4 is followed by rules 2 and 3, that propagate any changes to  $u$  back to  $x$  and  $y$ . Before considering rule 5 for application, the schedule specifies that first rule 1 should be considered, so that any changes to the domains of  $x$  and  $y$  are propagated to the domain of  $u$ .

To see that an appropriate scheduling of the rules can be beneficial compared to cycling through the rules, suppose that all rules are pending application, and that  $D_x = D_y = [1..100]$ , and  $D_u = \mathcal{Z}$ . If we iterate the rules in their original order 1,2,3,4,5 then we first reduce  $D_u$  to  $[1..100^4]$  by means of rule 1. Next, rules 2 and 3 are executed without making any changes. Rule 4 then reduces  $D_u$  to  $[1..140]$ , which makes rules 2 and 3 pending application again. Next, rule 5 is executed without reducing  $D_x$ . Because  $x$  and  $y$  have not changed, rule 1 is not set to pending application, and rule 2 is the first rule that is applied in the second cycle, which reduces  $D_x$  to  $[1..5]$ . If use the generated schedule 4,2,3,1,5 instead, the same reduction is achieved immediately after applying the first two rules, instead of the six rules that are applied if we just cycle through the rules.

For full decomposition, there can be hierarchical dependencies between auxiliary variables. Much like the HC4revise procedure of [4], the generated schedule then specifies a bottom-up traversal of this hierarchy in a forward evaluation phase, and a top-down traversal in a backward propagation phase. These phases are performed before and after applying a DRF of a user constraint, respectively. In the forward evaluation phase, the DRFs that are executed correspond to the *MULTIPLICATION 1* and *EXPONENTIATION* rules. The DRFs of the backward propagation phase correspond to the *MULTIPLICATION 2* and *3*, and *ROOT EXTRACTION* rules. The HC4revise procedure is part of the HC4 algorithm, which enforces hull consistency for constraints on the reals using an implicit decomposition. For a discussion of this algorithm in the context of controlled constraint propagation, see [12].

### Constraint Rewriting

The proposed approaches were implemented by first rewriting arithmetic constraints to polynomial constraints, and then to a sequence of DRFs that correspond to the rules of the approach used. We implemented the following variants:

- $\mathbf{d_u}$  (direct, unoptimized): the direct approach, discussed in Section 5, where we isolate all variable occurrences in the original constraints without decomposing them first;

- d<sub>o</sub>** (direct, optimized): the optimization of the direct approach, discussed at the end of Section 5, which involves dividing out common powers of variables in the extended arithmetic expressions that arise from isolating the variable occurrences;
- p<sub>u</sub>** (partial, unoptimized): partial decomposition into simple constraints, as discussed in Section 6. The decomposition is implemented by introducing an auxiliary variable for every nonlinear power product. This procedure may introduce more auxiliary variables than necessary;
- p<sub>o</sub>** (partial, optimized): an optimized version of variant **p<sub>u</sub>**, where we stop introducing auxiliary variables as soon as the constraints contain no more duplicate occurrences of variables;
- f<sub>m</sub>** (full, multiplication): full decomposition into atomic constraints, as discussed in Section 7, allowing only linear constraints and multiplication as atomic constraints;
- f<sub>s</sub>** (full, squaring): idem, but also allowing  $x = y^2$  as an atomic constraint;
- f<sub>e</sub>** (full, exponentiation): idem, allowing  $x = y^n$  for all  $n > 1$  as an atomic constraint.

If the distinction between the different variants of an approach is irrelevant, we will sometimes omit the subscripts to the names **d**, **p**, and **f**.

Full and partial decomposition are implemented as a rewrite step, where the auxiliary variables are introduced. The resulting CSP is then rewritten using the direct approach. During the first rewrite step the hierarchical relations between the auxiliary variables are recorded, and the schedules are generated as a part of the second rewrite step. For variants **p<sub>o</sub>** and **f** the question of which auxiliary variables to introduce is an optimization problem in itself. Some choices result in more auxiliary variables than others. We have not treated this issue as an optimization problem but relied on the (somewhat arbitrary) heuristics described below. For this reason we have to consider the possibility that performance of variants **p<sub>o</sub>** and **f** can be further improved because in our experiments we used a suboptimal decomposition. The heuristics are as follows.

- For variant **p<sub>o</sub>** we replace nonlinear power products from left to right, so the rightmost nonlinear term of a polynomial constraint is always left intact.
- For the full decomposition approach, nonlinear power products are processed in the order in which they occur in the problem statement, after normalization to polynomial constraints. On the first occurrence of a nonlinear power product, we start introducing auxiliary variables for terms that divide the power product by multiplying or exponentiating existing variables, and keep doing so until we have introduced an auxiliary variable that corresponds to the full power product. When there are several choices

for which existing variables to multiply or exponentiate, we introduce an auxiliary variable for a term with the largest possible sum of exponents, thereby giving preference to exponentiation over multiplication, insofar as it is allowed by the variant. For variant  $\mathbf{f}_e$  we first introduce auxiliary variables for all exponentiations in the power product. For variant  $\mathbf{f}_s$ , we first introduce auxiliary variables for all exponentiations that divide the power product, and whose exponent is a power of 2. Unused auxiliary variables are deleted at a later stage.

To illustrate the latter heuristic, suppose we want to introduce an auxiliary variable for the term  $x^5 \cdot y^3 \cdot z$ . If we allow exponentiation, we start by introducing auxiliary variables  $u_1$  and  $u_2$  for the exponentiations in the term, and constrain them as follows:  $u_1 = x^5$ ,  $u_2 = y^3$ . Next we can introduce an auxiliary variable  $u_3$  for  $x^5 \cdot y^3$ ,  $x^5 \cdot z$ , or  $y^3 \cdot z$  by adding a constraint that multiplies two of the variables  $u_1$ ,  $u_2$ , and  $z$ . Because the sum of exponents is highest for the first option, we add  $u_3 = u_1 \cdot u_2$ . Finally  $u_4$  is introduced to replace the original term:  $u_4 = u_3 \cdot z$ . With only squaring allowed, we would be making these decisions in the presence of auxiliary variables for  $x^2$ ,  $x^4$ , and  $y^2$ , where  $x^4$  is obtained by squaring  $x^2$ . In this case, the first auxiliary variable introduced would be for  $x^4 \cdot y^2$ . With only multiplication allowed, after introducing  $u_1 = x \cdot x$  and  $u_2 = u_1 \cdot u_1$ , we would be expanding the term by repeatedly multiplying it with  $x$ ,  $y$ , or  $z$ .

Except for the optimized version of the direct approach, our current implementation can be optimized further by adopting the two-step propagation of linear constraints described in [14], as discussed at the end of Section 4. Because linear constraints are never decomposed, the effect is essentially the same for all alternatives that we discussed, so we have not considered this technique in our evaluation.

## 11 Experiments

### 11.1 Problems

For evaluating the alternative approaches, we used the integer problems described below. Problems with only integer variables and arithmetic constraints are rare in practice, and in that sense, our benchmark problems are artificial, but they serve well to generate a purely integer workload. The approach that works best on these problems can also be expected to work well in a hybrid setting, where integer variables and arithmetic constraints are mixed with other types of variables and constraints. In that case, only a fraction of the workload will be devoted to integer arithmetic, but with the results of this study we can be confident that we are dealing with these constraints in an efficient way.

**Cubes** The problem is to find all natural numbers  $n \leq 10^5$  that are a sum of four different cubes, for example

$$1^3 + 2^3 + 3^3 + 4^3 = 100.$$

This problem is modeled as follows:

$$\langle 1 \leq x_1, x_1 \leq x_2 - 1, x_2 \leq x_3 - 1, x_3 \leq x_4 - 1, x_4 \leq n, \\ x_1^3 + x_2^3 + x_3^3 + x_4^3 = n; n \in [1..10^5], x_1, x_2, x_3, x_4 \in \mathcal{Z} \rangle$$

**Opt** We are interested in finding a solution to the constraint  $x^3 + y^2 = z^3$  in the integer interval  $[1..10^5]$  for which the value of  $2x \cdot y - z$  is maximal.

**Fractions** This problem is taken from [19]: find distinct nonzero digits such that the following equation holds:

$$\frac{A}{BC} + \frac{D}{EF} + \frac{G}{HI} = 1$$

There is a variable for each letter. The initial domains are  $[1..9]$ . To avoid symmetric solutions an ordering is imposed:

$$\frac{A}{BC} \geq \frac{D}{EF} \geq \frac{G}{HI}$$

Also two redundant constraints are added:

$$3\frac{A}{BC} \geq 1 \quad \text{and} \quad 3\frac{G}{HI} \leq 1$$

Because division is not present in our arithmetic expressions, the above constraints are multiplied by the denominators of the fractions to obtain arithmetic constraints. We studied a representation for this problem using one equality and four inequalities for the ordering and the redundant constraints, and 36 disequalities  $A \neq B, A \neq C, \dots, H \neq I$ .

**Kyoto** The problem (see [10]) is to find the number  $n$  such that the alphanumeric equation

$$\begin{array}{r} \text{K Y O T O} \\ \text{K Y O T O} \\ + \text{K Y O T O} \\ \hline \text{T O K Y O} \end{array}$$

has a solution in the base- $n$  number system. Our representation uses a variable for each letter and one variable for the base number. The variables  $K$  and  $T$  may not be zero. There is one large constraint for the addition, 6 disequalities  $K \neq Y \dots T \neq O$  and four constraints stating that the individual digits  $K, Y, O, T$ , are smaller than the base number. To spend some CPU time, we searched base numbers 2..100.

**Sumprod** This is the problem cited in Subsection 1.1, for  $n = 14$ . We use the following representation:

$$\langle \begin{aligned} x_1 + \dots + x_n &= c_1 + \dots + c_n, \\ x_1 \cdot \dots \cdot x_n &= c_1 \cdot \dots \cdot c_n, \\ x_1 \leq x_2, x_2 \leq x_3, \dots, x_{n-1} &\leq x_n ; \\ x_1, \dots, x_n &\in [1..n], \\ c_1 \in \{1\}, c_2 \in \{2\}, \dots, c_n &\in \{n\} \end{aligned} \rangle$$

For  $n = 14$ , the value of the expression  $\prod_{i=1}^n i$  equals  $14!$ , which exceeds  $2^{32}$ , and to avoid problems with the input of large numbers, we used bound variables  $c_1, \dots, c_n$  and constraint propagation to evaluate it.

## 11.2 Results

Tables 2 and 3 compare the implemented variants of our approaches on the problems defined in the previous subsection. The first two columns of table 2 list the number of variables and DRFs that were used. Column nodes lists the size of the search tree, including failures and solutions. The next two columns list the number of times that a DRF was applied, and the percentage of these applications that the domain of a variable was actually modified. For the *opt* problem, the DRF that implements the optimization is not counted, and its application is not taken into account. The reported CPU times are user time in seconds, as reported by the UNIX `time` command on a 1200 MHz Athlon CPU. The last column compares the performance of our implementation to that of ECL<sup>2</sup>PS<sup>e</sup>, and will be discussed at the end of this section.

Table 3 lists measured numbers of basic interval operations. Note that for variant  $\mathbf{d}_o$ , there are two versions of the division and addition operations: one for integer intervals, and one for intervals of reals of which the bounds are rational numbers (marked  $\mathcal{Q}$ ). Columns multI and multF list the numbers of multiplications of two integer intervals, and of an integer interval and an integer factor, respectively. These are different operations in our implementation.

For the *cubes*, *opt*, and *sumprod* problems, the constraints are already in simple form, so variants  $\mathbf{d}_u$ ,  $\mathbf{d}_o$  and  $\mathbf{p}_o$  are identical. For *cubes* and *opt* all nonlinear terms involve a single multiplication or exponentiation, so for these experiments also variants  $\mathbf{p}_u$  and  $\mathbf{f}_e$  are the same. For the *fractions* problem, and for *sumprod*, no exponentiations are used, so all three variants of the full decomposition approach that we implemented are identical.

The results of these experiments clearly show the disadvantage of implementing exponentiation by means of multiplication: there is less domain reduction because we increase the number of variable occurrences (see the dependency problem, discussed in Section 8). For *opt* and variant  $\mathbf{f}_m$ , the run did not complete within reasonable time and was aborted.

For *fractions* the symbolic manipulation of variant  $\mathbf{d}_o$  reduces the search tree by a factor 0.70. However, this reduction is not reflected in the timings, and in fact the CPU time even increases. The reason is that computing the domain

	nvar	nDRF	nodes	DRFs applied	%eff.	CPU (sec.)	ECL <sup>i</sup> PS <sup>e</sup>
<i>cubes</i>							6.54s
<b>d<sub>u</sub>,p<sub>o</sub></b>	5	14	169,755	1,876,192	9.52	9.69	=
<b>p<sub>u</sub>,f<sub>e</sub></b>	9	22	169,755	2,237,590	16.28	6.53	=
<b>f<sub>m</sub></b>	13	34	206,405	3,011,749	20.02	8.53	-
<b>f<sub>s</sub></b>	13	34	178,781	2,895,717	20.62	8.80	-
<i>opt</i>							5752.70s
<b>d<sub>u</sub>,p<sub>o</sub></b>	4	7	115,469	5,187,002	42.16	21.55	+
<b>p<sub>u</sub>,f<sub>e</sub></b>	8	15	115,469	9,800,017	60.00	22.75	+
<b>f<sub>m</sub></b>	10	21	?	?	?	?	-
<b>f<sub>s</sub></b>	10	21	5,065,195	156,906,444	46.49	422.93	-
<i>fractions</i>							6.90s
<b>d<sub>u</sub></b>	9	154	11,289	1,193,579	3.65	15.40	=
<b>d<sub>o</sub></b>	9	154	7,879	734,980	3.45	17.38	=
<b>p<sub>u</sub></b>	37	210	11,289	1,410,436	23.27	4.89	=
<b>p<sub>o</sub></b>	32	200	11,289	1,385,933	21.65	5.25	=
<b>f</b>	43	208	11,131	1,426,204	27.76	4.98	=
<i>kyoto</i>							302.73s
<b>d<sub>u</sub></b>	5	37	87,085	3,299,814	6.09	21.84	=
<b>d<sub>o</sub></b>	5	37	87,085	3,288,461	5.94	44.56	+
<b>p<sub>u</sub></b>	13	53	87,085	3,781,514	23.02	10.93	=
<b>p<sub>o</sub></b>	12	51	87,085	3,622,461	21.45	11.24	=
<b>f<sub>m</sub></b>	16	60	87,087	4,276,066	26.70	10.40	=
<b>f<sub>s</sub></b>	16	60	87,085	4,275,957	26.70	10.39	=
<b>f<sub>e</sub></b>	16	59	87,085	3,746,532	23.26	9.42	=
<i>sumprod</i>							23.25s
<b>d<sub>u</sub>,p<sub>o</sub></b>	28	82	230,233	10,910,441	7.91	102.49	=
<b>p<sub>u</sub></b>	30	86	230,233	9,196,772	9.39	80.59	=
<b>f</b>	54	134	55,385	3,078,649	18.01	23.75	=

Table 2: Statistics and comparison with ECL<sup>i</sup>PS<sup>e</sup>

	root	exp	div	multI	multF	sum	total
<i>cubes</i>							
<b>d,p<sub>o</sub></b>	1,182	4,224	0	0	4,756	4,245	14,408
<b>p<sub>u</sub>,f<sub>e</sub></b>	180	181	0	0	4,756	4,245	9,363
<b>f<sub>m</sub></b>	0	0	589	438	4,927	4,363	10,317
<b>f<sub>s</sub></b>	192	198	384	198	4,842	4,305	10,121
<i>opt</i>							
<b>d,p<sub>o</sub></b>	2,299	4,599	1,443	1,444	11,064	5,187	26,037
<b>p<sub>u</sub>,f<sub>e</sub></b>	1,636	1,538	2,150	738	8,138	4,445	18,645
<b>f<sub>m</sub></b>	?	?	?	?	?	?	?
<b>f<sub>s</sub></b>	21,066	18,106	54,172	18,285	106,652	57,470	275,751
<i>fractions</i>							
<b>d<sub>u</sub></b>	0	0	868	28,916	14,238	13,444	57,466
<b>d<sub>o</sub></b>	0	0	51	11,892	8,010	6,727	29,584
			1,550 $\mathcal{Q}$			1,355 $\mathcal{Q}$	
<b>p<sub>u</sub></b>	0	0	734	933	4,736	4,669	11,071
<b>p<sub>o</sub></b>	0	0	776	1,509	5,292	5,147	12,725
<b>f</b>	0	0	693	339	4,835	4,769	10,636
<i>kyoto</i>							
<b>d<sub>u</sub></b>	735	11,041	1,963	13,853	10,853	13,946	52,390
<b>d<sub>o</sub></b>	735	8,146	218	8,955	12,516	10,592	48,749
			4,310 $\mathcal{Q}$			3,277 $\mathcal{Q}$	
<b>p<sub>u</sub></b>	383	759	1,591	484	5,324	7,504	16,044
<b>p<sub>o</sub></b>	383	759	1,597	1,360	5,756	8,008	17,863
<b>f<sub>m</sub></b>	0	0	1,991	578	5,324	7,505	15,398
<b>f<sub>s</sub></b>	< 0.5	< 0.5	1,990	578	5,324	7,504	15,397
<b>f<sub>e</sub></b>	1	1	1,554	484	5,324	7,504	14,868
<i>sumprod</i>							
<b>d,p<sub>o</sub></b>	0	0	4,032	100,791	85,419	149,479	339,721
<b>p<sub>u</sub></b>	0	0	2,186	27,948	81,728	149,479	261,340
<b>f</b>	0	0	609	205	25,799	46,960	73,573

Table 3: Measured numbers (thousands) of interval operations

updates involves adding intervals of real numbers. The arithmetic operations on such intervals are more expensive than their counterparts on integer intervals, because the bounds have to be maintained as rational numbers. Arithmetic operations on rational numbers are more expensive because they involve the computation of greatest common divisors. For *kyoto* the symbolic manipulation did not reduce the size of the search tree, so the effect is even more severe.

In general, the introduction of auxiliary variables leads to a reduction of the number of interval operations compared to the direct approach. As discussed at the end of Section 6, this is because auxiliary variables prevent the evaluation of subexpressions that did not change. This effect is strongest for *fractions*, where the main constraint contains a large number of different power products. Without auxiliary variables all power products are evaluated for every *POLYNOMIAL EQUALITY* rule defined by this constraint, even those power products the variable domains of which did not change. With auxiliary variables the intervals for such unmodified terms are available immediately, which leads to a significant reduction of the number of interval multiplications. For *sumprod*, the difference between variants **d** and **p<sub>u</sub>** is a bit artificial, because the operations that are saved involve the computation of the constant term  $c_1 \cdot \dots \cdot c_n$ . A comparable number of interval additions can be saved if we introduce a variable for the constant term  $c_1 + \dots + c_n$ . If we add these variables to the CSP all variants of the direct and partial decomposition approaches are essentially the same.

That stronger reduction is achieved as a result of full decomposition, mentioned in Section 7, is seen for the *fractions* benchmark and more prominently for *sumprod*. In the latter benchmark, this effect depends on a decomposition of the term  $\prod_{i=1}^n x_i$  as  $x_1 \cdot (x_2 \cdot (\dots (x_{n-1} \cdot x_n) \dots))$ , with an auxiliary variable for each pair of matching brackets. The decomposition then matches the chronological ordering used to select the variable for branching. If the ordering is reversed, the number of nodes is equal to that of the other approaches. The effect described in Section 6 is not demonstrated by these experiments.

If we do not consider the symbolic manipulation of variant **d<sub>o</sub>**, variant **f<sub>e</sub>** leads to the smallest total number of interval operations in all cases, but the scheduling mechanism discussed in Section 10 is essential for a consistent good performance. If for example the schedule is omitted for *opt*, the number of interval operations almost triples, and performance of variants **p<sub>u</sub>** and **f<sub>e</sub>** is then much worse than that of **d<sub>u</sub>**. This is conform the observations of [12], where it is demonstrated that for constraints on reals, enforcing hull consistency for a decomposition through repeated application of the HC4revise procedure yields superior performance compared to the basic HC3 algorithm. Based on these observations, we expect that the benefit of using the schedule will grow with the number of variables.

The total numbers of interval operations in table 3 do not fully explain all differences in elapsed times. One of the reasons is that different interval operations have different costs. Also some overhead is involved in applying a reduction rule, so if the number of applications differs significantly for two experiments, this influences the elapsed times as well (*opt*, **d**, **p<sub>u</sub>**). The elapsed times are

not the only measure that is subject to implementation details. For example, we implemented division by a constant interval  $[-1.. -1]$  as multiplication by a constant, which is more efficient in our implementation. Such decisions are reflected in the numbers reported in table 3.

For each of the benchmarks, the last column of Table 2 compares the performance of the variants that we implemented with that of an ECL<sup>i</sup>PS<sup>e</sup> [7] program that directly encodes the problem statement of Subsection 11.1 using the ic library. For each problem, the first entry in this column lists the CPU time reported by ECL<sup>i</sup>PS<sup>e</sup> for an all-solution search, where we applied the same branching scheme as we used in OpenSolver. The other entries compare propagation strength, for which we ran the solvers without search, and compared the resulting domains of the variables. A mark ‘=’ means that the computed domains are the same, ‘+’ that our variant achieved stronger reduction, and ‘-’ that constraint propagation is weaker than with ECL<sup>i</sup>PS<sup>e</sup>.

In addition, for *cubes* we verified that the number of nodes in the ECL<sup>i</sup>PS<sup>e</sup> search tree is identical to that for all variants except  $\mathbf{f}_m$  and  $\mathbf{f}_s$ , which nicely fits with the comparable CPU times. In contrast, for the *kyoto* benchmark, the number of nodes in the search tree is substantially lower for our approaches than for ECL<sup>i</sup>PS<sup>e</sup>, and so is the CPU time. For the *opt* problem the CPU time for our approaches (except  $\mathbf{f}_m$ ) is also substantially lower than for ECL<sup>i</sup>PS<sup>e</sup>. We have not verified the number of nodes visited by the `minimize/2` built-in, but the sequence of suboptimal solutions is identical to that found by our approaches (not verified for  $\mathbf{f}_m$ ). For this comparison we used ECL<sup>i</sup>PS<sup>e</sup> version 5.10.

## 12 Conclusions

In this paper we discussed a number of approaches to constraint propagation for arithmetic constraints on integer intervals. To assess them we implemented them using the OpenSolver framework of [23], and compared their performance on a number of benchmark problems. We can conclude that:

- Implementation of exponentiation by multiplication gives weak reduction. In the full decomposition approach  $x = y^n$  should be used as an atomic constraint.
- The optimization of the direct approach, where common powers of variables are divided out, can significantly reduce the size of the search tree, but the resulting reduction steps rely heavily on the division and addition of rational numbers. These operations are more expensive than their integer counterparts, because they involve the computation of greatest common divisors. As a result, our implementation of this approach was inefficient.
- Introducing auxiliary variables can be beneficial in two ways: it may strengthen constraint propagation, as discussed in Sections 6 and 7, and it may prevent the evaluation of subexpressions the variable domains of which did not change.

- As a result, given an appropriate scheduling of the rules, the full and partial decomposition approaches perform better than the direct approach without the optimization, in terms of numbers of interval operations. Actual performance depends on many implementation aspects. However for our test problems the performance of variants  $\mathbf{p}_u$ ,  $\mathbf{p}_o$  and  $\mathbf{f}_e$  does not differ much, except for one case where the decomposition of a single multiplication of all variables significantly reduced the size of the search tree.

Because of the inherent simplicity of the reduction rules and the potential reduction of the search tree, full decomposition of arithmetic constraints into multiplication, exponentiation, and linear constraints is our method of choice. However, a hierarchical scheduling of the resulting reduction rules is essential for efficient constraint propagation, and if a solver does not provide facilities for controlling the propagation order, the direct approach seems preferable.

Given that the optimization of the direct approach can achieve a significant reduction of the search tree, it would be interesting to combine it with full decomposition. Depending on the effect of the symbolic manipulation, a selection of the optimized rules that enforce a particular constraint according to variant  $\mathbf{d}_o$  could be used as redundant rules. In this case, the internal computations need not be precise, and we could maintain the rational bounds as floating-point numbers, thus avoiding the expensive computation of greatest common divisors.

## Acknowledgements

We would like to thank Maarten van Emden and the referees of earlier versions for useful comments.

## References

- [1] K. R. Apt. A proof theoretic view of constraint programming. *Fundamenta Informaticae*, 33(3):263–293, 1998. Available via <http://arXiv.org/archive/cs/>.
- [2] K. R. Apt. *Principles of Constraint Programming*. Cambridge University Press, 2003.
- [3] K. R. Apt and P. Zoetewij. A comparative study of arithmetic constraints on integer intervals. In K. R. Apt, F. Fages, F. Rossi, P. Szeredi, and J. Vánca, editors, *Recent Advances in Constraints*, volume 3010 of *Lecture Notes in Artificial Intelligence*, pages 1–24, 2004. Available via <http://xxx.lanl.gov/abs/cs.DS/0403016>.
- [4] F. Benhamou, F. Goualard, L. Granvilliers, and J.-F. Puget. Revising hull and box consistency. In *Proceedings of the 16th International Conference on Logic Programming (ICLP'99)*, pages 230–244. The MIT Press, 1999.

- [5] F. Benhamou, D. McAllester, and P. Van Hentenryck. CLP(intervals) revisited. In *Proceedings of the 1994 International Symposium on Logic programming*, pages 124–138. MIT Press, 1994.
- [6] F. Benhamou and W. J. Older. Applying interval arithmetic to real, integer, and Boolean constraints. *Journal of Logic Programming*, 32(1):1–24, 1997.
- [7] A. M. Cheadle, W. Harvey, A. J. Sadler, J. Schimpf, K. Shen, and M. G. Wallace. ECL<sup>i</sup>PS<sup>e</sup>: a Tutorial Introduction. Available from <http://www.eclipse-clp.org/>.
- [8] P. Codognet and D. Diaz. Compiling constraints in clp(fd). *Journal of Logic Programming*, 27(3):185–226, 1996.
- [9] H. Collaviza, F. Delobel, and M. Rueher. Comparing partial consistencies. *Reliable Computing*, 5(3):213–228, 1999.
- [10] V. Dubrovsky and A. Shvetsov. *Quantum Cyberteaser*. Available from <http://www.nsta.org/quantum/kyotoarc.asp>, May/June 1995.
- [11] *GNU MP: The GNU Multiple Precision Library*, September 2006. Edition 4.2.1, available from <http://www.gmp.lib.org>.
- [12] F. Goualard and L. Granvilliers. Controlled propagation in continuous numerical constraint networks. In *Proceedings of the 2005 ACM symposium on Applied computing*, pages 377–382. ACM Press, 2005.
- [13] L. Granvilliers. *RealPaver User’s Manual: Solving Nonlinear Constraints by Interval Computations*, August 2004. Edition 0.4, for RealPaver Version 0.4, available from <http://sourceforge.net/projects/realpaver>.
- [14] W. Harvey and P. J. Stuckey. Improving linear constraint propagation by changing constraint representation. *Constraints*, 8(2):173–207, 2003.
- [15] T. J. Hickey, Q. Ju, and M. H. van Emden. Interval arithmetic: from principles to implementation. *Journal of the ACM*, 48(5):1038–1068, 2001.
- [16] K. Marriott and P. Stuckey. *Programming with Constraints: An Introduction*. The MIT Press, 1998.
- [17] R. E. Moore. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1966.
- [18] D. Ratz. Inclusion isotone extended interval arithmetic: a toolbox update. Technical Report D-76128, University of Karlsruhe, 1996.
- [19] C. Schulte and G. Smolka. Finite domain constraint programming in Oz. A tutorial. Available from <http://www.mozart-oz.org/documentation/fdt/>, August 2002. Version 1.3.1 (20040616).
- [20] C. Schulte and P. J. Stuckey. When do bounds and domain propagation lead to the same search space? *Transactions on Programming Languages and Systems*, 27(3):388–425, May 2005.

- [21] P. Van Hentenryck, L. Michel, and Y. Deville. *Numerica: a Modeling Language for Global Optimization*. The MIT Press, 1997.
- [22] R. J. Wallace and E. F. Freuder. Ordering heuristics for arc consistency algorithms. Available from <http://4c.ucc.ie/web/pubs.jsp>, 1992.
- [23] P. Zoetewij. Opensolver: A coordination-enabled abstract branch-and-prune tree search engine (abstract). In F. Rossi, editor, *Proceedings of CP 2003*, volume 2833 of *LNCS*, page 1002. Springer-Verlag, 2003.

## Appendix

We provide here the proofs of the Bounds consistency Theorems 9.2 and 9.3, and the *MULTIPLICATION* Theorem 10.1.

### Proof of the Bounds consistency Theorem 9.2.

Let  $\phi := \langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle$ . Call a variable  $u$  of  $\phi$  *bounds consistent* if the bounds of its domain satisfy the condition of the bounds consistency (see Definition 9.1).

Given an integer interval  $[l..h]$  denote by  $\overline{[l..h]}$  the corresponding real interval  $[l, h]$ . Suppose that  $D_x = [l_x..h_x], D_y = [l_y..h_y], D_z = [l_z..h_z]$ . To show that  $\phi$  is closed under the applications of the *MULTIPLICATION 1* rule it suffices to prove that

$$\{l_z, h_z\} \subseteq \text{int}(D_x \cdot D_y). \quad (6)$$

So take  $c \in \{l_z, h_z\}$ . By the bounds consistency of  $z$  we have  $c = a \cdot b$  for some  $a \in \overline{D_x}$  and  $b \in \overline{D_y}$ . Since  $D_x$  and  $D_y$  are integer intervals we have  $\lfloor a \rfloor, \lceil a \rceil \in D_x$  and  $\lfloor b \rfloor, \lceil b \rceil \in D_y$ . To prove (6), by the definition of  $D_x \cdot D_y$ , we need to find  $a_1, a_2 \in D_x$  and  $b_1, b_2 \in D_y$  such that

$$a_1 \cdot b_1 \leq c \leq a_2 \cdot b_2.$$

The choice of  $a_1, a_2, b_1$  and  $b_2$  depends on the sign of  $a$  and of  $b$  and is provided in the following table:

condition	$a_1$	$b_1$	$a_2$	$b_2$
$a = 0$	$a$	$\lfloor b \rfloor$	$a$	$\lfloor b \rfloor$
$b = 0$	$\lfloor a \rfloor$	$b$	$\lfloor a \rfloor$	$b$
$a > 0, b > 0$	$\lfloor a \rfloor$	$\lfloor b \rfloor$	$\lceil a \rceil$	$\lceil b \rceil$
$a > 0, b < 0$	$\lceil a \rceil$	$\lfloor b \rfloor$	$\lfloor a \rfloor$	$\lceil b \rceil$
$a < 0, b > 0$	$\lfloor a \rfloor$	$\lceil b \rceil$	$\lceil a \rceil$	$\lfloor b \rfloor$
$a < 0, b < 0$	$\lceil a \rceil$	$\lceil b \rceil$	$\lfloor a \rfloor$	$\lfloor b \rfloor$

To prove that  $\phi$  is closed under the applications of the *MULTIPLICATION 2* and *3* rules it suffices to prove

$$\{l_x, h_x\} \subseteq \text{int}(D_z/D_y) \text{ and } \{l_y, h_y\} \subseteq \text{int}(D_z/D_x). \quad (7)$$

We need to distinguish a number of cases. The case analysis depends on the position of 0 w.r.t. each of the intervals  $D_x$  and  $D_y$ . This leads to 9 cases, which by symmetry between  $x$  and  $y$  can be reduced to 6 cases. We present here the proofs for representative 3 cases.

*Case 1.*  $l_x \geq 0$ ,  $l_y \geq 0$ .

By the bounds consistency of  $x$  for some  $b \in [l_y, h_y]$  we have  $l_x \cdot b \in [l_z, h_z]$ . Then  $b \leq h_y$  and  $l_x \geq 0$ , so  $l_x \cdot b \leq l_x \cdot h_y$ . Also  $l_z \leq l_x \cdot b$ , so

$$l_z \leq l_x \cdot h_y.$$

Next, by the bounds consistency of  $y$  for some  $a \in [l_x, h_x]$  we have  $a \cdot h_y \in [l_z, h_z]$ . Then  $l_x \leq a$  and  $h_y \geq 0$ , so  $l_x \cdot h_y \leq a \cdot h_y$ . Also  $a \cdot h_y \leq h_z$ , so

$$l_x \cdot h_y \leq h_z.$$

So  $l_x \cdot h_y \in [l_z, h_z]$  and consequently by the definition of the integer intervals division

$$l_x \in D_z/D_y \text{ and } h_y \in D_z/D_x.$$

By a symmetric argument

$$h_x \in D_z/D_y \text{ and } l_y \in D_z/D_x.$$

*Case 2.*  $l_x \geq 0$ ,  $h_y \leq 0$ .

By the bounds consistency of  $x$  for some  $b \in [l_y, h_y]$  we have  $h_x \cdot b \in [l_z, h_z]$ . Then  $b \leq h_y$  and  $h_x \geq 0$ , so  $h_x \cdot b \leq h_x \cdot h_y$ . Also  $l_z \leq h_x \cdot b$ , so

$$l_z \leq h_x \cdot h_y.$$

Next, by the bounds consistency of  $y$  for some  $a \in [l_x, h_x]$  we have  $a \cdot h_y \in [l_z, h_z]$ . Then  $a \leq h_x$  and  $h_y \leq 0$ , so  $a \cdot h_y \geq h_x \cdot h_y$ . Also  $h_z \geq a \cdot h_y$ , so

$$h_x \cdot h_y \leq h_z.$$

So  $h_x \cdot h_y \in [l_z, h_z]$  and consequently by the definition of the integer intervals division

$$h_x \in D_z/D_y \text{ and } h_y \in D_z/D_x.$$

Further, by the bounds consistency of  $x$  for some  $b \in [l_y, h_y]$  we have  $l_x \cdot b \in [l_z, h_z]$ . Then  $l_y \leq b$  and  $l_x \geq 0$ , so  $l_x \cdot l_y \leq l_x \cdot b$ . Also  $l_x \cdot b \leq h_z$ , so

$$l_x \cdot l_y \leq h_z.$$

Next, by the bounds consistency of  $y$  for some  $a \in [l_x, h_x]$  we have  $a \cdot l_y \in [l_z, h_z]$ . Then  $l_x \leq a$  and  $l_y < 0$ , so  $l_x \cdot l_y \geq a \cdot l_y$ . Also  $a \cdot l_y \geq l_z$ , so

$$l_z \leq l_x \cdot l_y.$$

So  $l_x \cdot l_y \in [l_z..h_z]$  and consequently by the definition of the integer intervals division

$$l_x \in D_z/D_y \text{ and } l_y \in D_z/D_x.$$

*Case 3.*  $l_x < 0 < h_x$ ,  $l_y \geq 0$ .

The proof for this case is somewhat more elaborate. By the bounds consistency of  $x$  for some  $b \in [l_y, h_y]$  we have  $l_x \cdot b \in [l_z, h_z]$ . Then  $l_y \leq b$  and  $l_x < 0$ , so  $l_x \cdot l_y \geq l_x \cdot b$ . But also  $l_x \cdot b \geq l_z$ , so

$$l_z \leq l_x \cdot l_y.$$

Next, by the bounds consistency of  $y$  for some  $a \in [l_x, h_x]$  we have  $a \cdot l_y \in [l_z, h_z]$ . Then  $l_x \leq a$  and  $l_y \geq 0$ , so  $l_x \cdot l_y \leq a \cdot l_y$ . But also  $a \cdot l_y \leq h_z$ , so

$$l_x \cdot l_y \leq h_z.$$

So  $l_x \cdot l_y \in [l_z..h_z]$  and consequently by the definition of the integer intervals division

$$l_x \in D_z/D_y \text{ and } l_y \in D_z/D_x.$$

Further, by the bounds consistency of  $x$  for some  $b \in [l_y, h_y]$  we have  $h_x \cdot b \in [l_z, h_z]$ . Then  $l_y \leq b$  and  $h_x > 0$ , so  $h_x \cdot l_y \leq h_x \cdot b$ . But also  $h_x \cdot b \leq h_z$ , so

$$h_x \cdot l_y \leq h_z.$$

Next, we already noted that by the bounds consistency of  $y$  for some  $a \in [l_x, h_x]$  we have  $a \cdot l_y \in [l_z, h_z]$ . Then  $a \leq h_x$  and  $l_y \geq 0$ , so  $a \cdot l_y \leq h_x \cdot l_y$ . But also  $l_z \leq a \cdot l_y$ , so

$$l_z \leq h_x \cdot l_y.$$

So  $h_x \cdot l_y \in [l_z..h_z]$  and consequently by the definition of the integer intervals division

$$h_x \in D_z/D_y.$$

It remains to prove that  $h_y \in D_z/D_x$ . We showed already  $l_x \cdot l_y \leq h_z$ . Moreover,  $l_x < 0$  and  $l_y \leq h_y$ , so  $l_x \cdot h_y \leq l_x \cdot l_y$  and hence

$$l_x \cdot h_y \leq h_z.$$

Also we showed already  $l_z \leq h_x \cdot l_y$ . Moreover  $h_x > 0$  and  $l_y \leq h_y$ , so  $h_x \cdot l_y \leq h_x \cdot h_y$  and hence

$$l_z \leq h_x \cdot h_y.$$

So if either  $l_z \leq l_x \cdot h_y$  or  $h_x \cdot h_y \leq h_z$ , then either  $l_x \cdot h_y \in [l_z..h_z]$  or  $h_x \cdot h_y \in [l_z..h_z]$  and consequently  $h_y \in D_z/D_x$ .

If both  $l_x \cdot h_y < l_z$  and  $h_z < h_x \cdot h_y$ , then

$$[l_z..h_z] \subseteq [l_x..h_x] \cdot h_y.$$

In particular for some  $a \in D_x$  we have  $l_z = a \cdot h_y$ , so  $h_y \in D_z/D_x$ , as well.

This concludes the proof for this case.  $\square$

**Proof of the Bounds consistency Theorem 9.3.**

We consider each variable in turn. We begin with  $x$ . Suppose that  $D_x = [l_x..h_x]$ .  $\phi$  is closed under the applications of the *MULTIPLICATION 2* rule, so

$$\{l_x, h_x\} \subseteq \text{int}(D_z/D_y). \quad (8)$$

To show the bounds consistency of  $x$  amounts to showing

$$\{l_x, h_x\} \subseteq \overline{D_z} \otimes \overline{D_y}. \quad (9)$$

(Recall that given real intervals  $X$  and  $Y$  we denote by  $X \otimes Y$  their division, defined in Section 3.)

*Case 1.*  $\text{int}(D_z/D_y) = \mathcal{Z}$ .

This implies that  $0 \in D_z \cap D_y$ , so by the definition of real intervals division  $\overline{D_z} \otimes \overline{D_y} = \langle -\infty, \infty \rangle$ . Hence (9) holds.

*Case 2.*  $\text{int}(D_z/D_y) \neq \mathcal{Z}$ .

So  $\text{int}(D_z/D_y)$  is an integer interval, say  $\text{int}(D_z/D_y) = [l_{zy}..h_{zy}]$ . Two subcases arise.

*Subcase 1.*  $\overline{D_z} \otimes \overline{D_y}$  is a, possibly open ended, real interval.

By (8) for some  $b_1, b_2 \in D_y$  and  $c_1, c_2 \in D_z$  we have

$$l_{zy} \cdot b_1 = c_1,$$

$$h_{zy} \cdot b_2 = c_2.$$

Let

$$\underline{b} := \min(b_1, b_2), \bar{b} := \max(b_1, b_2), \underline{c} := \min(c_1, c_2), \bar{c} := \max(c_1, c_2).$$

So  $\{l_{zy}, h_{zy}\} \subseteq [\underline{c}, \bar{c}] \otimes [\underline{b}, \bar{b}]$ . Also  $[\underline{c}, \bar{c}] \otimes [\underline{b}, \bar{b}] \subseteq \overline{D_z} \otimes \overline{D_y}$ . Hence  $\{l_{zy}, h_{zy}\} \subseteq \overline{D_z} \otimes \overline{D_y}$  and consequently, by the assumption for this subcase,  $[l_{zy}, h_{zy}] \subseteq \overline{D_z} \otimes \overline{D_y}$ . This proves (9) since by (8)  $\{l_x, h_x\} \subseteq [l_{zy}, h_{zy}]$ .

*Subcase 2.*  $\overline{D_z} \otimes \overline{D_y}$  is not a, possibly open ended, real interval.

In what follows for an integer interval  $D := [l..h]$  we write  $D > 0$  if  $l > 0$ ,  $D < 0$  if  $h < 0$ . Also recall that  $\langle D \rangle := \{x \in \mathcal{Z} \mid l < x < h\}$ .

This subcase can arise only when  $D_z > 0$  and  $0 \in \langle D_y \rangle$  or  $D_z < 0$  and  $0 \in \langle D_y \rangle$ , see [18] (reported as Theorem 4.8 in [15]), where the definition of the division of real intervals is considered.

Since  $\phi$  is closed under the *MULTIPLICATION* rule 3

$$D_y \subseteq \text{int}(D_z/D_x).$$

So  $\text{int}(D_z/D_x) \neq \emptyset$  since by assumption  $D_y$  is non-empty. Also, since  $0 \notin D_z$ , we have  $\text{int}(D_z/D_x) \neq \mathcal{Z}$ . So  $\text{int}(D_z/D_x)$  is a non-empty integer interval such that  $0 \in \langle \text{int}(D_z/D_x) \rangle$ .

But  $D_z > 0$  or  $D_z < 0$ , so if  $D_x > 0$ , then  $\text{int}(D_z/D_x) > 0$  or  $\text{int}(D_z/D_x) < 0$  and if  $D_x < 0$ , then  $\text{int}(D_z/D_x) > 0$  or  $\text{int}(D_z/D_x) < 0$ , as well. So  $0 \in \langle D_x \rangle$ . Hence  $0 \in \langle D_x \rangle \cap \langle D_y \rangle$  while  $0 \notin D_z$ . This contradicts (5). So this subcase cannot arise.

The proof for the variable  $y$  is symmetric to the one for the variable  $x$ .

Consider now the variable  $z$ .  $\phi$  is closed under the applications of the *MULTIPLICATION 1* rule, so

$$D_z \subseteq \text{int}(D_x \cdot D_y).$$

Take now  $c \in D_z$ . Then there exist  $a_1, a_2 \in D_x$  and  $b_1, b_2 \in D_y$  such that  $a_1 \cdot b_1 \leq c \leq a_2 \cdot b_2$ . We can assume that both inequalities are strict, that is,

$$a_1 \cdot b_1 < c < a_2 \cdot b_2, \quad (10)$$

since otherwise the desired conclusion is established.

Let

$$\underline{a} := \min(a_1, a_2), \bar{a} := \max(a_1, a_2), \underline{b} := \min(b_1, b_2), \bar{b} := \max(b_1, b_2).$$

We now show that  $a \in [\underline{a}, \bar{a}]$  and  $b \in [\underline{b}, \bar{b}]$  exist such that  $c = a \cdot b$ . Since  $[\underline{a}, \bar{a}] \subseteq \overline{D_x}$  and  $[\underline{b}, \bar{b}] \subseteq \overline{D_y}$ , this will establish the bounds consistency of  $z$ .

The choice of  $a$  and  $b$  depends on the signs of  $a_1$  and  $b_2$ . When one of these values is zero, the choice is provided in the following table, where in each case on the account of (10) no division by zero takes place:

condition	$a$	$b$
$a_1 = 0$	$c/b_2$	$b_2$
$a_2 = 0$	$c/b_1$	$b_1$
$b_1 = 0$	$a_2$	$c/a_2$
$b_2 = 0$	$a_1$	$c/a_1$

It is straightforward to show that in each case the quotient belongs to the corresponding interval. For example, when  $a_1 = 0$  we need to prove that  $c/b_2 \in [\underline{a}, \bar{a}]$ . By (10)  $a_2 \neq 0$ . If  $a_2 > 0$ , then again by (10),  $b_2 > 0$ , so  $c/b_2 \in [0..a_2]$ . In turn, if  $a_2 < 0$ , then also by (10)  $b_2 < 0$ , so, yet again by (10),  $c/b_2 \in [a_2..0]$ .

When neither  $a_1$  nor  $b_2$  is zero, the choice of  $a$  and  $b$  has to be argued case by case.

*Case 1.*  $a_1 > 0, b_2 > 0$ .

Then by (10)  $b_1 < c/a_1$  and  $c/b_2 < a_2$ . Suppose that both  $b_2 < c/a_1$  and  $c/b_2 < a_1$ . Then  $a_1 \cdot b_2 < c < a_1 \cdot b_2$ , which is a contradiction. So either  $c/a_1 \leq b_2$  or  $a_1 \leq c/b_2$ , that is either  $c/a_1 \in [b_1..b_2]$  or  $c/b_2 \in [a_1..a_2]$ .

*Case 2.*  $a_1 > 0, b_2 < 0$ .

Then by (10)  $b_1 < c/a_1$  and  $a_2 < c/b_2$ . Suppose that both  $b_2 < c/a_1$  and  $a_1 < c/b_2$ . Then  $a_1 \cdot b_2 < c < a_1 \cdot b_2$ , which is a contradiction. So either  $c/a_1 \leq b_2$  or  $c/b_2 \leq a_2$ , that is either  $c/a_1 \in [b_1..b_2]$  or  $c/b_2 \in [a_2..a_1]$ .

*Case 3.*  $a_1 < 0$ ,  $b_2 > 0$ .

Then by (10)  $c/a_1 < b_1$  and  $c/b_2 < a_2$ . Suppose that both  $c/a_1 < b_2$  and  $c/b_2 < a_1$ . Then  $a_1 \cdot b_2 < c < a_1 \cdot b_2$ , which is a contradiction. So either  $b_2 \leq c/a_1$  or  $a_1 \leq c/b_2$ , that is either  $c/a_1 \in [b_2..b_1]$  or  $c/b_2 \in [a_1..a_2]$ .

*Case 4.*  $a_1 < 0$ ,  $b_2 < 0$ .

Then by (10)  $c/a_1 < b_1$  and  $a_2 < c/b_2$ . Suppose that both  $c/a_1 < b_2$  and  $a_1 < c/b_2$ . Then  $a_1 \cdot b_2 < c < a_1 \cdot b_2$ , which is a contradiction. So either  $b_2 \leq c/a_1$  or  $c/b_2 \leq a_1$ , that is either  $c/a_1 \in [b_2..b_1]$  or  $c/b_2 \in [a_2..a_1]$ .

So in each of the four cases we can choose either  $a := a_1$  and  $b := c/a_1$  or  $a := c/b_2$  and  $b := b_2$ .  $\square$

**Proof of the *MULTIPLICATION* Theorem 10.1.**

The weak interval division produces larger sets than the interval division. As a result the *MULTIPLICATION* rules 2w and 3w yield a weaker reduction than the original *MULTIPLICATION* rules 2 and 3. So it suffices to prove that  $\phi := \langle x \cdot y = z ; x \in D_x, y \in D_y, z \in D_z \rangle$  is closed under the applications of the *MULTIPLICATION* 1, 2 and 3 rules assuming that it is closed under the applications of the *MULTIPLICATION* 1, 2w and 3w rules. Suppose that  $D_x = [l_x..h_x]$ ,  $D_y = [l_y..h_y]$ ,  $D_z = [l_z..h_z]$ . The assumption implies

$$\{l_x, h_x\} \subseteq \text{int}(D_z : D_y) \tag{11}$$

and

$$\{l_y, h_y\} \subseteq \text{int}(D_z : D_x) \tag{12}$$

The proof is by contradiction. Assume that (11) and (12) hold, while  $\phi$  is not closed under application of *MULTIPLICATION* 2 and 3. Without loss of generality, suppose that *MULTIPLICATION* 2 is the rule that can make a further reduction. This is the case iff

$$\text{int}(D_z/D_y) \subset \text{int}(D_z : D_y).$$

By definition, the proper inclusion implies that  $l_y \geq 0$  or  $h_y \leq 0$ . Assume  $l_y \geq 0$ , the case for  $h_y \leq 0$  is similar. Let  $l'_y := \max(1, l_y)$ , and let  $A := \{l_z/l'_y, l_z/h_y, h_z/l'_y, h_z/h_y\}$ , and  $B := \{l_z/l_x, l_z/h_x, h_z/l_x, h_z/h_x\}$ . A further implication of the proper inclusion is that one or both of  $l'_y$  and  $h_y$  do not have a multiple in  $D_z$ : otherwise  $\min(A)$  and  $\max(A)$  would be elements of  $D_z/D_y$ , and we would have  $\text{int}(D_z : D_y) = \text{int}(D_z/D_y)$ . The cases for  $l'_y$  and  $h_y$  can be seen in isolation, and their proofs are similar, so here we only consider the case that  $l'_y$  does not have a multiple in  $D_z$ . In what follows we can assume  $0 \notin D_z$ , since otherwise  $l'_y$  and  $h_y$  do have a multiple in  $D_z$ .

*Case 1.*  $l_z > 0$ .

From (11) it follows that  $h_x \leq \lfloor \max(A) \rfloor$ , which for the case  $l'_y, h_y, l_z, h_z > 0$

that we consider here implies  $h_x \leq \lfloor h_z/l'_y \rfloor$ . Because  $\lfloor l_z \cdot h_z \rfloor$  does not contain a multiple of  $l'_y$ , we have  $\lfloor h_z/l'_y \rfloor = \lfloor l_z/l'_y \rfloor$ , so

$$h_x \leq \lfloor l_z/l'_y \rfloor.$$

A further consequence of (11) is that  $l_x, h_x > 0$ . From (12) it follows that  $l'_y \geq \lceil \min(B) \rceil$ , which for  $l_x, l_z > 0$  implies

$$l'_y \geq \lceil l_z/h_x \rceil \geq l_z/h_x \geq l_z/\lfloor h_z/l'_y \rfloor.$$

Because  $l'_y$  is no divisor of  $l_z$ , and both numbers are positive, we have  $\lfloor l_z/l'_y \rfloor < l_z/l'_y$ , and consequently  $l'_y > l_z/(l_z/l'_y)$ , leading to  $l'_y > l'_y$ , which is a contradiction.

*Case 2.  $h_z < 0$ .*

Similarly, because  $l'_y, h_y > 0$  and  $l_z, h_z < 0$ , it follows from (11) that  $l_x \geq \lceil \min(A) \rceil = \lceil l_z/l'_y \rceil$ , and  $l_x, h_x < 0$ . Because  $\lfloor l_z \cdot l_h \rfloor$  does not contain a multiple of  $l'_y$ , we have  $\lceil l_z/l'_y \rceil = \lceil h_z/l'_y \rceil$ , so

$$l_x \geq \lceil h_z/l'_y \rceil.$$

We use this information in the following implication of (12):

$$l'_y \geq \lceil \min(B) \rceil = \lceil h_z/l'_x \rceil \geq h_z/l'_x$$

to get  $l'_y \geq h_z/\lceil h_z/l'_y \rceil$ . Because  $|\lceil h_z/l'_y \rceil| < |h_z/l'_y|$ , we have  $l'_y > h_z/(h_z/l'_y)$ , leading to  $l'_y > l'_y$ , which is a contradiction.  $\square$