Integer Arithmetic

Syntax and Semantics

The integer arithmetic (IA) is the first order theory of integer numbers.

The alphabet of the integer arithmetic consists of:

- function symbols $+, \cdot, s$ (s is the successor function $n \mapsto n+1$)
- constant symbol 0

The semantics of IA is defined in the structure $\mathfrak{Z} = \langle \mathbb{Z}, +, \cdot, n \mapsto n+1 \rangle$.

Examples

Ex: Write a formula $pos(x)$ that holds if and only if $x \geq 0$

The order relation is defined as

$$x \leq y \ : \ \exists z \ . \ pos(z) \wedge x + z = y$$

The set of even numbers is defined by

$$even(x) \ : \ \exists y \ . \ x = y + y$$

The divisibility relation is defined as

$$x | y \ : \ \exists z \ . \ y = x \cdot z$$

Examples

The set of prime numbers is defined by

$$prime(x) \ : \ \forall y \forall z \ . \ x = y \cdot z \longrightarrow (y = 1 \vee z = 1)$$

The least common multiple is defined as

$$z = lcm(x, y) \ : \ \forall t \ . \ x|t \ \wedge \ y|t \leftrightarrow z|t$$

Goldbach's Conjecture

$$\forall x \ . \ 2 \leq x \wedge even(x) \longrightarrow \exists y \exists z \ . \ prime(y) \wedge prime(z) \wedge x = y + z$$

Peano Arithmetic

An axiomatic theory is a set of formulae in which truth is derived from a (possibly infinite) set of axioms, e.g. Euclid's geometry is an axiomatic theory.

1. $0 \neq s(x)$
2. $s(x) = s(y) \rightarrow x = y$
3. $x + 0 = x$
4. $x + s(y) = s(x + y)$
5. $x \cdot 0 = 0$
6. $x \cdot s(y) = x \cdot y + x$
7. $\varphi(0) \wedge \forall x \;.\; [\varphi(x) \rightarrow \varphi(s(x))] \rightarrow \forall x \;.\; \varphi(x)$

Notice that the last point defines an infinite number of axioms.

Undecidability of Integer Arithmetic

Follows directly from Gödel's Incompleteness Theorem:

Kurt Gödel. *Uber formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme* I. Monatshefte für Mathematik und Physik, 38:173 198, 1931.

Alonzo Church. *An unsolvable problem of elementary number theory*. American Journal of Mathematics, 58:345 363, 1936.

Undecidability of Integer Arithmetic

The quantifier-free fragment is also undecidable:

Yuri Matiyasevich. *Enumerable sets are diophantine.* Journal of Sovietic Mathematics, (11):354 358, 1970.

Undecidability of Hilbert's Tenth Problem:

Given a Diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: To devise a process according to which it can be determined in a finite number of operations whether the equation is solvable in rational integers.

Undecidability of Integer Arithmetic

Undecidability of the arithmetic of addition and divisibility:

$$
\begin{array}{rcl}
z = lcm(x,y) & : & \forall t \,.\, x|t \ \wedge \ y|t \leftrightarrow z|t \\
x^2 & = & lcm(x, x+1) - x \\
4 \cdot x \cdot y & = & (x+y)^2 - (x-y)^2
\end{array}
$$

Consequently, the arithmetic of addition and

- least common multiple
- square function

are undecidable.

Presburger Arithmetic

Definition

PA is the additive theory of natural numbers $\langle \mathbb{N}, 0, s, + \rangle$

PA is decidable

Mojzesz Presburger. *Über die Vollstandigkeit eines gewissen Systems der Arithmetik*. Comptes rendus du I Congrès des Pays Slaves, Warsaw 1929.

Examples

Even/Odd:

$$\begin{aligned} even(x) &: \exists y \,.\, x = y + y \\ odd(x) &: \exists y \,.\, even(y) \;\wedge\; x = s(y) \end{aligned}$$

Order:

$$x \leq y \;:\; \exists z \,.\, x + z = y$$

Zero/One:

$$\begin{aligned} zero(x) &: \forall y \,.\, x \leq y \\ one(x) &: \exists z \,.\, zero(z) \wedge \neg x = z \;\wedge\; \forall y \,.\, y = z \vee x \leq y \end{aligned}$$

Modulo constraints:

$$x \equiv_m y \;:\; \exists z \,.\, (x \leq y \wedge y - x = mz) \;\vee\; (x > y \wedge x - y = mz)$$

Quantifier Elimination in PA

A theory admits quantifier elimination if any formula of the form $Q_1 x_1 \ldots Q_n x_n \,.\, \phi(x_1, \ldots, x_n, y_1, \ldots, y_m)$ is equivalent (modulo the theory) to a quantifier-free formula $\psi(y_1, \ldots, y_m)$.

We consider the (equivalent) theory of addition and modulo constraints

$$x \equiv_m y \;:\; \exists z \,.\, (x \leq y \wedge y - x = mz) \;\vee\; (x > y \wedge x - y = mz)$$

Given a PA formula $\exists x \,.\, \phi(x, y_1, \ldots, y_m)$, we build an equivalent formula $\psi(y_1, \ldots, y_m)$ in the new language (with modulo constraints)

Quantifier Elimination in PA

1. **Eliminate the negations**

- replace $\neg(t_1 = t_2)$ by $t_1 < t_2 \vee t_2 < t_1$,
- replace $\neg(t_1 < t_2)$ by $t_1 = t_2 \vee t_2 < t_1$, and
- replace $\neg(t_1 \equiv_m t_2)$ by $\bigvee_{i=1}^{m-1} t_1 \equiv_m t_2 + i$.

Then rewrite the formula into DNF, i.e. a disjunction of $\exists x \,.\, \beta_1 \wedge \ldots \wedge \beta_n$, where each β_i is one of the following forms:

$$\begin{array}{rcl} nx & = & u - t \\ nx & \equiv_m & u - t \\ nx & < & u - t \\ u - t & < & nx \end{array}$$

Quantifier Elimination in PA

2. **Uniformize the coefficients of** x

Let p be the least common multiple of the coefficients of x.

Multiply each atomic formula containing nx by $\frac{p}{n}$.

In particular, $nx \equiv_m u - t$ becomes $px \equiv_{\frac{p}{n}m} \frac{p}{n}(u - t)$.

Quantifier Elimination in PA

Eliminate the coefficients of x Replace all over the formula px by x and add the new conjunct $x \equiv_p 0$

Special case If $x = u - t$ occurs in the formula, eliminate directly x by replacing it with $u - t$.

Quantifier Elimination in PA

Assume $x = u - t$ **does not occur**.

We have a formula of the form

$$\exists x \ . \ \bigwedge_{j=1}^{l} r_j - s_j < x \ \wedge \ \bigwedge_{i=1}^{k} x < t_i - u_i \ \wedge \ \bigwedge_{i=1}^{n} x \equiv_{m_i} v_i - w_i$$

Let $M = [m_i]_{i=1}^{n}$. The formula is equivalent to:

$$\bigvee_{q=1}^{M} \Big[\bigwedge_{i=1}^{l} \Big(\bigwedge_{j=1}^{k} (r_j - s_j) + q < t_i - u_i \ \wedge \ \bigwedge_{i=1}^{n} (r_j - s_j) + q \equiv_{m_i} v_i - w_i \Big) \Big]$$

Example

$$\exists x \ . \ 1 < x \ \wedge \ x < 100 \ \wedge \ x \equiv_2 1 \ \wedge \ x \equiv_3 2$$

$$x \in [2, 99] \wedge x \equiv_2 1 \quad : \qquad 3 \quad \color{red}{5} \quad 7 \quad 9 \quad \color{red}{11} \quad 13 \quad 15 \quad \color{red}{17} \ \ldots$$

$$x \in [2, 99] \wedge x \equiv_3 2 \quad : \quad 2 \qquad \color{red}{5} \qquad 8 \qquad \color{red}{11} \qquad 14 \qquad \color{red}{17} \ \ldots$$

$$\bigvee_{q=1}^{6} \big(1 + q < 100 \ \wedge \ 1 + q \equiv_2 1 \ \wedge \ 1 + q \equiv_3 2\big)$$

Decidability of PA

The result quantifier elimination in a Presburger formula is equivalent to a disjunction of conjunctions of atomic propositions of the following forms:

$$
\begin{aligned}
\sum_{i=1}^{n} a_i x_i + b &\quad \geq \quad 0 \\
\sum_{i=1}^{n} a_i x_i + b &\quad \equiv_n \quad m
\end{aligned}
$$

If all quantifiers are eliminated from a formula with no free variables, the result is either true of false.

Semilinear Sets

Preliminaries

Let $\mathbf{x}, \mathbf{y} \in \mathbb{N}^{\mathbf{n}}$, for some $n > 0$

$$\begin{aligned} \mathbf{x} &= \langle x_1,\ x_2,\ \ldots,\ x_n \rangle \\ \mathbf{y} &= \langle y_1,\ y_2,\ \ldots,\ y_n \rangle \end{aligned}$$

We define the following operations:

$$\mathbf{x} + \mathbf{y} \quad = \quad \langle x_1 + y_1,\ x_2 + y_2,\ \ldots,\ x_n + y_n \rangle$$

$$a\mathbf{x} \quad = \quad \langle ax_1,\ ax_2,\ \ldots,\ ax_n \rangle,\ a \in \mathbb{N}$$

$$\mathbf{x} \leq \mathbf{y} \quad \Longleftrightarrow \quad x_1 \leq y_1 \ \wedge\ x_2 \leq y_2 \ \wedge\ \ldots\ \wedge\ x_n \leq y_n$$

Preliminaries

Lemma 1 *Each set of pairwise incomparable elements of $\mathbb{N}^n$ is finite. In consequence, each set $M \subseteq \mathbb{N}^n$ has a finite number of minimal elements.*

A strict order $\prec$ is called well-founded if there are no infinite descending chains $x_1 \succ x_2 \succ \ldots$. For example, $<$ is well-founded on $\mathbb{N}^n$.

Principle 1 (Well-founded Induction) *Let $\langle W, \preceq \rangle$ be a well-founded set, and P a property of the elements of W. If both the following hold:*

1. *P is true for all minimal elements of W,*
2. *for all $x \in W$: if $P(y)$ is true for all $y \prec x$ then $P(x)$ is true*

then, for all $x \in W$, $P(x)$ is true.

Linear Sets

$\mathcal{L}(C,P) = \{c + p_1 + \ldots + p_m \mid c \in C,\ p_1,\ldots,p_m \in P\}$ for some $C, P \in \mathbb{N}^n$

- C = set of constants (bases)
- P = set of periods (generators)

An element $x \in \mathcal{L}(C,P)$ is of the form $x = c + \sum_{i=1}^{m} \lambda_i p_i$, where $c \in C$, $\lambda_i \in \mathbb{N}$ and $p_i \in P$, for all $1 \leq i \leq m$.

A set $M \in \mathbb{N}^n$ is said to be linear if $M = \mathcal{L}(\{c\}, P)$ where:

- $c \in \mathbb{N}^n$
- $P \subseteq \mathbb{N}^n$ is finite

Examples

$$\mathcal{L}(\{(1,0)\},\{(1,2),(3,2)\}) = \{(2,2),(4,2),(3,4),(5,4),(7,4),\ldots\}$$

$$\{(x,y) \mid x \geq 1\} = \mathcal{L}(\{(1,0)\},\{(1,0),(0,1)\})$$

Semilinear Sets

A set S is semilinear if it is a finite union of linear sets.

Example 1 *$\mathcal{L}(C, P)$ is semilinear iff $C, P \subseteq \mathbb{N}^n$ are finite.*

A function $f : \mathbb{N}^n \longrightarrow \mathbb{N}^m$ is said to be linear if for all $x, y \in \mathbb{N}^n$ we have $f(x + y) = f(x) + f(y)$.

Lemma 2 *If $M \subseteq \mathbb{N}^m$ is a semilinear set and $f : \mathbb{N}^m \longrightarrow \mathbb{N}^n$ is a linear function, $m, n > 0$, then $f(M)$ is a semilinear set.*

Lemma 3 *If $M \subseteq \mathbb{N}^m$ is a semilinear set and $c \in \mathbb{N}^m$, then the set $c + M = \{c + x \mid x \in M\}$ is semilinear.*

Counterexample

$M = \{(x,y) \mid y \leq x^2\}$ is not semilinear

Suppose $M = \bigcup_{i=1}^{k} \mathcal{L}(c_i, P_i)$

Let $m = \max\{\frac{y}{x} \mid (x,y) \in \bigcup_{i=1}^{k} P_i\}$

Take $x_1, x_2 > m$. The slope of the line connecting (x_1, x_1^2) and (x_2, x_2^2) is $x_1 + x_2 > 2m > m$. Hence at most one of (x_1, x_1^2), (x_2, x_2^2) can be generated by $P_i, i = 1, \ldots, k$, contradiction.

Closure Properties of Semilinear Sets

Theorem 1 *The class of semilinear subsets of* $\mathbb{N}^n$, $n > 0$ *is effectively closed under union, intersection and projection.*

The most difficult is to show closure under intersection. It is enough to show that the intersection of two linear sets is semilinear.

Closure Properties of Semilinear Sets

Let $M = \mathcal{L}(c, \{p_1, \ldots, p_k\})$ and $M' = \mathcal{L}(c', \{p'_1, \ldots, p'_\ell\})$.

$$A \triangleq \{\langle \lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_\ell \rangle \mid c + \sum_{i=1}^{k} \lambda_i p_i = c' + \sum_{j=1}^{\ell} \mu_j p'_j\}$$

$$B \triangleq \{\langle \lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_\ell \rangle \mid \sum_{i=1}^{k} \lambda_i p_i = \sum_{j=1}^{\ell} \mu_j p'_j,\ \ \lambda_i > 0,\ \ \mu_j > 0\}$$

$$f(\langle \lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_\ell \rangle) \triangleq \sum_{i=1}^{k} \lambda_i p_i$$

f is a linear function, and $M \cap M' = c + f(A)$.

It is enough to prove that A is semilinear.

Closure Properties of Semilinear Sets

Let C, P be the sets of minimal elements of A, B.

Proposition 1 *Each element of B is a sum of elements of P.*

By well-founded induction. If $\mathbf{x} \in \mathbf{B}$ is a minimal element, then $\mathbf{x} \in \mathbf{P}$.

Else, $\mathbf{x} = \langle \lambda_{\mathbf{1}}, \ldots, \lambda_{\mathbf{k}}, \mu_{\mathbf{1}}, \ldots, \mu_\ell \rangle$ has a minimal element $\mathbf{x}' = \langle \lambda'_{\mathbf{1}}, \ldots, \lambda'_{\mathbf{k}}, \mu'_{\mathbf{1}}, \ldots, \mu'_\ell \rangle \in \mathbf{P}$ s.t. $\mathbf{x}' < \mathbf{x}$.

$$\begin{aligned}
\sum_{i=1}^{k} \lambda_i p_i &= \sum_{j=1}^{\ell} \mu_j p'_j \\
\sum_{i=1}^{k} \lambda'_i p_i &= \sum_{j=1}^{\ell} \mu'_j p'_j \\
\sum_{i=1}^{k} (\lambda_i - \lambda'_i) p_i &= \sum_{j=1}^{\ell} (\mu_j - \mu'_j) p_j
\end{aligned}$$

Closure Properties of Semilinear Sets

Hence $\mathbf{x}'' = \langle \lambda_{\mathbf{1}} - \lambda'_{\mathbf{1}}, \ldots, \lambda_{\mathbf{k}} - \lambda'_{\mathbf{k}}, \mu_{\mathbf{1}} - \mu'_{\mathbf{1}}, \ldots, \mu_{\ell} - \mu'_{\ell} \rangle \in \mathbf{B}$

Since $\mathbf{x}'' < \mathbf{x}$, we can apply the induction hypothesis.

Since $\mathbf{x} = \mathbf{x}' + \mathbf{x}''$, we conclude. $\square$

Closure Properties of Semilinear Sets

Proposition 2 $A = \mathcal{L}(C, P)$

"$\subseteq$" For each $\mathbf{x} = \langle \lambda_{\mathbf{1}}, \ldots, \lambda_{\mathbf{k}}, \mu_{\mathbf{1}}, \ldots, \mu_{\ell} \rangle \in \mathbf{A} \setminus \mathbf{C}$ there exists $\mathbf{x}' = \langle \lambda'_{\mathbf{1}}, \ldots, \lambda'_{\mathbf{k}}, \mu'_{\mathbf{1}}, \ldots, \mu'_{\ell} \rangle \in \mathbf{C}$ such that $\mathbf{x}' < \mathbf{x}$.

It is enough to show that $\mathbf{x} - \mathbf{x}' \in \mathbf{B}$:

$$\begin{aligned}
\sum_{i=1}^{k} (\lambda_i - \lambda'_i) p_i &= \sum_{i=1}^{k} \lambda_i p_i - \sum_{i=1}^{k} \lambda'_i p_i \\
&= (c' - c) + \sum_{j=1}^{\ell} \mu_j p'_j - \left[(c' - c) + \sum_{j=1}^{\ell} \mu'_j p'_j \right] \\
&= \sum_{j=1}^{\ell} (\mu_j - \mu'_j) p'_j
\end{aligned}$$

□

Semilinear sets = Presburger-definable sets

Theorem 2 (Ginsburg-Spanier) *The class of semilinear subsets of* $\mathbb{N}^n$ *coincides with the class of Presburger definable subsets of* $\mathbb{N}^n$*.*

"$\subseteq$" $M = \bigcup_{i=1}^{n} \mathcal{L}(\{c_i\}, \{p_{i1}, \ldots, p_{im_i}\})$

The formula defining M is the following:

$$M(x) \;\equiv\; \exists y_{11} \ldots \exists y_{nm_n} \,.\, \bigvee_{i=1}^{n} x = c_i + \sum_{j=1}^{m_i} y_{ij} p_{ij}$$

Semilinear sets = Presburger-definable sets

"$\supseteq$" Let $\phi(x_1, \ldots, x_k)$ be a Presburger formula, i.e. a disjunction of conjunctions of atomic propositions of the following forms:

$$\begin{aligned} \sum_{i=1}^{n} a_i x_i + b &\geq 0 \\ \sum_{i=1}^{n} a_i x_i + b &\equiv_n m \end{aligned}$$

Each atomic proposition describes a semilinear set, hence their intersections and unions are again semilinear sets.