Markus Müller-Olm, Helmut Seidl

Computing Polynomial Program Invariants
Computing Polynomial Program Invariants

Markus Müller-Olm
FernUniversität Hagen, LG Praktische Informatik 5
58084 Hagen, Germany
e-mail: mmo@ls5.cs.uni-dortmund.de

Helmut Seidl
TU München, Informatik, I2
85748 München, Germany
e-mail: seidl@informatik.tu-muenchen.de

October 2, 2003

Keywords: program analysis, polynomial relation, abstract interpretation, computable algebra, program correctness

1 Introduction

Invariants and intermediate assertions are the key to deductive verification of programs. Correspondingly, techniques for automatically checking and finding invariants and intermediate assertions have been studied (cf., e.g., [3, 2]). In this paper we present analyses that check and find valid polynomial relations in programs. A polynomial relation is a formula \( p(x_1, \ldots, x_k) = 0 \) where \( p(x_1, \ldots, x_k) \) is a multi-variate polynomial in the program variables \( x_1, \ldots, x_k \).\(^1\) Our analyses combine techniques from abstract interpretation and computable algebra and fully interpret assignment statements with polynomial expressions on the right hand sides while considering other assignments as non-deterministic. Polynomial non-equality guards are also treated precisely\(^2\) while other conditions at branches are ignored. The first analysis automatically checks whether a polynomial relation holds among the program variables whenever control reaches a given target program point. Our second analysis extends this testing procedure to compute precisely the set of all polynomial relations of bounded degree that are valid at the target among the program variables under the above abstraction.

The following is known as an undecidable problem in non-deterministic flow graphs, if the full standard signature of arithmetic operators (addition, subtraction, multiplication, and division) is available [8, 14]: decide whether a given variable holds a constant value at a given program point in all executions. Clearly, constancy of a variable is a polynomial relation: \( x \) is a constant at program point \( n \) if and only if the polynomial relation \( x - c = 0 \) is valid at \( n \) for some \( c \in F \). Moreover, with polynomials we can write all expressions involving addition, subtraction, and multiplication. Thus, our result allows us to find constants in non-deterministic flow graphs in which just these three operators are used (“polynomial constants”) and indicates that division is the real cause of undecidability of constant propagation.

The current paper extends and simplifies an earlier conference paper [13] that considered just detection of polynomial constants. We improve over this conference paper in a number of respects:

- A modest generalization is that we show how to check validity of arbitrary polynomial relations (Section 3), while in [13] only particular polynomial relations of the form \( x - c = 0 \) were checked for validity. While checking arbitrary polynomial relations can be done with essentially the same technique this was not made explicit in [13].
- We treat polynomial non-equality guards.

---

\(^1\) More generally our analyses can handle positive Boolean combinations of polynomial relations.

\(^2\) Again, positive Boolean combinations of such guards can be handled.
• Most importantly, we are now able not just to check polynomial relations but to derive valid polynomial relations of bounded degree (Section 4). Without a systematic way of derivation, we must guess candidate relations by some heuristic or ad-hoc method. In [13], for instance, the constant $e$ for the candidate relation $x - c = 0$ is determined in an ad-hoc way by executing a single program path.

The main idea of our checking algorithm is to compute a polynomial ideal that represents the weakest precondition for the validity of the given polynomial relation at the given target program point. We rely on results from computable algebra in order to ensure that this computation can be done effectively, most notably, on Hilbert’s Basis Theorem and on Buchberger’s Algorithm. The polynomial relation in question is valid at the target program point if and only if the computed weakest precondition is valid for all states. The latter can easily be checked.

In the case of derivation, we compute the weakest precondition of a generic polynomial relation at the target program point. In this generic relation coefficients are replaced by variables. Again an ideal that represents the weakest-precondition can be computed effectively. We can then characterize the set of values of the new variables for which the weakest precondition is universally valid by means of a linear equation system. The space of solutions of this linear equation system characterizes the coefficients of all polynomial relations (up to the chosen degree bound) which are valid at the target program point.

Looking for valid polynomial relations is a rather general question with many applications. First of all, many classical data flow analysis problems can be seen as problems about polynomial relations. Some examples are: finding definite equalities among variables like $x = y$; constant propagation, i.e., detecting variables or expressions with a constant value at run-time; discovery of symbolic constants like $x = 5y + 2$ or even $x = yz^2 + 42$; detection of complex common sub-expressions where even expressions are sought which are syntactically different but have the same value at run-time such as $xy + 42 = y^2 + 5$; and discovery of loop induction variables.

Polynomial relations found by an automatic analysis are also useful in program verification contexts, as they provide non-trivial valid assertions about the program. In particular, loop invariants can be discovered fully automatically. As polynomial relations express quite complex relationships among variables, the discovered assertions may form the backbone of the program proof and thus significantly simplify the verification task.

As an illustration of the kind of programs and properties our analysis can handle, consider the program in Figure 1. After initializing $x_1$ with 1 and $x_2$ with $x_3$, the program iteratively executes the two assignments $x_1 := x_1 \cdot x_3 + 1$; $x_2 := x_2 \cdot x_3$ in sequence. It is not difficult to see that after $n$ iterations of the loop, $x_1$ holds the value $\sum_{i=0}^{n} q^i = \frac{q^{n+1} - 1}{q - 1}$ (computed by Horner’s method) and $x_2$ holds the value $q^{n+1} + 1$ if $q$ is the (unknown) initial value of $x_3$. Therefore, after leaving the loop and multiplying $x_1$ with $x_3 - 1$ (i.e., the value $q - 1$), we can easily convince ourselves that the equation: $x_1 - x_2 + 1 = 0$ holds at program point 2.

2 Polynomial Programs

We model programs by non-deterministic flow graphs as in Figure 1. Let $X = \{x_1, \ldots, x_k\}$ be the set of (global) variables the program operates on. We use $x$ to denote the vector of variables $x = (x_1, \ldots, x_k)$. We assume that the variables take values in a fixed field $F$. For simplicity, we suppose that $F$ has characteristic zero. In practice, $F$ is the field of rational numbers. Then a state assigning values to the variables is conveniently modeled by a $k$-dimensional vector $x = (x_1, \ldots, x_k) \in F^k$; $x_i$ is the value assigned to variable $x_i$. Note that we distinguish variables and their values by using a different font.

We assume that the basic statements in the program are either polynomial assignments of the form $x_j := p$ or non-deterministic assignments of the form $x_j := ?$ where $x_j \in X$ and $p$ is a polynomial in $F[X]$; the
A program is given by a control flow graph $G = (N, E, A, s)$ that consists of:

- a set $N$ of program points;
- a set of edges $E \subseteq N \times N$;
- a mapping $A : E \rightarrow \text{Lab}$ that annotates each edge with a basic statement or negated polynomial guard;
- a special entry (or start) point $s \in N$.

The core part of our algorithm can be understood as a precise abstract interpretation of a constraint system characterizing the program executions that reach a given target program point $t \in N$. We represent program executions or runs by finite sequences

$$r = r_1; \ldots ; r_m$$

where each $r_i$ is of the form $p \neq 0$ or $x_j := p$ where $x_j \in X$ and $p \in F[X]$. We write Runs for the set of runs. The set of runs reaching $t$ from some program point $u \in N$ can be characterized as the least solution of a system of subset constraints on run sets. We start by defining the program executions of base edges $e$ in isolation. If $e$ is annotated by a guard, i.e., $A(e) \equiv p \neq 0$, or a polynomial assignment, i.e., $A(e) \equiv x_j := p$, it gives rise to a single execution: $R(e) = \{A(e)\}$. The effect of base edges $e$ annotated by a non-deterministic assignment $x_j := ?$ is captured by all runs that assign some value from $F$ to $x_j$:

$$R(e) = \{x_j := e | e \in F\}$$

Thus, we capture the effect of non-deterministic assignments by collecting all constant assignments. The runs reaching $t$ from program nodes are the smallest solution of the following constraint system $R$:

$$[R1] \quad R(t) \supseteq \{e\}$$
$$[R2] \quad R(u) \supseteq f_e(R(v)) \quad \text{if} \quad e = (u, v) \in E$$

where "$e$" denotes the empty run, and $f_e(R) = \{r; t | r \in R(e) \land t \in R\}$. By [R1], the set of runs reaching the target when starting from the target contains the empty run. By [R2], a run starting from $u$ is obtained by considering an outgoing edge $e = (u, v)$ and concatenating a run corresponding to $e$ with a run starting from $v$.

So far, we have furnished flow graphs with a symbolic operational semantics only by describing the sets of runs possibly reaching program points. Each of these runs gives rise to a partial transformation of the underlying program state $x \in F^k$; for states outside the domain the run is not executable, because some of the conditions in the run are not satisfied. Every guard $p \neq 0$ induces a partial identity function with domain

$$\text{dom}([p \neq 0]) = \{x \in F^k | p(x) \neq 0\}$$

Polynomial assignments are always executable. Thus, a polynomial assignment $x_j := p$ gives rise to the transformation with domain $\text{dom}([x_j := p]) = F^k$ and

$$[x_j := p] x = (x_1, \ldots, x_{j-1}, p(x), x_{j+1}, \ldots, x_k)$$

These definitions are inductively extended to runs:

$$[e] = \text{id}, \text{ where } \text{id} \text{ is the identity function and } [r; a] = [a] \circ [r] \text{ where } \circ \text{ denotes composition of partial functions.}$$

The partial transformation $f = [r]$ induced by a run $r$ can always be represented by polynomials.
$g_0, \ldots, g_k \in \mathbb{F}[X]$ such that $\text{dom}(f) = \{ x \in \mathbb{F}^k \mid g_0(x) \neq 0 \}$ and $f(x) = (g_1(x), \ldots, g_k(x))$ for every $x \in \text{dom}(f)$. This is clearly true for the identity transformation induced by the empty path $\varepsilon$ (take the polynomials $1, x_1, \ldots, x_k$). It is also not hard to see that the transformations induced by polynomial assignments or guards can be represented this way. Moreover, transformations of the given form are closed under composition. To see this, consider a second transformation $f'$ which is given by polynomial $g'_0, \ldots, g'_k \in \mathbb{F}[X]$. Then we have:

$$ x \in \text{dom}(f' \circ f) \iff g_0(x) \neq 0 \land g'_0(q_1(x), \ldots, q_k(x)) \neq 0 \land (g_0 \cdot g'_0(q_1/x_1, \ldots, q_k/x_k))(x) \neq 0 $$

such that $f' \circ f$ is given by the polynomials $g'_0 \cdot q_1/x_1, \ldots, g'_k/x_k$, where the $q'_i$ are obtained by substituting the polynomials $q_j$ for $x_j$ in $q'_i$, i.e.,

$$ q'_i = g'_i[x_i/x_1, \ldots, q_k/x_k] $$

3 Polynomial Relations and Weakest Preconditions

A polynomial relation over a vector space $\mathbb{F}^k$ is an equation $p = 0$ for some $p \in \mathbb{F}[X]$. Such a relation can be represented as the polynomial $p$ alone. The vector $y \in \mathbb{F}^k$ satisfies the polynomial relation $p$ if $p(y) = 0$. We also write $\vdash x \models p$ to denote this fact.

The polynomial relation (denoted by) $p$ is guaranteed to hold after a single run $r$ for those initial states $x \in \text{dom}([r])$ that satisfy $[r]x \models p$. For states $x \notin \text{dom}([r])$, $p$ is trivially guaranteed after run $r$ as $r$ is not executable for those states. Thus,

$$ x \notin \text{dom}([r]) \lor [r]x \models p $$

represents the weakest precondition of the validity of $p = 0$ after run $r$. Assuming that the transformation induced by the run $r$ is represented by the polynomials $g_0, \ldots, g_k$, we have for each $x \in \mathbb{F}^k$:

$$ x \notin \text{dom}([r]) \lor [r]x \models p \iff \begin{cases} g_0(x) = 0 \lor \forall q_1(x), \ldots, q_k(x) = 0 \\ g_0(x) = 0 \lor \forall q_1/x_1, \ldots, q_k/x_k = 0 \\ (g_0 \cdot p[q_1/x_1, \ldots, q_k/x_k])(x) = 0 \end{cases} $$

From this calculation, we deduce that the weakest precondition is again a polynomial relation. Even better: the mapping $[r]^T$ that assigns to each polynomial relation (represented by a single polynomial) its weakest precondition before run $r$ is the total function defined by:

$$ [r]^T p = g_0 \cdot p[q_1/x_1, \ldots, q_k/x_k] \tag{1} $$

The only polynomial relation which is true for all program states is the relation $0 = 0$. Thus, the polynomial relation $p$ is valid after run $r$ if $[r]^T p = 0$, because the initial state is arbitrary. Accordingly, the polynomial relation $p$ is valid at the target node $t$ if it is valid after all runs $r \in \mathbb{R}(s)$. Summarizing, we have:

**Lemma 1** The polynomial relation $p_t \in \mathbb{F}[X]$ is valid at the target node $t$ iff $[r]^T p_t = 0$ for all $r \in \mathbb{R}(s)$.

We conclude that the set $S = \{ [r]^T p_t \mid r \in \mathbb{R}(s) \} \subseteq \mathbb{F}[X]$ of polynomials gives us a handle to solve the validity problem for the polynomial relation $p_t$ at the target node $t$: $p_t$ is valid at $t$ iff $S \subseteq \{0\}$. The problem is that we need a representation of this set which is finitary — and find a way to compute it. In this case, we recall that the set $\mathbb{F}[X]$ of all polynomials forms a commutative ring. A subset $I$ of a commutative ring $R$ satisfying the two conditions:

(i) $a + b \in I$ whenever $a, b \in I$ (closure under sum) and

(ii) $r \cdot a \in I$ whenever $r \in R$ and $a \in I$ (closure under product with arbitrary ring elements)

is called *ideal*. Ideals (in particular those in polynomial rings) enjoy interesting and useful properties. For a subset $G \subseteq R$, the least ideal containing $G$ is given by

$$ \langle G \rangle = \{ r_1 g_1 + \ldots + r_n g_n \mid n \geq 0, r_i \in R, g_i \in G \} $$

In this case, $G$ is also called the set of *generators* of $\langle G \rangle$. In particular,

$$ \langle G \rangle = \{0\} \iff G \subseteq \{0\} $$

for every $G \subseteq R$. Thus, in our scenario, we can equivalently check $\langle S \rangle = \{0\}$ instead of $S \subseteq \{0\}$. We conclude that we can work with ideals of polynomials instead of sets without losing interesting information. The set $\mathcal{I}_X$ of ideals of $\mathbb{F}[X]$, ordered by subset inclusion, forms a complete lattice. In particular,

- The least element of $\mathcal{I}_X$ is $\{0\}$. 

4
The greatest element of $\mathcal{I}_X$ equals $\langle 1 \rangle = F[X]$.

The least upper bound $I_1 \cup I_2$ of two ideals is defined by $I_1 + I_2 = \langle I_1 \cup I_2 \rangle$.

Moreover, we recall Hilbert’s famous basis theorem for polynomial ideals over a field:

**Theorem 1 (Hilbert, 1888)** Every ideal $I \subseteq F[X]$ of a commutative polynomial ring in finitely many variables is finitely generated, i.e., $I = \langle G \rangle$ for some finite subset $G \subseteq F[X]$.

This means that each ideal can be effectively represented. For testing validity of the polynomial relation $I$, we need an abstract transformer $f^I_e : \mathcal{I}_X \to \mathcal{I}_X$ corresponding to edges $e = (u, v)$ which exactly abstracts $f_e$, i.e., the effect of concatenating the fixed run set of the edge $e$ with run sets. We define:

$$f^I_e I = \langle \{ [r]^T p \mid r \in R(e), p \in I \} \rangle$$

We prove:

**Lemma 3** For every subset $G$ of polynomials,

$$f^I_e(G) = \langle \{ [r]^T p \mid r \in R(e), p \in G \} \rangle$$

**Proof:** Since $G \subseteq \langle G \rangle$, we trivially have the inclusion “$\subseteq$” by monotonicity. For the reverse inclusion, consider a polynomial $p \in f^I_e(G)$. Then $p$ can be written as $p = \sum q_i p_i$ for polynomials $p_i = [r_i]^T p_i'$ for some $r_i \in R(e)$ and $p_i' \in G$, and $q_i \in F[X]$. Each $p_i'$ can turn can be written as $p_i' = \sum q'_{ij} g_{ij}$ for $g_{ij} \in G$ and arbitrary polynomials $q'_{ij}$. In particular,

$$p_i = \sum q'_{ij} g_{ij}$$

for some polynomials $q_{ij}$ (unfold the definition of $q'_{ij}$ for seeing the last step). Therefore, $p_i \in \langle \{ [r]^T p \mid r \in R(e), p \in G \} \rangle$ for all $i$. But then also $p \in \langle \{ [r]^T p \mid r \in R(e), p \in G \} \rangle$ since ideals are closed under sums and products with arbitrary polynomials.

Using Lemma 3, we calculate:

$$f^I_e(\alpha(R)) = f^I_e(\langle \{ [r]^T p \mid r \in R \} \rangle) = \langle \{ [r]^T (\{ [r]^T p \mid r' \in R(e), r \in R(e) \}) \mid r' \in R(e), r \in R \} \rangle = \langle \{ [r]^T p \mid r \in f_e(R) \} \rangle = \alpha(f_e(R))$$

Therefore, $f^I_e$ is indeed an exact abstraction of $f_e$. It remains to prove that the application of $f^I_e$ can be effectively computed. This is easy if $A(e)$ is either a guard or a polynomial assignment. Then the set $R(e)$ consists of a single element, namely, a guard or a polynomial.
assignment. For any generating system $G \subseteq \mathbb{F}[X]$, we therefore obtain by Lemma 3,

$$f^2_\mathcal{E}(G) = \left\{ \langle p \cdot q \mid q \in G \rangle \mid A(e) = p \neq 0 \right\}$$

In particular, we conclude that for every finite set of generator polynomials $G$, a finite generating system for the ideal $f^2_\mathcal{E}(G)$ is effectively computable.

Not quite as obvious is the case where the edge $e$ is labeled with an unknown assignment $x_j : = ?$. Then the run set $R(e) = \{ x_j := c \mid c \in \mathbb{F} \}$ is infinite. Still, however, the effect of concatenating this run set turns out to be computable. To see this, recall that every polynomial $p \in \mathbb{F}[X]$ can be uniquely written as a sum

$$p = \sum_{i=0}^{d} p_i \cdot x_j^i$$

where the $\{ x_j \}$-coefficient polynomials $p_i$ of $x_j^i$ do not contain occurrences of $x_j$. Then define $\pi_j : \mathbb{F}[X] \to 2^{\mathbb{F}[X]}$ as the mapping which maps $p$ to the set $\{ p_0, \ldots, p_d \}$ of its $\{ x_j \}$-coefficient polynomials. We prove:

**Lemma 4** Assume $A(e) \equiv x_j := ?$. Then for every set $G$ of generator polynomials,

$$f^2_\mathcal{E}(G) = \bigcup \left\{ \langle \pi_j(q) \mid q \in G \rangle \right\}$$

The lemma and its proof are similar to Lemma 8 in [12].

**Proof:** By definition and Lemma 3, we have:

$$f^2_\mathcal{E}(G) = \langle \{ x_j := c \} \mid c \in \mathbb{F}, q \in G \rangle$$

$$= \langle \{ q[c/x_j] \mid c \in \mathbb{F}, q \in G \rangle$$

Obviously, each polynomial $\{ q[c/x_j] \}$ is contained in the ideal generated from the $\{ x_j \}$-coefficient polynomials of $q$. Therefore, a generator set of the left-hand side $f^2_\mathcal{E}(G)$ is included in the right-hand side $\langle \bigcup \{ \pi_j(q) \mid q \in G \} \rangle$ of the equation, and hence also the generated ideal. This proves the inclusion “$\subseteq$”.

For the reverse inclusion, it suffices to prove for an arbitrary polynomial $q \in G$, that the set $\pi_j(q)$ of $\{ x_j \}$-coefficient polynomials of $q$ is contained in $\langle q[c/x_j] \mid c \in \mathbb{F} \rangle$. Assume that $q = \sum_{i=0}^{d} q_i \cdot x_j^i$ where $q_i$ do not contain occurrences of $x_j$. Consider the square matrix $A$ defined by:

$$A = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
1 & 1 & \cdots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & d & \cdots & d^d
\end{pmatrix}$$

It is not hard to see that

$$A \begin{pmatrix} q_0 \\ \vdots \\ q_d \end{pmatrix} = \begin{pmatrix} q[0/x_j] \\ \vdots \\ q[d/x_j] \end{pmatrix}.$$ 

The determinant of $A$ is an instance of what is known as Vandermonde’s determinant and has the value $\prod_{0 \leq i < m \leq (m - 1)}$. As the characteristic of $\mathbb{F}$ is 0, the determinant is different from 0. Therefore, matrix $A$ is invertible and for the inverse matrix, we have

$$\begin{pmatrix} q_0 \\ \vdots \\ q_d \end{pmatrix} = A^{-1} \begin{pmatrix} q[0/x_j] \\ \vdots \\ q[d/x_j] \end{pmatrix}.$$ 

Thus, the coefficient polynomials of $q$ are even linear combinations of the polynomials $q[0/x_j], \ldots, q[d/x_j]$ which shows that $\pi_j(q) = \{ q_0, \ldots, q_d \}$ is contained in the ideal generated by the polynomials $q[c/x_j], c \in \mathbb{F}$.

Since for every polynomial $p$, the set of its $\{ x_j \}$-coefficient polynomials is effectively computable, we conclude that also the generator set $\bigcup \{ \pi_j(q) \mid q \in G \}$ of the ideal $f^2_\mathcal{E}(G)$ is effectively computable—given only that the set $G$ is finite.

For a given target node node $t \in N$ and polynomial relation $p \in \mathbb{F}[X]$ let $R^2_t$ denote the following abstracted constraint system over the complete lattice $\mathcal{L}_X$:

$$[R1]^t \quad R^2(t) \supseteq \langle \{ p_t \} \rangle$$

$$[R2]^t \quad R^2(u) \supseteq f^2_{\mathcal{E}}(R^2(v)) \quad \text{if } e = (u, v) \in E$$

We find:

**Lemma 5** The constraint system $R^2_u$ has a unique least solution $R^2(u), u \in N$, with the following properties:

1. $R^2(u), u \in N$, is effectively computable.
2. $R^2(u) = \alpha(R(u))$ for every $u \in N$.

**Proof:** Since $\mathcal{L}_X$ is a complete lattice and all transfer functions $f^2_{\mathcal{E}}$ on right-hand sides of constraints are monotonic, the constraint system $R^2_u$ has a unique least solution. Moreover, recall that Hilbert’s basis Theorem 1 implies that every ascending sequence of ideals:

$$I_0 \subseteq \cdots \subseteq I_m \subseteq \cdots$$
is ultimately stable, i.e., \( I_{m'} = I_m \) for some \( m \in \mathbb{N} \) and all \( m' \geq m \). We conclude that the least solution can be computed by a finite number of fixpoint iterations. Since each intermediate occurring ideal is finitely generated, each individual fixpoint iteration is computable. By Buchberger’s algorithm (cf., e.g., [5, 6, 15]) it is decidable whether or not a polynomial \( p \) is contained in the ideal \( \langle G \rangle \) for a finite set of generators \( G \). It follows that it is also decidable whether an ideal \( I_1 \) is included in another ideal \( I_2 \) — given only finite generator sets \( G_i \) for the involved ideals \( I_i \). Thus, we can effectively decide when fixpoint iteration for \( R_{\alpha}^f \) reaches the least fixpoint. This completes the proof of Assertion 1.

For the second assertion, we apply the Transfer Lemma of general fixpoint theory (see, e.g., [1, 4]). By Lemma 2 the abstraction function \( \alpha \) is completely distributive, and by Lemma 3, the transfer functions are exact. Under these assumptions, the Transfer Lemma ensures that the least solution of the abstracted constraint system \( R_{\alpha}^f \) is the abstraction of the least solution of the concrete constraint system \( R \). This is Assertion 2. Thus, the proof is complete.

We can now put the pieces together and prove the main theorem of this section.

**Theorem 2** There is an effective procedure to decide whether a polynomial relation \( p_t \) is valid at a given program point \( t \) or not.

**Proof:** The polynomial relation \( p_t \) is valid at \( t \) if and only if \( \alpha(\mathbb{R}(s)) = \{0\} \), where \( s \) is the entry point of the program. By Lemma 5 we can effectively compute a generating system for \( \alpha(\mathbb{R}(s)) \) by computing the least solution of constraint system \( R_{\alpha}^f \). As the ideal \( \{0\} \) has only the two sets of generators \( \emptyset \) and \( \{0\} \) it is easy to check, whether the set of generators computed for \( R_{\alpha}^f(s) \) generates \( \{0\} \) or not.

Consider the example program from Figure 1. We want to verify that the relation given by the polynomial \( p = x_1 - x_2 + 1 \) holds at program point 2. Starting from the ideal \( \langle \{p\} \rangle \) for program point 2, we obtain a set of generators for the ideal \( R_{\alpha}^f(1) \) of preconditions at program point 1 by first computing:

\[
q_1 = \left[ x_1 := x_1 \cdot (x_3 - 1) \right]^T p = p[x_1 \cdot (x_3 - 1)/x_1] = x_1 \cdot x_3 - x_1 - x_2 + 1
\]

and then iteratively adding to the ideal \( \langle q_1 \rangle \) further preconditions for the loop until stabilization is reached. We have:

\[
\begin{align*}
[x_1 := x_1 \cdot x_3 + 1; x_2 := x_2 \cdot x_3] & \in \langle q_1 \rangle \\
& = q_1 \left[ x_1 \cdot x_3 + 1/x_1, x_2 \cdot x_3/x_2 \right] \\
& = (x_1 \cdot x_3 + 1) \cdot x_3 - (x_1 \cdot x_3 + 1) - x_2 \cdot x_3 + 1 \\
& = x_1 \cdot x_3^2 - x_1 \cdot x_3 - x_2 \cdot x_3 + x_3 \\
& = x_3 \cdot q_1 \\
& \in \langle q_1 \rangle
\end{align*}
\]

Thus, the value of the fixpoint for program point 1 is given by \( R_{\alpha}^f(1) = \langle \{q_1\} \rangle \). For the entry point 0 of the program we then calculate the set of preconditions for the set \( \{q_1\} \):

\[
[x_1 := 1; x_2 := x_3]^T q_1 = 1 \cdot x_3 - 1 - x_3 + 1 = 0
\]

Therefore, \( R_{\alpha}^f(0) = \langle \{0\} \rangle = \{0\} \) — implying that the relation \( x_1 - x_2 + 1 = 0 \) indeed holds at program point 2. \( \square \)

The considerations of this section can easily be extended to checking finite sets of polynomials. A set \( G \subseteq \mathbb{F}[x] \) is valid for a state \( y \in \mathbb{F}^k \), \( y \models G \) for short, iff \( y \models p \) for all \( p \in G \). Thus, a set represents the conjunction of its members. We can clearly check validity of a finite set \( G_t \) at a program point \( t \) by applying the above procedure for each relation in \( G_t \). We can do better, however, by checking all of them at once. Clearly, we obtain from Lemma 1:

**Corollary 1** The set of polynomial relations \( G_t \subseteq \mathbb{F}[x] \) is valid at the target node \( t \) iff \( \forall p \in G \) for all \( r \in \mathbb{R}(s), p \in G_t \).

Accordingly, we work with the abstraction mapping \( \alpha' : 2^{\text{Runs}} \rightarrow \mathcal{I}_x \):

\[
\alpha'(R) = \langle \{[r]^T p \mid r \in R, p \in G_t \} \rangle.
\]

This leads to the a slightly modified constraint \([R1]^\alpha\):

\[
[R1]^\alpha \quad R_{\alpha}^f(t) \supseteq \langle G_t \rangle
\]

The rest works as before. We conclude:

**Theorem 3** There is an effective procedure to decide whether a (finite) set of polynomial relations \( G_t \) is valid at a given program point \( t \) or not.
Note that we can represent *disjunctions* of polynomial relations by products: \( p = 0 \lor p' = 0 \) is valid for a state \( \sigma \) iff \( \sigma \models p \land p' \). Thus, by considering sets of polynomials and using products, we can indeed handle arbitrary *positive Boolean combinations* of polynomial relations.

4 Inferring Valid Polynomial Relations

It seems that the algorithm of testing whether a certain given polynomial relation \( p_0 = 0 \) is valid at some program point contains no clue on how to infer so far unknown valid polynomial relations. This, however, is not quite true. Let the degree of a polynomial be the maximal degree of a monomial occurring in \( p \) where the degree of a monomial \( b \cdot x_1^{i_1} \cdots x_k^{i_k}, b \in \mathbb{F} \), equals \( i_1 + \cdots + i_k \), i.e., the sum of the exponents of the variables. For a fixed maximal degree \( d \in \mathbb{N} \), let \( D_d = \{(i_1, \ldots, i_k) \mid i_1 + \cdots + i_k \leq d \} \). We introduce a new set of variables \( A_d \) given by:

\[
A_d = \{ a_\sigma \mid \sigma \in D_d \}
\]

Then we introduce the *generic* polynomial of degree \( d \) as the polynomial

\[
p_d = \sum_{\sigma = (i_1, \ldots, i_k) \in D_d} a_\sigma \cdot x_1^{i_1} \cdots x_k^{i_k}
\]

The polynomial \( p_d \) is an element of the polynomial ring \( \mathbb{F}[X \cup A_d] \). Note that every concrete polynomial \( p \in \mathbb{F}[X] \) of degree at most \( d \) can be obtained from the generic polynomial \( p_d \) simply by substituting concrete values \( a_\sigma, \sigma \in D_d \), for the variables \( a_\sigma \). If \( a : \sigma \mapsto a_\sigma \) and \( a : \sigma \mapsto a_\sigma \), we write \( p_d(a|a) \) for this substitution. We have:

**Lemma 6** Let \( p = \sum_{\sigma = (i_1, \ldots, i_k) \in D_d} a_\sigma \cdot x_1^{i_1} \cdots x_k^{i_k} \in \mathbb{F}[X] \) denote a polynomial of degree at most \( d \) with coefficients \( a : \sigma \mapsto a_\sigma \). Then for every run \( r \),

\[
[r]^T p = (\langle [r]^T p_d \rangle)[a|a]
\]

where \([r]^T\) on the left-hand side of the equation is computed over \( \mathbb{F}[X] \) whereas on the right-hand side it is computed over \( \mathbb{F}[X \cup A_d] \).

**Proof:** By Equation (1), there are polynomials \( q_0, \ldots, q_k \in \mathbb{F}[X] \) such that \( [r]^T p' = q_0 \cdot p'[q_0/x_1, \ldots, q_k/x_k] \) for every polynomial \( p' \). Therefore,

\[
\begin{align*}
[r]^T p &= \langle [r]^T p_d \rangle[a|a] \\
&= q_0 \cdot \langle [r]^T p_d[q_0/x_1, \ldots, q_k/x_k] \rangle[a|a] \\
&= \langle (q_0 \cdot p_d[q_0/x_1, \ldots, q_k/x_k])[a|a] \rangle[a|a] \\
&= \langle ([r]^T p_d)[a|a] \rangle[a|a]
\end{align*}
\]

which proves the asserted equality. \( \square \)

Lemma 6 tells us that instead of computing the weakest precondition of each polynomial of degree at most \( d \) separately, we as well may compute the weakest precondition of the single generic polynomial \( p_d \) once and for all and substitute the concrete coefficients \( a_\sigma \) of the polynomials \( p \) into the precondition of \( p_d \) later. In particular, we conclude that the following statements are equivalent:

1. \( p \) is valid at the target program point \( t \);
2. \( \langle [r]^T p \rangle = 0 \) for all \( r \in \mathbb{R}(s) \);
3. \( \langle [r]^T p_d \rangle[a|a] = 0 \) for all \( r \in \mathbb{R}(s) \);
4. \( q[a|a] = 0 \) for all \( q \in \{[r]^T p_d \mid r \in \mathbb{R}(s)\} \);
5. \( q[a|a] = 0 \) for all \( q \in \{\langle [r]^T p_d \rangle[r \in \mathbb{R}(s)]\} \).
6. \( q[a|a] = 0 \) for all \( q \in G \) in a (finite) generator \( G \) of the ideal \( \langle \{[r]^T p_d \mid r \in \mathbb{R}(s)\} \rangle \).

Now it should be clear how an algorithm may find all polynomial relations of degree at most \( d \) which are valid at program point \( t \): first, we construct the abstract constraint system \( \mathbb{R}^d_{\mathbb{X} \cup A_d} \), now over \( \mathbb{R}(X \cup A_d) \), which for each program point \( u \) computes a (finite) generator set of the ideal \( \mathbb{R}^d(u) = \langle \{[r]^T p_d \mid r \in \mathbb{R}(u)\} \rangle \). Then it remains to determine the set of all coefficient maps \( a : D_d \mapsto \mathbb{F} \) such that \( q[a|a] = 0 \) for all \( q \in \mathbb{R}^d(s) \). Recall that each such polynomial \( q[a|a] \) is a polynomial in \( \mathbb{F}[X] \). Any such polynomial equals \( 0 \) iff each coefficient \( b \) of each occurring monomial \( b \cdot x_1^{i_1} \cdots x_k^{i_k} \) equals \( 0 \). The polynomial \( q \in \mathbb{F}[X \cup A_d] \), on the other hand, can uniquely be written as a finite sum

\[
q = \sum_{\sigma = (i_1, \ldots, i_k)} q_\sigma \cdot x_1^{i_1} \cdots x_k^{i_k}
\]

where each \( X \)-coefficient \( q_\sigma \) is in \( \mathbb{F}[A_d] \), i.e., may only contain occurrences of variables from \( A_d \). Thus, \( q[a|a] = 0 \) iff \( q_\sigma[a|a] = 0 \) for all index tuples \( \sigma \) occurring in the sum. Summarizing our considerations so far, we have shown:
Lemma 7 Let $G$ denote any finite generator set for the ideal $R^d(s)$. The set of coefficient maps $a : D_d \rightarrow F$ of polynomials of degree at most $d$ which are valid at program point $t$ equals the set of solutions of the equation system having an equation:

$$q_{\sigma} = 0$$

for each $X$-coefficient $q_{\sigma}$ of a polynomial $q \in G$. □

We are not yet done, since in general we are not able to determine the precise set of solutions of an arbitrary polynomial equation system algorithmically. Therefore, we need the following extra observation:

Lemma 8 Every ideal $R^d(u)$, $u \in N$, of the least solution of the abstract constraint system $R^d_{\Sigma}$ has a finite generator set $G$ consisting of polynomials $q$ whose $X$-coefficients are of degree at most 1, i.e., are of the form:

$$\sum_{\sigma \in D_u} b_{\sigma} \cdot a_{\sigma}$$

for $b_{\sigma} \in F$. Moreover, such a generator set can be effectively computed.

Proof: The polynomial $p_d$ has $X$-coefficients which trivially have degree 1, since these consist of individual variables $a_{\sigma}$. Also, applications of the least upper bound operation as well as of the abstract transformers $f_2^+$ when applied to (ideals represented through) finite sets of generators with $X$-coefficients of degree at most 1 again result in finite sets of generators with this property. Therefore, the assertion of the lemma follows by fixpoint induction. □

Together Lemma 7 and Lemma 8 show that the set of (coefficient maps) of polynomials of degree at most $d$ which are valid at our target program point $t$ can be characterized as the set of solutions of a linear equation system. Such equation systems can be algorithmically solved, i.e., finite representations of their sets of solutions can be constructed explicitly. We conclude our second main theorem:

Theorem 4 The set of all polynomials of degree at most $d$ which are valid at some target program point $t$ can be effectively computed.

Consider again the example program from Figure 1. We want to determine for program point 2, all valid polynomial relations up to degree 1, i.e., all valid polynomial relations of the form $a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 = 0$. Let $p_t \equiv a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3$ denote the generic polynomial of degree 1. Starting from the ideal $\langle \{p_t\} \rangle$ for program point 2, we determine a set of generators for the ideal $R^2(1)$ of preconditions at program point 1. First, we compute:

$$q_1 = \left[ x_1 := x_1 \cdot (x_3 - 1) \right] p_1$$

$$= p_1 \cdot x_1 \cdot (x_3 - 1)/x_1$$

$$= a_0 - a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + a_1 \cdot x_1 \cdot x_3$$

Next, we add the preconditions for the body of the loop:

$$\left[ x_1 := x_1 \cdot x_3 + 1; x_2 := x_2 \cdot x_3 \right] q_1$$

$$= q_1 \cdot x_1 \cdot x_3 + 1/x_1 \cdot x_2 \cdot x_3/x_2$$

$$= a_0 - a_1 \cdot x_1 \cdot a_1 \cdot x_1 \cdot x_3 + x_3 - x_3 \cdot x_1$$

$$= x_3 \cdot q_1$$

$$= a_0 - a_1 \cdot x_1 \cdot (-a_0 + a_1 + a_3) \cdot x_3 - a_3 \cdot x_3^2$$

The polynomial $q_2 \equiv a_0 - a_1 + (-a_0 + a_1 + a_3) \cdot x_3 - a_3 \cdot x_3^2$ is independent of $x_1$ and $x_2$. Thus, the ideal $\langle \{q_1, q_2\} \rangle$ remains stable under further iteration and therefore equals $R^2(1)$. A generator set for $R^2(0)$ is obtained by computing:

$$\left[ x_1 := x_1; x_2 := x_2 \right] q_1$$

$$= a_0 - a_1 + (a_1 + a_2 + a_3) \cdot x_3$$

$$\left[ x_1 := x_1; x_2 := x_2 \right] q_2$$

$$= q_2$$

$$= a_0 - a_1$$

$$= (-a_0 + a_1 + a_3) \cdot x_3 - a_3 \cdot x_3^2$$

The $\{x_1, x_2, x_3\}$-coefficients of these two polynomials now must equal 0. This gives us the following linear equations:

$$a_0 - a_1 = 0$$

$$a_0 - a_1 = 0$$

$$a_1 + a_2 + a_3 = 0$$

$$-a_0 + a_1 + a_3 = 0$$

$$-a_3 = 0$$

Thus, $a_3 = 0$, $a_1 = a_0$, and $a_2 = -a_0$. We conclude that $1 + x_1 - x_2 = 0$ is (up to constant multiples) the only polynomial relation of degree at most 1 which is valid at program point 2. □
5 Conclusion

We have presented two analysis algorithms. The first analysis determines for a given program point of a polynomial program with negative guards whether a given polynomial relation is valid or not. The second analysis generalizes this algorithm to infer all polynomial relations of bounded degree.

Linear algebra techniques have been used in program analysis for a long time. In his seminal paper [9], Karr presents an analysis that determines valid affine relations by a forward propagation of affine spaces. His analysis is precise for affine programs, i.e., it interprets assignments with affine right-hand sides precisely. In [11] we observe that checking a given affine relation for validity at a program point can be performed by a simpler backward propagating algorithm. This idea of backward propagation has lead to an interprocedural generalization of Karr’s result [12] and also underlies the current paper. In comparison with Karr’s result, we have a more general space of properties, namely polynomial relations instead of affine relations. Secondly, our analysis is precise for a larger class of programs, namely polynomial programs instead of affine programs. Finally, we leave the realm of linear algebra and rely on results from computable algebra instead.

We are not aware of much work on using techniques from computable algebra in program analysis, like we do here. In the work of Michel le Borgne et. al. (cf., e.g., [10]) and Gunnarsson et. al. (cf., e.g., [7]) polynomials over a finite field are used for representing state spaces in a forward reachability analysis of polynomial dynamical systems or discrete event dynamical systems, respectively. However, they actually work in a finite factorization of a polynomial ring over a finite field and use polynomials for representing state spaces of finite systems and not for treating arithmetic properties. Thus, they use polynomials as a convenient data structure but not to gain new decidability insights.

It is a challenging open problem whether or not the degree bound can be lifted, i.e., whether or not the set of all valid polynomial relations can be computed. It is not hard to see that this set is an ideal of $\mathbb{F}[X]$. Hence, by Hilbert’s basis theorem it can be represented by a finite set of generators such that this is a well-posed problem. Another challenge is to treat the inter-procedural case, i.e., to detect or even infer polynomial relations in programs with polynomial assignments and procedures.

References