On Variable-Bounded Non-Linear Expansions of Presburger Arithmetic

Piotr Bacik
Oxford University and Max Planck Institute for Software Systems
UK, Germany

Joris Nieuwveld
Oxford University
UK

Joël Ouaknine
Max Planck Institute for Software Systems
Germany

Mihir Vahanwala
Max Planck Institute for Software Systems
Germany

Madhavan Venkatesh
Max Planck Institute for Software Systems
Germany

Emil Rugaard Wieser
Max Planck Institute for Software Systems
Germany

ABSTRACT

We consider expansions of Presburger arithmetic with families of monadic polynomial predicates. (Examples of such predicates are the set of perfect squares, or the set of integers of the form $2n^3 - 5n + 3$, etc.) Although the full attendant first-order theories are well known to be undecidable, very little is known when one restricts the number of variables. For single-variable theories, we obtain positive results for the following two families of predicates: (i) for perfect fixed powers, decidability of the corresponding theory follows from the solvability of hyperelliptic Diophantine equations; and (ii) for polynomials of degree at most three, we establish decidability by relying on the low genus of the resulting algebraic curves. Finally, we discuss limitations and hardness results (via encodings of longstanding open Diophantine problems) as soon as any of the above restrictions are lifted.

CCS CONCEPTS

• **Theory of computation → Logic and verification**.

KEYWORDS

Presburger arithmetic, Diophantine equations, decidability, Büchi's conjecture

ACM Reference Format:
Piotr Bacik, Joris Nieuwveld, Joël Ouaknine, Mihir Vahanwala, Madhavan Venkatesh, and Emil Rugaard Wieser. 2026. On Variable-Bounded Non-Linear Expansions of Presburger Arithmetic. In . ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Presburger arithmetic was introduced and proven decidable in 1929 as a preliminary step towards Hilbert's goal of mechanising all of number theory, and in particular algorithmically determining the satisfiability of arbitrary Diophantine equations. Unfortunately, the famous works of Gödel, Church, and Turing in the 1930s brought the Hilbert program to a screeching halt, and Matiyasevich dealt the final blow in 1970 by proving, building on a large body of work by himself and others, that solving polynomial equations over the integers was in general algorithmically infeasible; in other words that Hilbert's tenth problem was undecidable.

Somewhat paradoxically, the demise of Hilbert's program did not dampen the scientific community's appetite for investigating the decidability of various logical theories of arithmetic and beyond: research into non-linear expansions and fragments of Presburger arithmetic, for example, remains a topic of active interest; see, for instance, the surveys [4, 10, 17]. A recent breakthrough by Hieronymi and Schulz shows that expanding Presburger arithmetic by two or more power predicates over multiplicatively independent bases leads to undecidability; in other words, for example, the first-order theory $FO\langle\mathbb{Z}; 0, 1, +, <, 2^{\mathbb{N}}, 3^{\mathbb{N}}\rangle$ is undecidable, where $2^{\mathbb{N}}$ and $3^{\mathbb{N}}$ stand for the sets of powers or 2 and powers of 3, respectively. Decidability can however be recovered when restricting to the *existential* fragment, viz. $\exists FO\langle\mathbb{Z}; 0, 1, +, <, 2^{\mathbb{N}}, 3^{\mathbb{N}}\rangle$ [14]. (The problem remains wide open when three or more power predicates are simultaneously in play.)

Let us turn to monadic *polynomial* predicates, i.e., sets of the form $\mathcal{R} = \{f(x) \mid x \in \mathbb{Z}\}$, where $f \in \mathbb{Q}[x]$ is an integer-valued polynomial with rational coefficients.[1] It is folklore that, whenever $\mathcal{R}$ corresponds to a polynomial of degree at least 2, the theory $FO\langle\mathbb{Z}; 0, 1, +, <, \mathcal{R}\rangle$ is automatically undecidable, via a simple encoding of multiplication within. In the 1970s, Büchi considered specifically the case of perfect squares, i.e., the predicate $\mathbf{Z}^2 := \{n^2 \mid n \in \mathbb{Z}\}$, and asked about the decidability of the *existential* fragment $\exists FO\langle\mathbb{Z}; 0, 1, +, <, \mathbf{Z}^2\rangle$. As we describe in greater detail in Sec. 5, Büchi in fact formulated a conjecture implying undecidability of this theory; a proof of Büchi's conjecture was recently announced by Xiao [24], finally establishing undecidability of the corresponding logical theory after some five decades! Note that Xiao's proof only concerns the perfect-square predicate, and the general question of the decidability of $\exists FO\langle\mathbb{Z}; 0, 1, +, <, \mathcal{R}\rangle$, where $\mathcal{R}$ is an arbitrary non-linear polynomial predicate, remains open.

In addition to restricting the number and use of quantifiers, another classical means of attempting to recover decidability involves

[1] Note that, whilst all polynomials with integer coefficients are automatically integer valued over $\mathbb{Z}$, the converse does not hold; consider, for example, the polynomial $f(x) = \frac{x^2+x}{2}$.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

bounding the number of variables; standard references on bounded-variable logics include [7–9, 15, 19].

We are now in a position to describe our main contributions. We focus on variable- and quantifier-bounded expansions of Presburger arithmetic with families of monadic polynomial predicates. Since variables and quantifiers are now in short supply, we expand our base signature to maximise expressiveness and flexibility,[2] by considering the following, where k is the bound on the number of allowable distinct variables and the binary relation symbol $\equiv_m$ refers to congruence modulo m:

- $FO^k\langle\mathbb{Z}; 0, 1, +, -, <, (\equiv_m)_{m\geq 2}\rangle$ denotes the first-order fragment with no restrictions on quantifiers;
- $\exists FO^k\langle\mathbb{Z}; 0, 1, +, -, <, (\equiv_m)_{m\geq 2}\rangle$ denotes the existential fragment;
- $SMT^k\langle\mathbb{Z}; 0, 1, +, -, <, (\equiv_m)_{m\geq 2}\rangle$ denotes the *satisfiability modulo theories* fragment, i.e., existential formulas in prenex normal form: $\exists x_1, \ldots, x_k \,.\, \varphi(x_1, \ldots, x_k)$, where $\varphi(x_1, \ldots, x_k)$ is quantifier free.

Somewhat surprisingly, even restricting to a *single* variable (i.e., $k = 1$) immediately leads to well-known open problems: let $\mathcal{R}_1$ and $\mathcal{R}_2$ be predicates corresponding to polynomials f_1 and f_2; in general the decidability of whether there are integers u and v such that $f(u) = g(v)$ — a severely restricted instance of Hilbert's tenth problem — is open. But such a query is easily encodable within the bare theory $SMT^1\langle\mathbb{Z}; \mathcal{R}_1, \mathcal{R}_2\rangle$, by asking for the truth value of $\exists x \,.\, \mathcal{R}_1(x) \wedge \mathcal{R}_2(x)$. Even in the case of a *single* predicate $\mathcal{R}$ (with underlying polynomial f), decidability remains open. Consider, for arbitrary integer constants a, b, c, d, the sentence $\exists x \,.\, \mathcal{R}(ax + b) \wedge \mathcal{R}(cx + d)$, which is readily expressible in $SMT^1\langle\mathbb{Z}; 0, 1, +, -, \mathcal{R}\rangle$. This formula asserts the existence of integers u and v such that $cf(u) + ad = af(v) + cb$; however in general it is not known whether such Diophantine equations can always be solved.

Our main positive results exclusively concern single-variable theories (in which case the first-order, existential, and SMT fragments essentially all coincide). We establish decidability of the following:

- Single-variable expansions of Presburger arithmetic by arbitrarily many polynomial predicates corresponding to perfect fixed powers (Thm. 3.1):
$$FO^1\langle\mathbb{Z}; 0, 1, +, -, <, (\equiv_m)_{m\geq 2}, (\mathbf{Z}^k)_{k\geq 2}\rangle \,.$$
- Single-variable expansions of Presburger arithmetic by arbitrarily many polynomial predicates of degree at most 3 (Thm. 4.1):
$$FO^1\langle\mathbb{Z}; 0, 1, +, -, <, (\equiv_m)_{m\geq 2}, (\mathcal{R}_i)_i\rangle \,.$$

Amongst other ingredients, these theorems are obtained by making use of deep results on the solvability of hyperelliptic Diophantine equations and equations corresponding to algebraic curves of low genus.

Finally, we establish undecidability of expansions involving the perfect-square predicate $\mathbf{Z}^2$ when several variables are allowed (Thm. 5.2), and discuss various other limitations and hardness results in Sec. 5. It is worth noting that the famous perfect-Euler-brick problem, which asks whether there exists a rectangular box with integer sides, and all of whose diagonals are moreover also integers, is easily encodable within $SMT^3\langle\mathbb{Z}; 0, +, \mathbf{Z}^2\rangle$; as this problem has been open for over two centuries, decidability of this three-variable fragment should therefore be considered well out of reach. On the other hand, the two-variable fragments of the various theories considered in this paper remain open and fascinating avenues for further research.

[2]For example, the subtraction operator is typically not included in the signature of Presburger-arithmetic theories, since a term such as $-x$ can be recovered through existential quantification: $\exists y \,.\, y + x = 0$. Likewise, modular-arithmetic constraints are usually implicit: x is an even number if and only if $\exists y \,.\, x = y + y$, etc.

2 TECHNICAL PRELIMINARIES

Integer-Valued Polynomials

We refer to [5] for basic properties of univariate integer-valued polynomials. In particular, a univariate integer-valued polynomial of degree k can uniquely be written as an integer linear combination $\sum_{r=0}^{k} f_r \binom{x}{r}$, where $\binom{x}{r}$ is the binomial polynomial $\frac{x(x-1)\cdots(x-r+1)}{r!}$ and is always integer valued. By convention, we take $\binom{x}{0} = 1$ and $\binom{x}{1} = x$.

Linear Recurrence Sequences

Linear recurrence sequences enable us to describe solution sets of certain Diophantine equations that arise in our analysis.

A linear recurrence relation over $\mathbb{Q}$ is an equation of the form
$$u_{n+d} = a_1 u_{n+d-1} + \cdots + a_d u_n \,, \tag{1}$$
where $a_1, \ldots, a_d \in \mathbb{Q}$ and $a_d \neq 0$. The initial values $u_0, u_1, \ldots, u_{d-1}$ and (1) together uniquely define a sequence of rationals $\langle u_n\rangle_{n=0}^{\infty}$ as well as a bi-sequence $\langle u_n\rangle_{n=-\infty}^{\infty}$. We refer to the former as a *linear recurrence sequence* (LRS) and the latter as a *linear recurrence bi-sequence* (LRBS). The smallest integer d for which a sequence obeys a relation of the form (1) is the *order* of the sequence.

Note that if $a_1, \ldots, a_d, u_0, \ldots, u_{d-1} \in \mathbb{Z}$ and $a_d = \pm 1$, then (1) implies that the LRBS $\boldsymbol{u} = \langle u_n\rangle_{n=-\infty}^{\infty}$ is entirely contained in $\mathbb{Z}$. In fact, an old result of Fatou [6] (see also [3, Chapter 7]) implies that an LRBS $\boldsymbol{u}$ satisfying (1), with $a_1, \ldots, a_d, u_0, \ldots, u_{d-1} \in \mathbb{Z}$, is contained in $\mathbb{Z}$ if and only if $a_0 = \pm 1$. In this case, we say that $\boldsymbol{u}$ is *reversible*.

Note that $\boldsymbol{u}$ has an exponential-polynomial form
$$u_n = \sum_{i=1}^{s} P_i(n)\lambda_i^n$$
where λ_i are *characteristic roots*, that is, roots of the *characteristic polynomial*
$$g(x) = x^d - a_{d-1}x^{d-1} - \cdots - a_0 \,,$$
and P_i are polynomials with algebraic coefficients with degree one less than the multiplicity of λ_i as a root of g. We say that $\boldsymbol{u}$ is *simple* if none of the roots of g are repeated, which in turn is equivalent to each P_i being constant.

Diophantine Equations

A Diophantine equation is a multivariate polynomial equality with integer coefficients for which one seeks integer solutions. In this section we detail some Diophantine equations that arise later, and how to obtain their solution sets.

Definition 2.1. A *Pell equation* is a Diophantine equation of the form $w^2 - nz^2 = 1$, where the coefficient $n > 0$ is required not to be

a perfect square. The solution (w_0, z_0) with $w_0, z_0 > 0$ which minimises w is called its *fundamental solution*. Diophantine equations of the form $w^2 - nz^2 = N$ (where $n > 0$ is not a perfect square and $N \neq 0$) are called *generalised Pell equations*.

The history of Pell equations goes back to ancient times, and it is well known that the fundamental solution exists and can be computed. We refer the reader to [11] for a comprehensive account and modern developments. We are specifically interested in [11, Chap. 16.3], which shows that the fundamental solution (w_0, z_0) is the one for which w_0, z_0 are positive and $w_0 + z_0\sqrt{n}$ is minimal.

Lemma 2.2. *[11, Thm. 16.3] Consider the generalised Pell equation $w^2 - nz^2 = N$, and let (w_0, z_0) be the fundamental solution of $w^2 - nz^2 = 1$. We can compute a finite set S of* generating *pairs (w_i, z_i) such that every solution (w', z') to the above generalised Pell equation satisfies $w' + z'\sqrt{n} = (w_i + z_i\sqrt{n})(w_0 + z_0\sqrt{n})^m$ for some $(w_i, z_i) \in S$ and $m \in \mathbb{Z}$.*

Corollary 2.3. *The set of solutions (w, z) to a generalised Pell equation $w^2 - nz^2 = N$ is obtained as a finite union of pairs of simple reversible LRBS.*

Proof. By Lem. 2.2 every solution $(w_{i,m}, z_{i,m})$ satisfies

$$w_{i,m} + z_{i,m}\sqrt{n} = (w_i + z_i\sqrt{n})(w_0 + z_0\sqrt{n})^m \tag{2}$$

for $m \in \mathbb{Z}$ and a finite set of pairs (w_i, z_i). By equating coefficients of 1 and $\sqrt{n}$ in (2) and using $w_0^2 - nz_0^2 = 1$, one verifies that

$$w_{i,m+2} = 2w_0 w_{i,m+1} - w_{i,m}$$
$$z_{i,m+2} = 2w_0 z_{i,m+1} - z_{i,m}\,,$$

so each $\langle w_{i,m}\rangle_{m=-\infty}^{\infty}$ and $\langle z_{i,m}\rangle_{m=-\infty}^{\infty}$ define reversible LRBS. These could only fail to be simple if the discriminant $4w_0^2 - 4 = 0$, i.e., if $w_0 = 1$. But this would imply $z_0 = 0$, contradicting the strict positivity of z_0. Therefore each $\langle w_{i,m}\rangle_{m=-\infty}^{\infty}$ and $\langle z_{i,m}\rangle_{m=-\infty}^{\infty}$ define simple reversible LRBS. □

Lemma 2.4. *Consider the system of simultaneous generalised Pell equations $w^2 - n_1 z_1^2 = N_1, w^2 - n_2 z_2^2 = N_2$, where $n_1 n_2$ is not a perfect square and $N_1 \neq N_2$. This system has only finitely many solutions which can moreover be effectively enumerated.*

Proof. Writing $z = n_1 n_2 z_1 z_2$, it suffices to prove that $z^2 = n_1 n_2 (w^2 - N_1)(w^2 - N_2)$ has only finitely many solutions which can moreover be effectively enumerated. This is done by a direct application of [1] or [2, Thm. 4.2][3] (see in particular the comment at the beginning of the proof, which clarifies that the case $Y^2 = c(X - \alpha_1)\cdots(X - \alpha_n)$ is also handled by the proof). □

We remark that algorithms to find solutions have been further refined, see e.g., [22, 23]. The work of Baker [1] implies the following lemma for so-called hyperelliptic equations.

Lemma 2.5. *Let $k \geq 2$, $j \geq 3$, $N \neq 0$, and n be integers. The Diophantine equation $w^k = nz^j + N$ has only finitely many solutions, which can moreover be effectively enumerated.*

[3]The original 1975 print makes a mistake of omission in the statement of the theorem, which was subsequently corrected in later editions.

3 FIXED-POWER PREDICATES

In this section, we prove the following theorem regarding single-variable Presburger arithmetic, where $\equiv_m$ is a binary relation symbol denoting congruence modulo m, and the predicate $\mathbf{Z}^k$ is the set $\{n^k \mid n \in \mathbb{Z}\}$ of perfect k-th powers.

Theorem 3.1. *The theory $FO^1\langle \mathbb{Z}; 0, 1, +, -, <, (\equiv_m)_{m\geq 2}, (\mathbf{Z}^k)_{k\geq 2}\rangle$ is decidable.*

The core subroutine in the decision procedure solves systems of Diophantine equations of the form $w^k - nz^j = N$. We begin by describing the pre-processing that leads to its invocation, which is summarised in the following proposition.

Proposition 3.2. *The decision problem in Thm. 3.1 Turing-reduces to deciding whether there exists $x \in \mathbb{Z}$ that satisfies a given set of constraints, which includes exactly one constraint of the form $x > c$, and constraints of the form $\mathbf{Z}^k(ax + b)$ and $\neg\mathbf{Z}^k(ax + b)$, where all coefficients are strictly positive.*

Proof. We first prove that deciding the theory indeed reduces to a constraint satisfaction problem. Any sentence in the theory may be written in prenex normal form as $Qx\,.\,\psi$ where Q is a quantifier and ψ is a quantifier-free formula. Note that since $\forall x\,.\,\psi$ is equivalent to $\neg\exists x\,.\,\neg\psi$, we may reduce to deciding the truth of existential sentences, i.e., when Q is $\exists$. By putting ψ in disjunctive normal form, we may rewrite $\exists x\,.\,\psi$ as $\exists x\,.\,\bigvee_i \varphi_i$ where each φ_i is a conjunction of literals, and this may further be rewritten as $\bigvee_i \exists x\,.\,\varphi_i$. In this way we reduce to deciding the satisfiability of a given conjunction of literals, and we proceed by analysing the constraints that may arise from literals in the theory.

By suitably rearranging and simplifying, we can assume that literals are of the form $x = c$, $x < c$, $x > c$, $x \equiv_m c$, $\mathbf{Z}^k(ax + b)$, and $\neg\mathbf{Z}^k(ax + b)$. Observe that this requires rewriting $\neg(x = c)$ as $(x < c) \vee (x > c)$, $\neg(ax \equiv_m b)$ as $\bigvee_{r=0, r\neq b}^{m-1} ax \equiv_m r$, and $ax \equiv_m b$ as $\bigvee_{r=0, ar\equiv_m b}^{m-1} x \equiv_m r$.

To eliminate the modular-arithmetic constraints, we use an extended version of the Chinese Remainder Theorem (see e.g., [12, Thm. 3.12]) to coalesce the modular-arithmetic constraints $x \equiv_{m_i} c_i$ into a single conjunct $x \equiv_M r$, or prove that they are infeasible. We now make this constraint implicit by replacing all occurrences of x by $My + r$, and simplifying the resulting expressions.

We can assume $x = c$ does not occur, and that the only inequality that occurs is $x > c$ for some $c \in \mathbb{Z}$. (If a term $x = c$ does occur, we simply perform the obvious substitution, reducing to a quantifier-free formula.) If the remaining conjuncts imply that x is in a bounded interval, i.e., there are terms $x > c_1$ and $x < c_2$, this case is readily solved by finite inspection. Thus at most one inequality appears. If no such constraint occurs, we simply case split by considering in turn $x < 0$, and $x = 0$, and $x > 0$. This proves our claim. If no such inequality occurs and $x < c$, the obvious linear substitution $x \mapsto -x$ turns this term into one of the form $x > c$.

We finally address the positivity of the coefficients of x in the power predicates. If k is odd, we can assume that in all instances (positive and negative) of $\mathbf{Z}^k(ax + b)$, the coefficient a is positive, by possibly replacing $\mathbf{Z}^k(ax + b)$ by $\mathbf{Z}^k(-ax - b)$. If k is even and $a < 0$ in a positive occurrence of such an atom, we have an upper bound on x (as x is assumed to be lower-bounded), and the conjunction

can be handled trivially. If $a < 0$ for a term $\neg \mathbf{Z}^k(ax+b)$, this term always holds when x exceeds a computable bound and can thus be disposed of straightforwardly. We can therefore reduce to the case where all coefficients of x in the power predicates are positive. □

We make a further observation: certain power constraints can be "redundant" in view of other power constraints. For example, consider the following three constraints: $\mathbf{Z}^2(x)$, $\mathbf{Z}^2(3x)$, and $\mathbf{Z}^4(16x)$. If $\mathbf{Z}^4(16x)$ holds, then $\mathbf{Z}^2(x)$ also holds and $\mathbf{Z}^2(3x)$ does not. We therefore say that $\mathbf{Z}^2(x)$ and $\mathbf{Z}^2(3x)$ are both *redundant with respect to* $\mathbf{Z}^4(16x)$, since the (positive) truth of $\mathbf{Z}^4(16x)$ uniquely determines the truth values of the other two constraints. We formalise this idea in the following definition:

Definition 3.3. The constraint $\mathbf{Z}^k(cx+d)$ is *redundant with respect to* $\mathbf{Z}^j(ax+b)$ if $k \mid j$ and $ad = bc$.

In general, if $\mathbf{Z}^k(cx+d)$ is redundant with respect to $\mathbf{Z}^j(ax+b)$ then the (positive) truth of $\mathbf{Z}^j(ax+b)$ determines the truth value of $\mathbf{Z}^k(cx+d)$. Indeed, if $ax+b$ is a perfect j-th power, then in particular it is a perfect k-th power, and so is $c^k(ax+b) = ac^{k-1}(cx+d)$ (using $ad = bc$). We thus have that $cx+d$ is a perfect k-th power if and only if ac^{k-1} is, and the latter can be effectively checked. The above discussion shows that we can identify and discard (positive or negative) constraints that are redundant with respect to some given positive constraint.

In the same vein, we define the notion of *similar* constraints.

Definition 3.4. The constraint $\mathbf{Z}^j(cx+d)$ is *similar* to a constraint $\mathbf{Z}^k(ax+b)$ if $ad = bc$.

Note that the notion of being similar is a transitive property. Though similar constraints cannot be as immediately discarded as redundant constraints, we will show that a conjunction of similar positive constraints can be coalesced into a single positive constraint.

For the proof, we require the notion of *p-adic valuation*. Recall that for a prime p, the p-adic valuation of a non-zero integer n, denoted $\mathfrak{v}_p(n)$, is equal to the highest power of p that divides n, e.g., $\mathfrak{v}_2(20) = 2$. We take $\mathfrak{v}_p(0) = \infty$. The valuation $\mathfrak{v}_p$ extends to rational numbers as $\mathfrak{v}_p(m/n) = \mathfrak{v}_p(m) - \mathfrak{v}_p(n)$.

Lemma 3.5. *Given similar constraints $\mathbf{Z}^{k_1}(a_1x+b_1), \ldots, \mathbf{Z}^{k_l}(a_lx+b_l)$, either they are not simultaneously satisfiable, or we may find a constraint $\mathbf{Z}^K(Ax+B)$ that is satisfied if and only if $\mathbf{Z}^{k_i}(a_ix+b_i)$ is satisfied for all $1 \le i \le l$. Furthermore, $K = \mathrm{lcm}(k_1, \ldots, k_l)$, where* lcm *denotes the least common multiple.*

Proof. Assume that $\mathbf{Z}^{k_1}(a_1x+b_1), \ldots, \mathbf{Z}^{k_l}(a_lx+b_l)$ are similar, and let b/a with $a > 0$ be the reduced form of the rational constant b_i/a_i (which is independent of i as $a_ib_j = b_ia_j$ for all $1 \le i, j \le l$).

If all constraints are satisfied, we have that $\mathfrak{v}_p(a_ix+b_i) \equiv 0 \mod k_i$ for all primes p and indices $i = 1, \ldots, l$. This can be rearranged as

$$\mathfrak{v}_p(ax+b) \equiv \mathfrak{v}_p(a) - \mathfrak{v}_p(a_i) \mod k_i \tag{3}$$

for each p, i. Note that $\mathfrak{v}_p(a) - \mathfrak{v}_p(a_i) \equiv 0 \mod k_i$ holds for all i for all but finitely many primes that divide some a_i — let us call such primes *interesting*. For each interesting prime p, we apply the (extended) Chinese Remainder Theorem [12, Thm. 3.12] to determine whether the constraints given by (3) for $i = 1, \ldots, l$ are simultaneously satisfiable, and if so, compute a residue r_p such that they hold if and only if $\mathfrak{v}_p(ax+b) \equiv -r_p \mod K$, where $K = \mathrm{lcm}(k_1, \ldots, k_l)$. This is equivalent to $(ax+b)\prod_p p^{r_p}$ being a perfect K-th power. By construction, $\mathbf{Z}^K((ax+b)\prod_p p^{r_p})$ holds if and only if $\mathbf{Z}^{k_i}(a_ix+b_i)$ holds for all $1 \le i \le l$. □

Remark 1. If $\mathbf{Z}^{k_1}(a_1x+b_1), \ldots, \mathbf{Z}^{k_l}(a_lx+b_l)$ are similar and simultaneously satisfiable, by Lem. 3.5 we may discard each $\mathbf{Z}^{k_i}(a_ix+b_i)$ to be replaced by the single constraint $\mathbf{Z}^K(Ax+B)$. We say that we *coalesce* $\mathbf{Z}^{k_1}(a_1x+b_1), \ldots, \mathbf{Z}^{k_l}(a_lx+b_l)$ into $\mathbf{Z}^K(Ax+B)$.

In the sequel, we shall assume there are no pairs of redundant constraints or similar positive constraints: we first discard redundant constraints, then coalesce similar positive constraints into a single positive constraint, and then again discard redundant constraints. As an example, if we had $\mathbf{Z}^2(5x), \mathbf{Z}^3(4x), \neg\mathbf{Z}^6(24x)$, the last constraint becomes redundant only after the first two are coalesced into $\mathbf{Z}^6(500x)$.

Once the pre-processing step is completed, our strategy for solving the satisfiability problem consists in handling the various constraints ($x > c$, positive constraints of the form $\mathbf{Z}^k(ax+b)$, and negative constraints of the form $\neg\mathbf{Z}^k(ax+b)$) sequentially in the order given. More precisely, starting with a candidate solution set for x of $\mathbb{Z}$, we iteratively shrink this solution set until a definitive conclusion can be drawn, i.e., the solution set either becomes empty (or finite, in which case we finalise the decision by inspection), or infinite and no further positive constraints remain, at which point we conclude that the system is satisfiable. Intuitively, the justification of correctness is as follows. Denote the solution set at a given stage by S. If S is infinite, then upon taking account of a further positive constraint of the form $\mathbf{Z}^k(ax+b)$, the resulting new solution set can be computed and furthermore is either finite, or is infinite and has relative density 0 in S. If the solution set is infinite and no further positive constraint remains, we can show that no conjunction of negative constraints can fully deplete the solution set, and the system is therefore automatically satisfiable.

To formalise this argument, we first analyse the solution sets arising from positive constraints.

Proposition 3.6. *The solution set of all $x \in \mathbb{Z}$ satisfying l non-similar constraints of the form $\mathbf{Z}^{k_i}(a_ix+b_i)$ with $a_i > 0$ for $1 \le i \le l$ is effectively computable and has the following structure.*

1. *If $l = 0$ then $S = \mathbb{Z}$.*
2. *If $l = 1$ then either $S = \emptyset$ or S is a finite union of sets of the form $f(\mathbb{Z})$ where f is a polynomial of degree k_1 with a positive leading coefficient.*
3. *If $l = 2$ then either $S = \emptyset$ or one of the following holds.*
   (a) *If $\max\{k_1, k_2\} > 2$ then S is finite.*
   (b) *If $k_1, k_2 = 2$ then S is a union of finitely many simple reversible LRBS.*
4. *If $l \ge 3$ then S is finite.*

Proof. **Case (1)** is obvious.

Case (2): If $l = 1$ then $S = \emptyset$ if b_1 is not a k_1-th power mod a_1, and this can be checked algorithmically by enumerating all k_1-th powers mod a_1. Conversely, if b_1 is a k_1-th power mod a_1 then there are infinitely many solutions and they can be parametrised

as follows. Pick $u \in \mathbb{N}$ such that $u^{k_1} \equiv b_1 \bmod a_1$. Then consider the polynomial $f_{u,k_1,a_1} \in \mathbb{Z}[t]$ that satisfies

$$(u + ta_1)^{k_1} = a_1 f_{u,k_1,a_1}(t) + b_1 .$$

Then f_{u,k_1,a_1} has degree k_1 and a positive leading coefficient (as $a_1 > 0$). Every $t \in \mathbb{Z}$ gives rise to a solution $x = f_{u,k_1,a_1}(t)$ of $\mathbf{Z}^k(a_1x + b_1)$. Conversely, whenever $x \in \mathbb{Z}$ is a solution, then $(a_1x + b_1)^{1/k_1}$ must be of the form $u + ta_1$ for some u such that $u^{k_1} \equiv b_1 \bmod a_1$ and some $t \in \mathbb{Z}$. Therefore the solution set is exactly

$$S = \bigcup_{\substack{0 \le u < a_1 \\ u^{k_1} \equiv b_1 \bmod a_1}} f_{u,k_1,a_1}(\mathbb{Z}) .$$

Case (3): Suppose $l = 2$ and (without loss of generality) $k_2 \ge k_1$. We have the system

$$a_1x + b_1 = y_1^{k_1} \ \wedge \ a_2x + b_2 = y_2^{k_2} . \tag{4}$$

By taking a linear combination, we eliminate x to obtain the system

$$a_2y_1^{k_1} - a_1y_2^{k_2} = a_2b_1 - a_1b_2 \tag{5}$$

$$y_1^{k_1} \equiv_{a_1} b_1 \wedge y_2^{k_2} \equiv_{a_2} b_2 . \tag{6}$$

By multiplying (5) by $a_2^{k_1-1}$, and setting $w \coloneqq a_2y_1$ and $z \coloneqq y_2$ we obtain the system

$$w^{k_1} - (a_1a_2^{k_1-1})z^{k_2} = a_2^{k_1-1}(a_2b_1 - a_1b_2) \tag{7}$$

$$\bigvee_{\substack{0 \le r_i < a_i \\ r_1^{k_1} \equiv_{a_1} b_1, r_2^{k_2} \equiv_{a_2} b_2}} w \equiv_{a_1a_2} a_2r_1 \ \wedge \ z \equiv_{a_2} r_2 . \tag{8}$$

Case (3a): If $k_2 \ge 3$ then (7) satisfies the conditions of Lem. 2.5 (note that $a_2b_1 - a_1b_2 \neq 0$ by non-similarity) and so there are finitely many effectively computable solutions, which may be checked against the modular constraints.

Case (3b): Suppose $k_1 = k_2 = 2$. If $a_1a_2^{k_1-1}$ is a perfect square, then the left-hand side of (7) may be factored using the difference of two squares, and by considering prime factorisations there are finitely many solutions. Otherwise, if $a_1a_2^{k_1-1}$ is not a perfect square, (7) comprises a generalised Pell equation, and by Cor. 2.3 the solutions (w, z) are exactly the value sets of finitely many pairs of simple reversible LRBS.

Since reversible LRBS are periodic modulo N for any $N \ge 1$, given an LRBS $\boldsymbol{u}$ one can effectively find an integer M such that each subsequence $\langle u_{Mn+s} \rangle_{n=-\infty}^{\infty}$ is constant mod a_1a_2 for each $0 \le s \le M - 1$. Therefore the solution set of pairs (w, z) satisfying (7) and (8) is comprised of a finite union of such subsequences (or possibly the empty set if none of the modular constraints are satisfied).

For every solution pair (w, z), we recover a solution x to the original system (4) by $x = \frac{z^{k_2} - b_2}{a_2}$. Since for any simple LRBS $\boldsymbol{u}$, we have $\frac{u^{k_2} - b_2}{a_2}$ is also a simple LRBS, and moreover for any z satisfying the modular constraints (8) we have $\frac{z^{k_2} - b_2}{a_2}$ is an integer, the set S of solutions x is a union of finitely many simple reversible LRBS.

Case (4): If $l \ge 3$ then if $\max\{k_1, k_2, k_3\} \ge 3$ then S is finite and effectively computable by Case (3). Otherwise, $k_1 = k_2 = k_3 = 2$, and by the same process as in Case (3) we obtain a system

$$a_2y_1^2 - a_1y_2^2 = a_2b_1 - a_1b_2 \tag{9}$$

$$a_3y_1^2 - a_1y_3^2 = a_3b_1 - a_1b_3 \tag{10}$$

along with some modular constraints which we omit. By multiplying (9) by $a_2a_3^2$ and (10) by $a_3a_2^2$, and setting $w \coloneqq a_2a_3y_1$, $z_2 \coloneqq y_2$, $z_3 \coloneqq y_3$, we obtain the simultaneous equations

$$w^2 - a_1a_2a_3^2z_2^2 = a_2a_3^2(a_2b_1 - a_1b_2) \tag{11}$$

$$w^2 - a_1a_2^2a_3z_3^2 = a_2^2a_3(a_3b_1 - a_1b_3) . \tag{12}$$

If either $a_1a_2a_3^2$ or $a_1a_2^2a_3$ is a perfect square then there are finitely many solutions by factoring the left-hand side of (11) or (12) using the difference of two squares. Otherwise, noting that $a_2b_1 - a_1b_2$ and $a_3b_1 - a_1b_3$ are non-zero by non-similarity, we have a system of simultaneous Pell equations in w, z_2, z_3, which has finitely many effectively computable solutions by Lem. 2.4, so S is finite and effectively computable. □

Prop. 3.6 already gives an algorithm to decide the satisfiability of any set of constraints of the form $x > c$ and positive constraints $\mathbf{Z}^k(ax + b)$. We now show that when the solution set is infinite, negative constraints can only be violated on a subset of relative density 0 within the solution set, meaning that arbitrarily many negative constraints will still leave infinitely many solutions overall.

Definition 3.7. Let $\varnothing \neq T \subseteq \mathbb{N}$ and $S \subseteq T$. Define the *(upper) density of S inside T* to be

$$\limsup_{n\to\infty} \frac{|S \cap [0, n]|}{|T \cap [0, n]|} .$$

First we need an elementary lemma. Given a function $f : \mathbb{Z} \to \mathbb{Z}$, define $S_f(c, n) = \mathrm{Im}(f) \cap [c, n]$.

Lemma 3.8. *Suppose $f \in \mathbb{Z}[x]$ has degree d and a positive leading coefficient, $c_1, c_2, N > 0$, and $g : \mathbb{Z} \to \mathbb{Z}$ is a function such that for all $|y| \ge N$ we have $|g(y)| > c_2y^{d+1}$. Then we have*

$$\limsup_{n\to\infty} \frac{|S_g(c_1, n) \cap S_f(c_1, n)|}{|S_f(c_1, n)|} = 0 .$$

Proof. It is sufficient to show that

$$\limsup_{n\to\infty} \frac{|S_g(0, n)|}{|S_f(0, n)|} = 0 . \tag{13}$$

There is a constant $C > 0$ such that $f(y) \le Cy^d$ for all $y \in \mathbb{N}$. Thus $|S_f(0, n)| \ge (n/C)^{1/d}$ while $|S_g(0, n)| \le N + (n/c_2)^{1/(d+1)}$. Then the result follows when considering $|S_f(0, n)| > 0$

$$\frac{|S_g(0, n)|}{|S_f(0, n)|} \le \frac{N + (n/c_2)^{1/(d+1)}}{(n/C)^{1/d}} .$$ □

Proposition 3.9. *Let S be the set of solutions x to a system of constraints given by $x > c$ and non-similar constraints $\mathbf{Z}^{k_i}(a_ix + b_i)$ for $i = 1, \ldots, l$, $a_i > 0$. If S is infinite, then the subset $S' \subseteq S$ for which any non-redundant negative constraint $\neg\mathbf{Z}^k(ax + b)$ is violated has density 0 relative to S.*

Proof. A negative constraint $\neg\mathbf{Z}^k(cx+d)$ being violated is equivalent to the positive constraint $\mathbf{Z}^k(cx+d)$ holding. It is sufficient to prove the result for a single negative constraint $\neg\mathbf{Z}^k(cx+d)$ as the union of finitely many null-density sets again has null density. We go through the cases given by Prop. 3.6.

Case (1): $l = 0$. Then $S = [c+1, \infty)$. The discarded set S' with the constraint $\mathbf{Z}^k(ax+b)$ added becomes a finite union of sets of the form $g(\mathbb{Z}) \cap [c+1, \infty)$ where g is a polynomial with $\deg g = k \geq 2$ and positive leading coefficient, by Prop. 3.6. Apply Lem. 3.8 with $f(y) = y$ and each g to conclude S' has density 0 relative to S.

Case (2): $l = 1$. First, suppose $\mathbf{Z}^k(ax+b)$ is similar (but not redundant) to $\mathbf{Z}^{k_1}(a_1x+b_1)$. By Lem. 3.5, these constraints get coalesced into $\mathbf{Z}^K(Ax+B)$ where K is the least common multiple of k and k_1. Then the solution set S to $x > c$ and $\mathbf{Z}^{k_1}(a_1x+b_1)$ is a finite union of sets of the form $f(\mathbb{Z}) \cap [c+1, \infty)$ for a polynomial f with positive leading coefficient and degree k_1, and the discarded set S' to $x > c$ and $\mathbf{Z}^K(Ax+B)$ is a finite union of sets of the form $g(\mathbb{Z}) \cap [c+1, \infty)$ for polynomials g of degree $K > k_1$. Therefore we may again apply Lem. 3.8 with f, g to conclude S' has density 0 relative to S.

Otherwise, suppose $\mathbf{Z}^k(ax+b)$ is not similar to $\mathbf{Z}^{k_1}(a_1x+b_1)$. Then by Prop. 3.6, if the discarded set S' to $x > c, \mathbf{Z}^{k_1}(a_1x+b_1), \mathbf{Z}^k(ax+b)$ is infinite, then $k = k_1 = 2$ and S' is a union of finitely many simple reversible LRBS restricted to $[c+1, \infty)$. For any simple reversible LRBS $\langle u_n \rangle_{n=-\infty}^{\infty}$, we have that $|u_n|$ grows exponentially as $|n| \to \infty$. Therefore we may apply Lem. 3.8 to f and the function $g(y) = u_y$ for each LRBS $\boldsymbol{u}$ forming part of the discarded set of S', to get that S' has density 0 relative to S.

Case (3): $l = 2$. The only way in which S is infinite is if $k_1 = k_2 = 2$. In that case, it is impossible for S' to be infinite. Indeed, by Prop. 3.6 we have that S' is infinite only if there are at most two non-similar constraints among $\mathbf{Z}^{k_1}(a_1x+b_1), \mathbf{Z}^{k_2}(a_2x+b_2), \mathbf{Z}^k(ax+b)$, meaning that (without loss of generality) $\mathbf{Z}^k(ax+b)$ is similar to $\mathbf{Z}^{k_1}(a_1x+b_1)$. By Lem. 3.5 we can coalesce these constraints into $\mathbf{Z}^K(Ax+B)$, where K is the least common multiple of k, k_1. But by non-redundancy, we cannot have $k \mid k_1$ nor $k_1 \mid k$ so $K > k_1 = 2$. Therefore by Prop. 3.6 S' is finite, and so trivially has density 0 relative to S. □

Prop. 3.9 was the final step in our proof of Thm. 3.1, to the effect that $FO^1\langle \mathbb{Z}; 0, 1, +, -, <, (\equiv_m)_{m>1}, (\mathbf{Z}^k)_{k>1} \rangle$ is decidable. Indeed, in summary, we use Prop. 3.2 to reduce to considering satisfiability of constraints of the form $x > c, \mathbf{Z}^k(ax+b), \neg\mathbf{Z}^k(ax+b)$. We may further reduce to the case in which the positive constraints $\mathbf{Z}^k(ax+b)$ are all non-similar, and all power constraints are non-redundant with respect to each other. Prop. 3.6 shows that the solution set to any number of positive constraints $\mathbf{Z}^k(ax+b)$ is effectively computable, and in the case for which the solution set is infinite, Prop. 3.9 shows that the addition of any negative constraints $\neg\mathbf{Z}^k(ax+b)$ removes at most a subset of null density, so the solution set remains infinite. Meanwhile if the solution set is finite, one solves the decision problem by simply enumerating every solution and checking against all constraints.

4 QUADRATIC AND CUBIC PREDICATES

In this section, we adapt the techniques used to prove Thm. 3.1 to decide single-variable Presburger arithmetic expanded with multiple predicates $(\mathcal{R}_i)_i$, where each predicate $\mathcal{R}_i$ corresponds to the value set of an integer-valued univariate polynomial f_i of degree at most 3, i.e., $\mathcal{R}_i(x)$ holds if and only if there exists an integer u such that $f_i(u) = x$. Formally, we prove the following.

Theorem 4.1. *Let $(\mathcal{R}_i)_i$ be predicates corresponding to value sets of integer-valued polynomials $(f_i)_i$ of degree at most 3. Then the theory $FO^1\langle \mathbb{Z}; 0, 1, +, -, <, (\equiv_m)_{m\geq 2}, (\mathcal{R}_i)_i \rangle$ is decidable.*

Before proceeding with the technical proof, we record a few simplifying assumptions. These assumptions establish an analogue of Prop. 3.2 (i.e., the pre-processing step) *mutatis mutandis*. Moreover, a polynomial f_i of degree 0 is constant and so $\mathcal{R}_i(ax+b)$ is equivalent to $ax+b = f_i(0)$, and for a polynomial $f_i(t) = dt+e$ of degree 1, $\mathcal{R}_i(ax+b)$ is equivalent to $ax+b \equiv e \pmod{d}$. Hence we can assume that the polynomials f_i are of degree 2 or 3. In summary, we obtain the following.

Lemma 4.2. *The decision problem in Thm. 4.1 Turing-reduces to deciding whether there exists $x > 0$ satisfying a conjunction of exactly one constraint $x > c$ together with other constraints of the form $\mathcal{R}_i(ax+b)$ and $\neg\mathcal{R}_i(ax+b)$, where $\mathcal{R}_i = f_i(\mathbb{Z})$ for a polynomial f_i of degree two or three.*

Next we want to restrict the kinds of polynomials that can appear. Recall that a polynomial $f(t) = c_dt^d + c_{d-1}t^{d-1} + \cdots + c_0$ is *depressed* if its second-highest coefficient, c_{d-1}, is zero.

Let $\mathcal{R}(ax+b, q, r)$ denote the predicate $\{f(qu+r) \mid u \in \mathbb{Z}\}$. We add this ternary predicate in our signature and henceforth always assume that all our polynomials are depressed and monic.

Lemma 4.3. *For any integer-valued polynomial f of degree at most 3 and corresponding predicate $\mathcal{R}$ and constants $a, b \in \mathbb{Z}$, we can compute a depressed monic polynomial $\tilde{f}$ with integer coefficients, together with constants $\tilde{a}, \tilde{b}, q, r$, such that for all x, $\mathcal{R}(ax+b)$ is equivalent to $\tilde{\mathcal{R}}(\tilde{a}x+\tilde{b}, q, r)$.*

Proof. We tackle the degree-3 case, the degree-2 case being similar and simpler.

The predicate $\mathcal{R}(ax+b)$ is equivalent to $\exists u\,.\, c_3u^3 + c_2u^2 + c_1u + c_0 = ax+b$. We can multiply through by an appropriate integer and assume without loss of generality that $c_3, \ldots, c_0$ are integers, and c_3 is positive. We "complete the cube" by multiplying through by $27c_3^2$ and write the equivalent statement

$$\exists u\,.\,(3c_3u+c_2)^3 + 27c_1c_3^2u + 27c_0c_3^2 - 9c_2^2c_3u - c_2^3 = 27c_3^2(ax+b),$$

which can be further rearranged as

$$\begin{aligned}&\exists u\,.\,(3c_3u+c_2)^3 + (9c_1c_3 - 3c_2^2)(3c_3u+c_2)\\ &= 27ac_3^2x + (27bc_3^2 - 27c_0c_3^2 + 9c_1c_2c_3 - 2c_2^3),\end{aligned}$$

or in other words $\exists u\,.\,\tilde{f}(3c_3u+c_2) = \tilde{a}x + \tilde{b}$. □

We continue following a very similar strategy to find a witness x that satisfies all constraints as in Sec. 3. Each predicate $\mathcal{R}_i$ can occur both positively and negatively, and we want to show that we can enumerate a solution set satisfying positive constraints when

it is finite; and when it is infinite, adding a further "non-similar" positive constraint would in all but one case restrict the solution set to a subset of relative null density. In the exceptional case, the solutions to the positive constraints are parametrised as an LRS, and the discarded indices form arithmetic progressions. Thus, in this case too, we can effectively determine whether there remains a value of x not discarded by the negative constraints.

As we did previously, we need to account for redundant constraints, but it is not immediately clear what a meaningful definition of redundancy is. Recall that a multivariate polynomial is *absolutely irreducible* if it is irreducible over the complex numbers.

Definition 4.4. The constraint $\mathcal{R}_1(a_1x + b_1)$ is *redundant with respect to* $\mathcal{R}_2(a_2x+b_2)$ if $\deg(f_1) \mid \deg(f_2)$ and $a_2f_1(u_1)-a_1f_2(u_2)-a_2b_1+a_1b_2$ is not absolutely irreducible.

As the polynomials defining our predicates have degrees 2 or 3, two constraints can only be in a redundancy relationship if the underlying polynomials have the same degree. We now have:

Proposition 4.5. *Let $\mathcal{R}_1(a_1x + b_1)$ be redundant with respect to $\mathcal{R}_2(a_2x + b_2)$. Then $a_1b_2 = a_2b_1$.*

Proof. Recall from Lem. 4.3, we assume that f_1 and f_2 are depressed and satisfy $f_i(0) = 0$. Thus, when $\deg(f_1) = \deg(f_2) = 2$, write $f_i(u_i) = c_iu_i^2$ for $i = 1, 2$. Then we have, by the definition of absolutely reducibility:

$$A_1u_1^2 - A_2u_2^2 + a_1b_2 - a_2b_1 = A_1(u_1 + C_2u_2 + D)(u_1 + C_2'u_2 + D'),$$

where $A_1 = a_2c_1$ and $A_2 = a_1c_2$ are non-zero. Then, $A_1C_2C_2' = -A_2$, $C_2 + C_2' = 0$, $D + D' = 0$, $DC_2' + D'C_2 = 0$, and $A_1DD' = a_1b_2 - a_2b_1$. Thus the second and third equations imply that $D = -D'$ and $C_2 = -C_2'$, and the first implies that $C_2 \neq 0$ as $A_2 \neq 0$. Hence the fourth equation implies that $D = 0$ and so the last equation yields $a_1b_2 - a_2b_1 = 0$.

When $\deg(f_1) = \deg(f_2) = 3$, let $f_i(u_i) = c_iu_i^3 + d_iu_i$ for $i = 1, 2$. Then we have by absolute reducibility:

$$\begin{aligned} &A_1u_1^3 + B_1u_1 - A_2u_2^3 - B_2u_2 + a_1b_2 - a_2b_1 \\ &= A_1(u_1 + C_2u_2 + D)(u_1^2 + E_1u_1u_2 + C_2'u_2^2 + E_2u_1 + E_3u_2 + D'), \end{aligned}$$

where $A_1 = a_2c_1$ and $A_2 = a_1c_2$ are non-zero, $B_1 = a_2d_1$ and $B_2 = a_1d_2$. Then, $A_1C_2C_2' = -A_2$ and $A_1DD' = a_1b_2 - a_2b_1$ and $E_1 + C_2$, $E_2 + D$, $C_2' + E_1C_2$, $E_3 + C_2E_2 + E_1D$, $C_2E_3 + DC_2'$ are all zero. Substituting $E_1 = -C_2$ and $E_2 = -D$ gives that $C_2' - C_2^2$, $E_3 - 2C_2D$, and $C_2E_3 + DC_2'$ are all zero. Hence, as $C_2 \neq 0$ as $A_2 \neq 0$, $E_3 = C_2D$, which forces that $C_2D = 0$ and thus that $D = 0$. Thus $a_1b_2 - a_2b_1 = 0$. □

In the case of redundancy with cubic polynomials, we can thus solve the (Diophantine) equation $a_2f_1(u_1) - a_1f_2(u_2) = 0$ via the factorisation $A_1x^3 + B_1x - A_2y^3 - B_2y = A_1(u_1 + mu_2)(u_1^2 - mu_1u_2 + m^2u_2^2 + B_1/A_1)$, where $m = -B_1/B_2 = -(A_2/A_1)^{1/3}$ is non-zero. This follows using the notation in the proof above. We shall assume that m is rational. This is of course the case when $B_2 \neq 0$, or when $B_1 = B_2 = 0$ and A_2/A_1 is a perfect cube. Otherwise, it is impossible for $A_1x^3 - A_2y^3 = 0$ to have integer solutions.

Looking at the proof of the lemma above, when two redundant predicates are both satisfied, a linear relationship between u_1 and u_2 has to hold, or in the cubic case, u_1 and u_2 have to lie on a certain conic. This conic is $u_1^2 - mu_1u_2 + m^2u_2^2 + B_1/A_1 = 0$, which represents an ellipse, which thus contains finitely many integer points (u_1, u_2) that we can effectively compute. Hence we conclude the following.

Lemma 4.6. *Let $\phi(x)$ be a conjunction of two redundant predicates $\mathcal{R}_1(a_1x + b_1, q_1, r_1)$ and either $\mathcal{R}_2(a_2x + b_2, q_2, r_2)$ or $\neg\mathcal{R}_2(a_2x + b_2, q_2, r_2)$. Then $\phi(x)$ can be written as the conjunction of one predicate $\mathcal{R}_3(a_3x + b_3, q_3, r_3)$ and a finite number of atoms definable in quantifier-free Presburger arithmetic.*

As we can observe from the factorisation, the solution set to $a_2f_1(u_1) - a_1f_2(u_2) = 0$ is the union of integer points on a line (passing through the origin and having rational slope $m = -s/t$), and finitely many integer points on a bounded conic. In other words, a solution u_1 corresponds to a solution u_2 if and only if the former takes one of finitely many values, or satisfies a set of divisibility constraints. The constraint on $a_2x + b_2$ is thus "redundant" in view of the constraint on $a_1x + b_1$ in the sense that it does not add "algebraic" information beyond modular-arithmetic annotation.

We can henceforth focus on the case where there is no redundancy, and at least one of the predicates corresponds to a cubic polynomial.

Proposition 4.7. *The solution set of all $x \in \mathbb{Z}$ satisfying l non-redundant constraints of the form $\mathcal{R}_i(a_ix + b_i, q_i, r_i)$ with $a_i > 0$ and $d_i = \deg(f_i)$ for $1 \leq i \leq l$ is effectively computable and has the following structure.*

1. *If $l = 0$ then $S = \mathbb{Z}$.*
2. *If $l = 1$ then either $S = \emptyset$ or S is a finite union of sets of the form $f(\mathbb{Z})$ where f is a polynomial of degree d_1 and with a positive leading coefficient.*
3. *If $l = 2$ then either $S = \emptyset$ or one of the following holds.*
   (a) *If $d_1 = d_2 = 3$ then S is finite.*
   (b) *If $d_1 = d_2 = 2$ then S is a union of finitely many simple reversible LRBS.*
   (c) *If $d_1 \neq d_2$ then either S is finite or S is a finite union of sets of the form $f(\mathbb{Z})$ where f is a polynomial of degree 6.*
4. *If $l = 3$ then either $S = \emptyset$ or one of the following holds.*
   (a) *If any two $i \neq j$ exist such that $d_i = 3 = d_j$, then S is finite.*
   (b) *If $d_1 = d_2 = d_3 = 2$ then S is finite.*
   (c) *Otherwise S is a union of finitely many simple reversible LRBS.*
5. *If $l \geq 4$ then S is finite.*

Proof. **Case (1)** is obvious.

Case (2): Write $f(u)$ for $f_1(q_1u + r_1)$. If $l = 1$ then $S = \emptyset$ if b_1 does not lie in the value set of f modulo a_1, and this can be checked algorithmically by enumerating all values of f modulo a_1. Conversely, if b_1 does lie in the value set of f modulo a_1 then there are infinitely many solutions and they can be parametrised as follows. Pick $u \in \mathbb{N}$ such that $f(u) \equiv b_1 \bmod a_1$. Then consider the polynomial $g_{u,d_1,a_1} \in \mathbb{Z}[t]$ that satisfies

$$f(u + ta_1) = a_1g_{u,k_1,a_1}(t) + b_1.$$

Then g_{u,k_1,a_1} has degree k_1 and a positive leading coefficient (as $a_1, q_1 > 0$). Every $t \in \mathbb{Z}$ gives rise to a solution $x = g_{u,k_1,a_1}(t)$ of $\mathcal{R}_1(a_1x + b_1, q_1, r_1)$. Conversely, whenever $x \in \mathbb{Z}$ is a solution, then there must exist some $y = u + ta_1$ such that $f(y) = a_1x + b_1$,

$f(u) \equiv b_1 \bmod a_1$, and $t \in \mathbb{Z}$. Therefore the solution set is exactly

$$S = \bigcup_{\substack{0 \le u < a_1 \\ f(u) \equiv b_1 \bmod a_1}} g_{u,k_1,a_1}(\mathbb{Z}) .$$

Case (3a) We claim that a pair of non-redundant cubic positive constraints $\mathcal{R}_1(a_1x+b_1, q_1, r_1), \mathcal{R}_2(a_2x+b_2, q_2, r_2)$ has finitely many solutions, which can moreover be effectively enumerated. Indeed, homogenising the cubic curve

$$a_2 f_1(u_1) - a_1 f_2(u_2) - a_2 b_1 + a_1 b_2 = 0$$

into the projective plane over an appropriate algebraic extension of the rationals gives an absolutely irreducible curve. This curve has three places at infinity, $([\rho_i : 1 : 0])_{i=1}^3$ where each ρ_i is a complex cube root of a_2/a_1 (because f_1 and f_2 are monic, evaluating this curve at $[u_1 : u_2 : 0]$ gives $a_2u_1^3 - a_1u_2^3 = 0$). If this curve has genus 1, then its finitely many integer points can be enumerated by [2, Thm. 4.3]. Otherwise, the curve has genus 0, in which case its finitely many integer points can be enumerated by [18].

Case (3b) Follows from Case 3b) of Prop 3.6 *mutatis mutandis*.

Case (3c) We now consider the case of two positive constraints, where the first $\mathcal{R}_1(a_1x + b_1, q_1, r_1)$ corresponds to a quadratic polynomial. The attendant curve $a_1f_2(u_2) - a_2f_1(u_1) = a_1b_2 - a_2b_1$ can then be simplified to have the form $(a_2u_1)^2 = a_2(a_1f_2(u_2) + a_2b_1 - a_1b_2) = a_1a_2g(u_2)$, where $g \in \mathbb{Q}[u_2]$ is monic, and $a_1g \in \mathbb{Z}[u_2]$. If g has three distinct roots (i.e., an elliptic curve has arisen), then the finitely many integer points on the curve can be enumerated by [1, Thm. 2].

Otherwise, g has a repeated root, which is necessarily rational because it corresponds to the common factor of g and its derivative. By Gauss's lemma, a_1g thus splits over $\mathbb{Z}$ as $(\alpha u_2 + \beta)^2(\gamma u_2 + \delta)$. We then make the substitution $v_1 = \frac{a_2u_1}{\alpha u_2+\beta}$, and observe that $v_1^2 = a_2(\gamma u_2+\delta)$. In this manner, the values of u_1 and u_2 are parametrised by v_1. If we have constraints that u_1 and u_2 are respectively r_1 modulo q_1 and r_2 modulo q_2, we enforce the following modular-arithmetic constraints on v_1:

$$v_1^2 \equiv a_2r_2\gamma + a_2\delta \mod a_2\gamma q_2,$$
$$\alpha v_1^3 + (a_2\beta\gamma - a_2\alpha\delta)v_1 \equiv a_2^2r_1\gamma \mod a_2^2\gamma q_1.$$

Inserting $v_1^2 = a_2(\gamma u_2 + \delta)$ into $(\alpha u_2 + \beta)^2(\gamma u_2 + \delta)$ we obtain that x is parametrised as a degree 6 polynomial in v_1.

We thus get solutions to our constraints whenever v_1 satisfies the above: such values for v_1, if they exist, form a union of finitely many arithmetic progressions by the Chinese Remainder Theorem.

Case (4a) immediately follows from Case (3a).

Case (4b) follows from Case 4 of Prop 3.6 *mutatis mutandis*.

Case(4c) We freely borrow notation from case (3c). Assume f_2 has degree 3 and that f_1 and f_3 have degree 2. Then, as in case (3c), we construct v_1 and v_3 that give the system

$$v_1^2 = a_2(\gamma_1u_2 + \delta_1),$$
$$v_3^2 = a_2(\gamma_3u_2 + \delta_3),$$

where as before, $v_i = \frac{a_iu_2}{\alpha_iu_2+\beta_i}$. Eliminating u_2 yields the equation $\gamma_3v_1^2 - \gamma_1v_3^2 = a_2(\gamma_3\delta_1 - \gamma_1\delta_3)$.

We argue that the constant on the right is non-zero. Suppose for the sake of deriving a contradiction that $\gamma_3\delta_1 - \gamma_1\delta_3 = 0$. This would imply that the polynomials $a_1a_3g_1 = a_3(a_1f_2 + a_2b_1 - a_1b_2)$ and $a_1a_3g_3 = a_1(a_3f_2 + a_2b_3 - a_3b_2)$ have the root δ_1/γ_1 in common. This would also have to be a root of their difference $a_2(a_3b_1 - a_1b_3)$, which by our assumption of non-redundancy, is a non-zero constant: a contradiction, as desired.

The equation $\gamma_3v_1^2 - \gamma_1v_3^2 = a_2(\gamma_3\delta_1 - \gamma_1\delta_3)$ thus has infinitely many solutions only if it is a generalised Pell equation. In this case, by Cor. 2.3 the values of v_1 for which all three constraints are satisfied form a union of finitely many exponentially-growing LRBS.

Case (5): Four non-redundant (positive) constraints will have finitely many solutions by virtue of containing two cubic constraints (Case (3a)) or three quadratic constraints (Case (4b)) and may be effectively enumerated. □

Proposition 4.8. *Let S be the set of solutions x to a system of constraints given by $x > c$ and non-redundant constraints $\mathcal{R}_i(a_ix + b_i, q_i, r_i)$ for $i = 1, \ldots, l$, $a_i > 0$. If S is infinite, then the subset $S' \subseteq S$ for which any non-redundant negative constraint $\neg\mathcal{R}(ax + b)$ is violated has density 0 relative to S, unless S is of the form (3b) and S is of the form (4c) in Prop. 4.7*

Proof. A negative constraint $\neg\mathcal{R}_i(cx+d)$ being violated is equivalent to the positive constraint $\mathcal{R}_i(cx + d)$ holding. We go through the cases given by Prop. 4.7. In each case, we show that a negative constraint discards either a subset of relative density 0, or that the set of parameters (of the solutions to the positive constraints) invalidated by forming finitely many arithmetic progressions. In either case, we can effectively determine whether there remain solutions after accounting for finitely many negative constraints.

Case (1): $l = 0$. Then $S = [c+1, \infty)$. The discarded set S' with the constraint $\mathcal{R}_i(ax + b)$ added becomes a finite union of sets of the form $h(\mathbb{Z}) \cap [c + 1, \infty)$ where h is a polynomial with $\deg h = d \ge 2$ and positive leading coefficient, by Prop. 4.7. Apply Lem. 3.8 with $g(y) = y$ and each h to conclude S' has density 0 relative to S.

Case (2): $l = 1$. Then the solution set S to $x > c$ and $\mathcal{R}_1(a_1x+b_1)$ is a finite union of sets of the form $g(\mathbb{Z}) \cap [c+1, \infty)$ for a polynomial g with positive leading coefficient and degree d_1. S' will satisfy the conclusion of either case (3a), (3b), or (3c). In case S' satisfies (3a) it is finite and thus has density 0 relative to S. The case of S' satisfying (3b) has been handled in Prop. 3.9 *mutatis mutandis* and S' thus has density 0 relative to S. And in case S' satisfies (3c), the discarded set S' is a finite union of sets of the form $h(\mathbb{Z}) \cap [c + 1, \infty)$ for polynomials h of degree 6, which is greater than d_1. Therefore we may again apply Lem. 3.8 with f, g to conclude S' has density 0 relative to S.

Case (3): $l = 2$. S is infinite only if $d_1 = 2, d_2 = 3$, or $d_1 = d_2 = 2$. In the former case, S' can be infinite only if $d_3 = 2$. Then by Prop. 4.7, if the discarded set S' is infinite, then S' is a union of finitely many simple reversible LRBS restricted to $[c + 1, \infty)$. As before we conclude by Lem. 3.8 that S' is a null-density subset of S. The case $d_1 = d_2 = 2, d_3 = 3$ results in finitely many solutions being discarded if a pair of constraints gives rise to an elliptic curve (see the first part of Case (3c) of Prop. 4.7); the case where it does not result in infinite LRBS of solutions being discarded. This discarded set can have positive relative density; however we have from Case (3c) of Prop. 4.7 that the solutions to the positive constraints are

themselves parametrised as LRBS. Prop. 4.10 shows that the indices of discarded solutions form arithmetic progressions, and hence we can effectively determine whether there exists a solution not discarded by the negative constraints.

Case (4) S' is finite and hence has density 0 relative to S. □

The following is a special case of [20, Prop. 2].

Lemma 4.9 (Parametrisation of x values ruled out). *Let α_n, β_n be a sequence of solutions to the generalised Pell equation $\alpha^2 - D_1 \cdot \beta^2 = N_1$, which are Pell multiples of a generating pair (α', β'). Assume further that, for a subsequence (α_k, β_k) with $k \in \mathcal{K} \subset \mathbb{N}$, we have*

$$(\alpha_k, \beta_k) = (\phi_1(\sigma_m), \phi_2(\tau_m)) \tag{14}$$

for rational polynomials ϕ_i and a sequence (σ_m, τ_m) of Pell values satisfying the equation $\sigma^2 - D_2\tau^2 = N_2$. Assume $\mathcal{K}$ is infinite and the D_i are square-free. Then

(i) $D_1 = D_2$.
(ii) $\deg \phi_1 = \deg \phi_2 =: \chi$.
(iii) *$\mathcal{K}$ is parametrisable, i.e., there exists an integer c such that $k = \chi \cdot m + c$ for all $k \in K$.*

Proof. It is clear from the polynomial relations that $\mathbb{Q}(\sqrt{D_1}) = \mathbb{Q}(\sqrt{D_2})$, which can happen only if $D_1 = D_2$.

We have

$$(\alpha_k, \beta_k) = (A_1\varepsilon^k + B_1\varepsilon^{-k}, \frac{1}{\sqrt{D}}(A_1\varepsilon_1^k + B_1\varepsilon^{-k}))$$

and

$$(\sigma_m, \tau_m) = (A_2\varepsilon^m + B_2\varepsilon^{-m}, \frac{1}{\sqrt{D}}(A_2\varepsilon^m + B_2\varepsilon^{-m}))\,,$$

where ε is the fundamental unit associated with the underlying Pell equation.

Substituting the latter equation into (14) and comparing growth rates, (i.e., for the polynomial identity to hold for infinitely many pairs of (k, m)) gives $\deg \phi_1 = \deg \phi_2$ (which we call χ). Further, for every such pair, one has $k = \chi \cdot m + c$ for a constant integer c not depending on k or m. This can be seen by dividing out by $\varepsilon^{\chi \cdot m}$ after making the substitution in (14), and taking the limit as $k, m \to \infty$. In particular, we must have $\frac{\varepsilon^k}{\varepsilon^{\chi \cdot m}}$ converges to a constant as $k, m \to \infty$. As $\varepsilon > 1$, we have that this sequence becomes eventually constant, i.e., $\varepsilon^k / \varepsilon^{\chi \cdot m} = \varepsilon^c$ for some fixed integer c. We moreover claim that this holds for all k, as one can construct the error-term Laurent polynomial in ε^m, of degree bounded independent of k or m, given by $\Lambda(\varepsilon^m) := \phi_1(\sigma_m, \tau_m) - \alpha_{\chi \cdot m + c}$. As this polynomial has infinitely many roots, it must be identically zero, giving $k = \chi \cdot m + c$ for all k.

Further, this constant integer c can be effectively determined from the constants in the equations. This enables us to parametrically rule out Pell-pairs (α_k, β_k) contributing to integral values of x. □

Proposition 4.10. *Suppose S is of the form (3b) and S' is of the form (4c) of Prop. 4.7. Then S' consists of indices of S in arithmetic progressions. These may be computed effectively.*

Proof. We borrow notation from the relevant cases. Note that S is parametrised by u_1, u_3 satisfying the generalised Pell equation $\gamma_3 u_1^2 - \gamma_1 u_3^2 = C$ for some C depending on the data, and that further $u_1^2 = (\alpha_1 u_2 + \beta_1)^2(\gamma_1 u_2 + \delta_1)$ and also $u_3^2 = (\alpha_3 u_2 + \beta_3)^2(\gamma_3 u_2 + \delta_3)$. Note that S' is parametrised by v_1, v_3 satisfying $\gamma_3 v_1^2 - \gamma_1 v_3^2 = a_2(\gamma_3\delta_1 - \gamma_1\delta_3)$ with $v_i^2 = a_2(\gamma_i u_2 + \delta_i)$ for $i = 1, 3$. hence $u_2 = \frac{v_i^2 - a_2\delta_i}{\gamma_i a_2}$ and thus $u_i^2 = (\alpha_i \frac{v_i^2 - a_2\delta_i}{\gamma_i a_2} + \beta_i)^2 \frac{v_i^2}{a_2}$ for $i = 1, 3$ so $u_i = \pm(\alpha_i \frac{v_i^2 - a_2\delta_i}{\gamma_i a_2} + \beta_i)\frac{v_i^2}{\sqrt{a_2}}$ for $i = 1, 3$. If $\sqrt{a_2}$ is irrational the claim is vacuous, otherwise we apply Lem. 4.9 four times for each choice of sign and get the desired conclusion. □

The above results justify the correctness of the following algorithm: We first compute S. Should S be of the form (3b) and any $n \geq 1$ negative constraint gives rise to a discarded set S' of the form (4c) we compute the resulting $S \setminus \bigcup_{i=1}^n S_i'$ using Prop 4.10. We set $\tilde{S} = S \setminus \bigcup_{i=1}^n S_i'$ in this case. Otherwise we set $\tilde{S} = S$. If $\tilde{S}$ is finite we enumerate $\tilde{S}$ and check all constraints. If $\tilde{S}$ is infinite we simply return true as any negative constraint will remove only a null-density subset.

5 UNDECIDABILITY

5.1 Büchi's Problem

Büchi formulated the following problem while studying the existential fragment of Presburger arithmetic expanded with the perfect-square predicate $\mathbf{Z}^2$: does there exist an M such that any integer sequence of M squares whose second difference is constant and equal to 2 is necessarily a sequence of *consecutive* squares? That is, is there an M such that for all $x_1, \ldots, x_M$ such that $x_{i+2}^2 - 2x_{i+1}^2 + x_i^2 = 2$ for $i = 1, \ldots, M-2$, we have that $x_i = x_{i+1} - 1$ for $i = 1, \ldots, M-1$? A positive answer to Büchi's problem enables one to define the squaring function from the perfect-square predicate without the need for quantifiers. Indeed, the assertion $y = x^2$ would be equivalent to $\bigwedge_{i=0}^{M-1} \mathbf{Z}^2(y + 2ix + i^2)$. Multiplication would in turn be positive-existentially defined using the identity $4xy = (x+y)^2 - (x-y)^2$. The undecidability of the existential fragment of Presburger arithmetic expanded with the perfect-square predicate $\mathbf{Z}^2$ would hence follow. Büchi himself conjectured that $M = 5$, and a proof has recently been announced by Xiao [24].

We show that the (negation of the) Büchi conjecture can be encoded in $SMT^2\langle\mathbb{Z}; +, 0, 1, \mathbf{Z}^2\rangle$. Indeed, any counterexample sequence must have the form $i^2 + c_1 i + c_0$ for $i = 1, \ldots, 5$, where c_0, c_1 are integers and the polynomial $g(t) = t^2 + c_1 t + c_0$ is not of the form $(x + b)^2$. By the contrapositive of [16, Cor. 1.7], there exists M such that $g(M)$ is not a perfect square: by appropriate shifting, we can assume $M = 6$. The negation of the Büchi conjecture is then simply the formula

$$\exists c_0, c_1 \,.\, \left(\bigwedge_{i=1}^{5} \mathbf{Z}^2(c_0 + ic_1 + i^2)\right) \wedge \neg\mathbf{Z}^2(c_0 + 6c_1 + 36)\,.$$

The following technical but elementary lemma shows how one can encode arbitrary Diophantine equations in the signature of $\langle\mathbb{Z}; 0, 1, +, -, \mathbf{Z}^2\rangle$ with a limited budget of first-order variables.

Lemma 5.1. *Let $h \in \mathbb{Z}[x_1, \ldots, x_n]$. The assertion $h(x_1, \ldots, x_n) = 0$, where $x_1, \ldots, x_n$ are integer-valued variables, can be encoded in $FO\langle\mathbb{Z}; 0, 1, +, -, \mathbf{Z}^2\rangle$ via a formula that uses at most 4 bound variables, all of which are existentially quantified.*

Proof. We need existentially quantified variables to implement multiplication using the identity $4xy = (x + y)^2 - (x - y)^2$, e.g.,

the proposition $t = 4xy$ is equivalent to $\exists u \exists v \,.\, (t = u - v) \wedge (u = (x + y)^2) \wedge (v = (x - y)^2)$, where the last conjunct is written as $\bigwedge_{i=0}^{4} \mathbf{Z}^2(v + 2i(x - y) + i^2)$, and similarly for the penultimate conjunct.

We momentarily leave aside the issue of the scarcity of existentially quantified variables and introduce rewrite rules (that replace polynomials with linear combinations of variables and simpler polynomials), with each application introducing fresh quantified variables. For convenience, we refer to the sum of monomials r through m of $h(x_1, \ldots, x_n)$ as h_r, and the r-th monomial g_{r1} of degree d_r is constructed through the intermediate monomials $g_{rs} = \frac{c_r}{4^{s-1}} \prod_{l=s}^{d_r} x_{jl}$. In the rewrite rules that follow, the subformulas are assumed to be minimal.

(1) A subformula of the form $\psi\,(h_r, \ldots)$ is rewritten as

$$\exists t \,.\, \psi\,[h_r/(t + g_{r1})] \wedge (t = h_{r+1})\,,$$

reducing the number of monomials.

(2) A subformula of the form $\psi\,(g_{rs}, \ldots)$ is rewritten as

$$\exists u, v \,.\, \psi\,[g_{rs}/(u - v)] \wedge\ u = \left(x_{j_s} + g_{r,s+1}\right)^2 \wedge\ v = \left(-x_{j_s} + g_{r,s+1}\right)^2,$$

reducing the degree of the monomial.

(3) A subformula of the form $w = T^2$ is rewritten as

$$\bigwedge_{i=0}^{4} \mathbf{Z}^2\left(w + 2iT + i^2\right),$$

where T is a linear combination of variables.

We observe that we can first repeatedly apply Rule 1 until the formula involves only monomials, then repeatedly apply Rule 2 until all propositions are either linear equations or assertions of a square relation, and finally apply Rule 3 to encode the latter in our signature.

Finally, we show that while applying the rewrite rules to the formula $h(x_1, \ldots, x_n) = 0$ and thus introducing existentially quantified variables, we can recycle these variables so that we only need 4 of them. The key observation is that if a variable t does not occur in a subformula ψ being rewritten, it can be recycled for the purpose.

We claim that we can alternate between introducing t_0 and t_1 as we repeatedly apply Rule 1. For instance, $h = 0$ gets rewritten to $\exists t_0 \,.\, t_0 + g_{11} = 0 \wedge t_0 = h_2$, which itself gets rewritten to $\exists t_0 \,.\, t_0 + g_{11} = 0 \wedge (\exists t_1 \,.\, t_0 = t_1 + g_{21} \wedge t_1 = h_3)$. We use our key observation that the previously quantified t_0 does not occur in $t_1 = h_3$, and can be recycled for this purpose of rewriting it. The intermediate formula after completing the applications of Rule 1 recycles t_0, t_1 in an alternating manner while introducing quantified variables.

We now have subformulas of the form $t_b = t_{1-b} + g_{r1}$ that we need to rewrite using Rule 2. We cannot use t_0, t_1, and hence must use t_2, t_3 to obtain

$$\exists t_2, t_3 \,.\, t_b = t_{1-b} + t_2 - t_3 \wedge t_2 = (x + g_{r2})^2 \wedge t_3 = (-x + g_{r2})^2,$$

to which Rule 2 may need to be reapplied. This time, however, we have access to t_0, t_1 while rewriting $t_2 = (x + g_{r2})^2$. In this manner, we can alternate between introducing t_0, t_1, and t_2, t_3 while applying Rule 2. This leaves us with subformulas of the form $t_a = (x + t_b - t_c)^2$ to rewrite using Rule 3. This is merely a syntactic rewrite, and we have indeed proven that we need only 4 existentially quantified variables. □

5.2 Universal Diophantine Equations and Undecidability

It is well known that Hilbert's tenth problem, i.e., deciding whether a given polynomial equation has integer solutions, is undecidable. Thanks to Xiao's proof of Büchi's conjecture [24], together with Lem. 5.1 and the bounds on the degree of the polynomial and number of variables, we now establish undecidability results for bounded-variable Presburger arithmetic expanded with the perfect-square predicate. More specifically, Jones [13] constructs *universal* Diophantine equations, i.e., polynomials h in several unknowns $x_1, \ldots, x_n$ and parameters $x, y, z, w \in \mathbb{N}$ such that

$$\exists x_1, \ldots, x_n \in \mathbb{N} \,.\, h(x, y, z, w, x_1, \ldots, x_n) = 0$$

if and only if x is contained in the recursively enumerable set indexed by $\langle y, z, w \rangle$.

Theorem 5.2. *The following theories are undecidable:*

(1) $\exists FO^{13}\langle \mathbb{Z}; 0, 1, +, -, <, \mathbf{Z}^2 \rangle$,
(2) $\exists FO^{14}\langle \mathbb{Z}; 0, 1, +, -, \mathbf{Z}^2 \rangle$,
(3) $SMT^{600}\langle \mathbb{Z}; 0, 1, +, -, <, \mathbf{Z}^2 \rangle$,
(4) $SMT^{2200}\langle \mathbb{Z}; 0, 1, +, -, \mathbf{Z}^2 \rangle$.

Proof. Item (1) follows from Matiyasevich's construction (see [13, Sec. 3]), which when given $x \in \mathbb{N}$ and a recursively enumerable set W, produces a Diophantine equation with 9 positive-integer unknowns that has a solution if and only if $x \in W$. Item (2) follows from the analogue due to Sun [21, Thm. 1.1(ii)], where the constructed equation has 9 integer unknowns and 1 nonzero-integer unknown.

Item (3) follows from Jones's concrete example of a universal Diophantine equation of degree 4 with 58 positive-integer unknowns, which can be implemented with at most 100 arithmetic operations [13, Thm. 5]. Note that each multiplication would introduce at most 5 fresh variables to be encoded in our SMT instance (as discussed in the proof of Lem. 5.1). We take $600 = 100 + 100 \cdot 5$ as a conservative estimate for the total number of variables.

Item (4) follows from converting each positive-integer unknown into a regular integer unknown by introducing fresh variables and using the Lagrange four-squares theorem, i.e., $\exists\, x > 0$ replaced by $\exists y_1, \ldots, y_4 \,.\, \bigwedge_{i=1}^{4} Sq(y_i)$, and every occurrence of x is replaced by $(y_1 + \cdots + y_4)$. A conservative upper bound on the number of variables introduced in this manner is $4 \cdot 4 \cdot 100$, and combining them with the original variables gives a sound estimate of 2200. □

REFERENCES

[1] A. Baker. Bounds for the solutions of the hyperelliptic equation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 65(2):439–444, 1969.
[2] Alan Baker. *Transcendental Number Theory.* Cambridge Mathematical Library. Cambridge University Press, 1975. 1994 Reprinted Edition.
[3] Jean Berstel and Christophe Reutenauer. *Rational series*, page 3–28. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2010.
[4] Alexis Bès. A survey of arithmetical definability. *A tribute to Maurice Boffa*, pages 1–54, 2002.
[5] Paul-Jean Cahen and Jean-Luc Chabert. What you should know about integer-valued polynomials. *The American Mathematical Monthly*, 123(4):311–337, 2016.
[6] P. Fatou. Sur les séries entières à coefficients entiers. *Comptes Rendus de l'Académie des Sciences, Paris*, 138(130):342–344, 1904.
[7] Erich Grädel, Phokion G. Kolaitis, and Moshe Y. Vardi. On the decision problem for two-variable first-order logic. *Bull. Symb. Log.*, 3(1):53–69, 1997.
[8] Erich Grädel and Martin Otto. On logics with two variables. *Theor. Comput. Sci.*, 224(1-2):73–113, 1999.

[9] Martin Grohe. Finite variable logics in descriptive complexity theory. *Bull. Symb. Log.*, 4(4):345–398, 1998.
[10] Christoph Haase. A survival guide to presburger arithmetic. *ACM SIGLOG News*, 5(3):67–82, 2018.
[11] Michael J. Jacobson Jr. and Hugh C. Williams. *Solving the Pell equation*. Springer, New York, 2009.
[12] Gareth A. Jones and Josephine M. Jones. *Elementary Number Theory*. Springer Undergraduate Mathematics Series. Springer London, 1998.
[13] James P. Jones. Universal Diophantine Equation. *The Journal of Symbolic Logic*, 47(3):549–571, 1982.
[14] Toghrul Karimov, Florian Luca, Joris Nieuwveld, Joël Ouaknine, and James Worrell. On the Decidability of Presburger Arithmetic Expanded with Powers. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2755–2778, 2025.
[15] Martin Otto. *Bounded variable logics and counting - a study in finite models*, volume 9 of *Lecture Notes in Logic*. Springer, 1997.
[16] Hector Pasten and Xavier Vidaux. Positive existential definability of multiplication from addition and the range of a polynomial. *Israel Journal of Mathematics*, 216:273–306, 10 2016.
[17] Françoise Point. On decidable extensions of Presburger arithmetic: from A. Bertrand numeration sytems to Pisot numbers. *The Journal of Symbolic Logic*, 65(3):1347–1374, 2000.
[18] Dimitrios Poulakis and Evaggelos Voskos. On the Practical Solution of Genus Zero Diophantine Equations. *Journal of Symbolic Computation*, 30(5):573–582, 2000.
[19] Ian Pratt-Hartmann. *Fragments of first-order logic*, volume 56. Oxford University Press, 2023.
[20] H. P. Schlickewei and W. M. Schmidt. Linear equations in members of recurrence sequences. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze*, 20(2):219–246, 1993.
[21] Zhi-Wei Sun. Further results on Hilbert's Tenth Problem. *Science China Mathematics*, 64(2):281–306, 2021.
[22] László Szalay. On the resolution of simultaneous Pell equations. *Annales Mathematicae et Informaticae*, 34:77–87, 2007.
[23] Nikos Tzanakis. Effective solution of two simultaneous Pell equations by the Elliptic Logarithm Method. *Acta Arithmetica*, 103, 10 2001.
[24] Stanley Yao Xiao. Hilbert's tenth problem for systems of diagonal quadratic forms, and Büchi's problem, 2025.