The Complexity of Disjunctive Linear Diophantine Constraints

Manuel Bodirsky[1]

Institut für Algebra, TU Dresden, Germany
manuel.bodirsky@tu-dresden.de

Barnaby Martin

Department of Computer Science, Durham University, U.K.
barnabymartin@gmail.com

Marcello Mamino[2]

Dipartimento di Matematica, largo Pontecorvo 5, 56127 Pisa, Italy
marcello.mamino@dm.unipi.it

Antoine Mottet[3]

Institut für Algebra, TU Dresden, Germany
antoine.mottet@tu-dresden.de

───── **Abstract** ─────

We study the Constraint Satisfaction Problem $\mathrm{CSP}(\mathbb{A})$, where $\mathbb{A}$ is first-order definable in $(\mathbb{Z}; +, 1)$ and contains $+$. We prove such problems are either in P or NP-complete.

1 Introduction

A *constraint satisfaction problem* (CSP) is a computational problem where the input consists of a finite set of variables and a finite set of constraints, and where the question is whether there exists a mapping from the variables to some fixed domain such that all the constraints are satisfied. When the domain is finite, and arbitrary constraints are permitted in the input, the CSP is NP-complete. However, when only constraints for a restricted set of relations are allowed in the input, it might be possible to solve the CSP in polynomial time. The set of relations that is allowed to formulate the constraints in the input is often called the *constraint language*. The question as to which constraint languages give rise to polynomial-time solvable CSPs has been the topic of intensive research over the past years. It was conjectured by Feder and Vardi [14] that CSPs for constraint languages over finite domains have a complexity dichotomy: they are in P or are NP-complete. This conjecture has recently been proved [12, 25].

A famous CSP over an infinite domain is the feasibility question for Integer Programs. It is of great importance in practice and theory of computing, and NP-complete. In order to obtain a systematic understanding of polynomial-time solvable restrictions and variations of this problem, Jonsson and Lööw [15] proposed to study the class of CSPs where the constraint language $\mathbb{A}$ is definable in *Presburger arithmetic*; that is, consists of relations that have a first-order definition over $(\mathbb{Z}; <, +, 1)$. Equivalently, each relation $R(x_1, \ldots, x_n)$ in $\mathbb{A}$ can be defined by a disjunction of conjunctions of the atomic formulas of the form $p \leq 0$ where p is a linear polynomial with integer coefficients and variables from $\{x_1, \ldots, x_n\}$. The constraint satisfaction problem for $\mathbb{A}$, denoted by CSP($\mathbb{A}$), is the problem of deciding whether a given conjunction of formulas of the form $R(y_1, \ldots, y_n)$, for some n-ary R from $\mathbb{A}$, is satisfiable in $\mathbb{A}$. By appropriately choosing such a constraint language $\mathbb{A}$, a great variety of problems over the integers can be formulated as CSP($\mathbb{A}$). Several constraint languages $\mathbb{A}$ over the integers are known where the CSP can be solved in polynomial time. Among the most famous of these is Linear Diophantine Equations, namely CSP($\mathbb{Z}; +, 1$). The first polynomial-time algorithms for the satisfiability of linear Diophantine equation systems have been discovered by Frumkin and, independently, Sieveking and von zur Gathen. Kannan and Bachem [17] presented a method based on first computing the Hermite Normal Form of the matrix given by the linear system (see discussion in the text-book of Schrijver [23]). Further improvements have been made in [13, 24, 19]. In the present parlance, CSP($\mathbb{Z}; <, +, 1$) is Integer Program feasibility itself. However, a complete complexity classification for the CSPs of Jonsson-Lööw languages appears to be a very ambitious goal.

Among the classes of constraint language that fall into the framework of Jonsson and Lööw are the *distance CSPs* of [5, 10] and the *temporal CSPs* of [8]. Temporal CSPs are those whose constraint language is first-order definable in $(\mathbb{Q}; <)$ and *discrete temporal CSPs* are those whose constraint languages is first-order definable in $(\mathbb{Z}; <)$. The classification for discrete temporal CSPs represents the join of the work on temporal CSPs and distance CSPs, and has only recently been accomplished [4].

Moving away from the discrete and non-dense, $(\mathbb{Q}; <)$ is not the only structure for which constraint languages that are first-order expansions have had their CSPs classified. The situation for such expansions of the language of linear programming, $(\mathbb{Q}; <, +, 1)$ was settled in [6]. Perhaps, more interesting for us is the simplified situation in which only first-order expansions of $(\mathbb{Q}; +)$ are considered, in [7]. Most recently, the work [16] delivers a classification for all first-order definitions in $(\mathbb{Q}; <, +, 1)$ that contain $+$, thus properly extending the result from [6]. In these works, the class of relations quantifier-free definable in Horn CNF plays a key role. In this context, the atomic relations are inequalities and equalities, and each clause may have no more than one equality or inequality. That is, additional disjuncts in clauses must be disequalities. For first-order expansions of $(\mathbb{Q}; +)$, the tractable constraint languages are precisely those that are quantifier-free Horn definable on $(\mathbb{Q}; +)$ [7].

However, the integers behave very differently from the rationals or reals and even simple types of Horn definitions engender intractable constraint languages, as documented in [15]. This article shows, depending on one's perspective, [un]surprisingly, that the tractability frontier for first-order definitions of $(\mathbb{Z}; +, 1)$, containing $+$, coincides with that for first-order expansions of $(\mathbb{Q}; +)$. Under a mild technical assumption on $\mathbb{A}$, either all of its relations are quantifier-free Horn definable, in the expansion of $(\mathbb{Z}; +, 1)$ associated with its quantifier elimination, and CSP($\mathbb{A}$) is solvable in P; or CSP($\mathbb{A}$) is NP-complete. From this we obtain the following dichotomy result.

▶ **Theorem 1.** *Let $\mathbb{A}$ be an expansion of $(\mathbb{Z}; +)$ by finitely many relations with a first-order definition in $(\mathbb{Z}; +, 1)$. Then* CSP($\mathbb{A}$) *is in P or NP-complete.*

Other related work

This work forms part of a growing body addressing infinite-domain CSPs. One line of that work concerns ω-categorical and finitely-bounded constraint languages and the other line considers constraint languages over ordinary structures of arithmetic. The two lines overlap in the foundational work on temporal CSPs [8]. The outstanding other result in the first line is [11] and recent progress can be seen in [3, 1]. The importance of the latter line is discussed in the survey [9].

The CSP for certain finite groups were studied already in the seminal [14]. $(\mathbb{Z}; +, 0)$ is a group *par excellence* and our work takes inspiration from that paper. One of our hardness results uses its Theorem 34 and our tractable cases include the situation when all relations are subgroups, or cosets of subgroups, of powers of $\mathbb{Z}$ (cf. [14], Theorem 33). However, not all first-order expansions of $(\mathbb{Z}; +, 1)$ are related to groups, and we have other sources of tractability too.

2   Preliminaries

We say a relational structure $\mathbb{A}$ is *first-order definable in* $(\mathbb{Z}; +, 1)$ (or a *first-order reduct of* $(\mathbb{Z}; +, 1)$) if it is over domain $\mathbb{Z}$ with relations specified by first-order formulas over $(\mathbb{Z}; +, 1)$. An *endomorphism* of $\mathbb{A}$ is a map $h \colon \mathbb{Z} \to \mathbb{Z}$ such that for every relation R of $\mathbb{A}$ and every tuple $(a_1, \ldots, a_k) \in \mathbb{Z}^k$, we have $\mathbf{a} \in R \Rightarrow h(\mathbf{a}) \in R$. We say that h is a *self-embedding* if the implication is an equivalence.

A formula over a relational signature σ is *primitive positive (pp)* if it is of the form $\exists x_1, \ldots, x_k (\psi_1 \wedge \cdots \wedge \psi_m)$ where each ψ_i is an atomic relation built from σ. Note that 0 is pp-definable in $(\mathbb{Z}; +)$. A *sentence* is a formula without free variables.

The *constraint satisfaction problem* for a structure $\mathbb{A}$ with finite relational signature σ, denoted CSP($\mathbb{A}$), is the following computational problem.

Input: A primitive positive σ-sentence Φ.
Question: $\mathbb{A} \models \Phi$?

All CSPs will be defined over strictly relational signatures, thus in this context $+$ must be considered a ternary relation and 1 a constant or singleton unary relation, depending on taste. Since we also use $+$ with its common meaning of binary operation, we concede guilt for overloading. However, the two uses will never conflict in meaning, so we will not dwell further on the matter. If $\mathbb{A}$ is first-order definable in $(\mathbb{Z}; <, +, 1)$ then CSP($\mathbb{A}$) is in NP (this is noted e.g. in [15]).

A *linear equation* is a formula of the form $\sum_{i=1}^{n} a_i x_i = b$ with $a_1, \ldots, a_n, b \in \mathbb{Z}$, whose free variables are $\{x_1, \ldots, x_n\}$. A *modular linear equation* is a formula of the form $\sum_{i=1}^{n} a_i x_i = b \bmod c$ with $a_1, \ldots, a_n, b, c \in \mathbb{Z}$. Let $\mathcal{L}_{(\mathbb{Z}; +, 1)}$ be the infinite relational language containing a relation symbol for each linear equation and modular linear equation. For convenience, we consider first-order logic to have native symbols for $\top$ (true) and $\bot$ (false). It is well-known that $(\mathbb{Z}; +, 1)$ admits quantifier elimination in the language $\mathcal{L}_{(\mathbb{Z}; +, 1)}$ (see [21], or [18, Corollary 3.1.21] for a more modern treatment). Call an $\mathcal{L}_{(\mathbb{Z}; +, 1)}$-formula *standard* if it does not contain a negated modular linear equation. Every $\mathcal{L}_{(\mathbb{Z}; +, 1)}$-formula is equivalent to a standard $\mathcal{L}_{(\mathbb{Z}; +, 1)}$-formula, since a negated modular linear equation is equivalent to a disjunction of modular linear equations (i.e., $k \neq b \bmod c \Leftrightarrow \bigvee_{0 \leq a \leq c, a \neq b} k = a \bmod c$). We say that an equation *appears* in a formula if it is a positive or negative literal in that formula.

Any subgroup G of $\mathbb{Z}^k$ can be given by a finite set of *generators*, i.e., k-tuples $\mathbf{g}^1, \ldots, \mathbf{g}^m$, such that for every $\mathbf{g} \in G$, there are $\lambda_1, \ldots, \lambda_m \in \mathbb{Z}$ such that $\mathbf{g} = \sum_i \lambda_i \mathbf{g}^i$, where we

write $\lambda \cdot \mathbf{g}$ for $(\lambda g_1, \ldots, \lambda g_k)$. A *coset* of a subgroup G of $\mathbb{Z}^k$ is any set of the form $\mathbf{a} + G := \{\mathbf{a} + \mathbf{g} \mid \mathbf{g} \in G\}$, where $\mathbf{a} \in \mathbb{Z}^k$. By moving to a standard formula, we are in a position to deduce the following.

▶ **Proposition 2.** *Suppose R is a unary relation first-order definable in $(\mathbb{Z}; +, 1)$. Then R has the form $(R^\circ \cup R^+) \setminus R^-$, where R° is a finite union of cosets of nontrivial subgroups of $\mathbb{Z}$, and R^+ and R^- are finite disjoint sets of integers.*

Proof. Consider a disjunction φ of equations (possibly negated and modular equations). If this disjunction contains a negated equation $ax \neq c$, then φ defines a relation that contains $\mathbb{Z} \setminus \{c/a\}$ and is therefore as in the statement. Otherwise, φ contains only positive linear equation and modular equations, and the relation that φ defines is clearly of the form $R^\circ \cup R^+$ for some finite set R^+ and some union R° of nontrivial subgroups of $\mathbb{Z}$.

Consider a quantifier-free formula φ in conjunctive normal form defining R. Each conjunct defines a relation of the right form, per the previous paragraph. It is easily checked that a conjunction of relations of this form is again a relation of the form $(R^\circ \cup R^+) \setminus R^-$, so that we have proved that every quantifier-free formula with one free variables defines a relation of the right form. The proposition then follows from quantifier-elimination. ◀

Note that if $R^+ \cap R^\circ = \emptyset$ and $R^- \subset R^\circ$, then R^+, R^-, and R° are unique. We use the terminology with this convention for all unary relations R that are first-order definable in $(\mathbb{Z}; +, 1)$ throughout the article.

▶ **Definition 3.** Let φ be an $\mathcal{L}_{(\mathbb{Z};+,1)}$-formula. We say that φ is *Horn* if it is a conjunction of clauses of the form

$$\bigvee_{i=1}^{n} \neg\varphi_i \vee \varphi_0$$

where $\varphi_1, \ldots, \varphi_n$ are linear equations and φ_0 is a linear or a modular linear equation.

▶ **Example 4.** Singletons, cofinite unary relations, and cosets of subgroups of $\mathbb{Z}^n$ are examples of Horn-definable relations.

3 Cores

If $\mathbb{A}$ is a first-order expansion of $(\mathbb{Z}; +)$, note that its endomorphisms are precisely of the form $x \mapsto \lambda x$ for some $\lambda \in \mathbb{Z}$. Therefore, we view in the following $\mathrm{End}(\mathbb{A})$ as a subset of $\mathbb{Z}$, where the monoid structure on $\mathrm{End}(\mathbb{A})$ implies that as a subset of $\mathbb{Z}$, it is closed under multiplication and contains 1. We say that $\mathbb{A}$ is a core if all its endomorphisms are self-embeddings, and that $\mathbb{B}$ is a core of $\mathbb{A}$ if $\mathbb{A}$ and $\mathbb{B}$ are homomorphically equivalent and $\mathbb{B}$ is a core.

▶ **Lemma 5.** *Let $\mathbb{A}$ be first-order definable in $(\mathbb{Z}; +, 1)$, and suppose that $\mathbb{A}$ contains $+$. There exists a structure which is a core of $\mathbb{A}$, and which is either a 1-element structure or first-order definable in $(\mathbb{Z}; +, 1)$ and containing $+$.*

Proof. If $0 \in \mathrm{End}(\mathbb{A})$ then the lemma is clearly true ($\mathbb{A}$ being homomorphically equivalent to the substructure of $\mathbb{A}$ induced by $\{0\}$), so let us assume that $0 \notin \mathrm{End}(\mathbb{A})$. Similarly we can assume that $\mathrm{End}(\mathbb{A}) \nsubseteq \{-1, 1\}$, otherwise $\mathbb{A}$ is already a core. For a quantifier-free formula ψ and an integer λ, define ψ/λ by induction on ψ as follows:
- if ψ is $\sum \lambda_i x_i = c$ and λ divides c, then ψ/λ is $\sum \lambda_i x_i = c/\lambda$,
- if ψ is $\sum \lambda_i x_i = c$ and λ does not divide c, then ψ/λ is $\bot$,

- if ψ is $\sum \lambda_i x_i = c \bmod d$ and $\ell := \gcd(\lambda, d)$ divides c, then ψ/λ is $\sum \lambda_i x_i = ec/\ell \bmod d/\ell$ where e is the inverse of λ/ℓ modulo d/ℓ,
- if ψ is $\sum \lambda_i x_i = c \bmod d$ and $\ell := \gcd(\lambda, d)$ does not divide c, then ψ/λ is $\perp$,
- extend to boolean combinations in the obvious fashion.

Note that for every tuple $\mathbf{a}$, we have that $\mathbf{a}$ satisfies ψ/λ iff $\lambda \cdot \mathbf{a}$ satisfies ψ. Indeed, if ψ is a linear equation then this is clear. Similarly, it is clear if ψ is a modular equation and $\ell := \gcd(\lambda, d)$ does not divide c. Suppose that ψ is a modular equation and $\ell := \gcd(\lambda, d)$ divides c. If $\sum \lambda \lambda_i x_i = c \bmod d$ then $\lambda/\ell \cdot (\sum \lambda_i x_i) = qd/\ell + c/\ell$ so that $e\lambda/\ell \cdot (\sum \lambda_i x_i) = (eq) \cdot d/\ell + ec/\ell$, where e is the inverse of λ/ℓ modulo d/ℓ and $q \in \mathbb{Z}$. We therefore obtain $\sum \lambda_i x_i = ec/\ell \bmod d/\ell$. Conversely if $\sum \lambda_i x_i = ec/\ell \bmod d/\ell$ then $\sum \lambda \lambda_i x_i = (\lambda e)c/\ell + (\frac{\lambda}{\ell}q)d = c \bmod d$.

Let ψ be any quantifier-free $\mathcal{L}_{(\mathbb{Z};+,1)}$-formula and suppose that $|\lambda| > 1$. The only cases where some magnitudes of the integers on the right-hand sides of terms in the formula ψ do not decrease by forming ψ/λ is when ψ only contains literals either of the form $\sum \lambda_i x_i = 0$ or of the form $\sum \lambda_i x_i = c \bmod d$ with λ and d coprime. Therefore, the sequence $\psi_0, \psi_1, \psi_2, \dots$ where ψ_0 is ψ and where ψ_{i+1} is ψ_i/λ for some $\lambda \in \mathrm{End}(\mathbb{A})$ with $|\lambda| > 1$ reaches in a finite number of steps a fixpoint where all the literals are either of the form $\sum \lambda_i x_i = 0$ or are modular equations whose modulus d is such that λ and d are coprime. Let $n \geq 1$ be such that for every ψ defining a relation of $\mathbb{A}$, the formula ψ_n is a fixpoint. Let $\mathbb{B}$ be the structure whose domain is $\mathbb{Z}$ and whose relations are $+$ and the relations defined by ψ_n for each ψ defining a relation of $\mathbb{A}$.

We claim that $\mathbb{B}$ is homomorphically equivalent to $\mathbb{A}$ and is a core. The first claim is clear, since $\mathbb{B}$ is isomorphic to the structure obtained from $\mathbb{A}$ by successive applications of endomorphisms $x \mapsto \lambda \cdot x$ (in particular $\mathbb{B}$ embeds into $\mathbb{A}$). Let now $x \mapsto \lambda \cdot x$ be an endomorphism of $\mathbb{B}$, and suppose that $\mathbf{a}$ is a tuple in a relation R of $\mathbb{B}$. Then we have that $\lambda \cdot \mathbf{a}$ in R since $x \mapsto \lambda \cdot x$ is an endomorphism. Conversely, note that λ is coprime to d or else we would not have reached a fixed point in the previous stage. Thus, $\lambda^{\varphi(d)} = 1 \bmod d$, where $\varphi(d)$ here is the totient of d. It follows then that $\lambda^{\varphi(d)}\mathbf{a} = \mathbf{a} \bmod d$. Suppose $\lambda\mathbf{a} \in R$, then by applying $\varphi(d) - 1$ times an endomorphism, we derive $\lambda^{\varphi(d)}\mathbf{a} \in R$. It follows that $\mathbf{a} \in R$, for both the cases that atoms are of the form $\sum \lambda_i x_i = 0$ or are modular equations whose modulus d is such that λ and d are coprime. Hence, $x \mapsto \lambda \cdot x$ is an embedding of $\mathbb{A}$.   ◄

We order the standard formulas lexicographically with respect to (in this order)

1. the number of non-Horn clauses,

2. the number of literals in clauses with at least two literals,

3. the number of all literals, and

4. the sum of the absolute values of all numbers appearing in an equation.

This order is used in a number of statements and proofs throughout the text, e.g., in Proposition 6, Lemma 17, and Theorem 18. A standard formula is *minimal* if no smaller formula is equivalent to it.

The following properties follow from the construction of cores in the previous proof.

▶ **Proposition 6.** *Let $\mathbb{A}$ be first-order definable in $(\mathbb{Z};+,1)$, and suppose that $\mathbb{A}$ contains $+$ and is a core. Let $\lambda \in \mathrm{End}(\mathbb{A})$. Let R be a relation of $\mathbb{A}$ and let φ be a minimal standard formula defining R.*

- *If $\sum \lambda_i x_i = c$ is a linear equation appearing in φ, then $c = 0$ or $|\lambda| = 1$.*
- *If $\sum \lambda_i x_i = c \bmod d$ is a modular linear equation in φ, then λ and d are coprime.*

Moreover, if $\mathrm{End}(\mathbb{A}) = 1 + d\mathbb{Z}$ for some $d \geq 2$, then every relation of $\mathbb{A}$ can be expressed with a minimal formula in which all modular linear equations are modulo a divisor of d.

Proof. The two items are clear from the proof of Lemma 5. For the last statement, let d' be a modulus appearing in a minimal definition of a relation of $\mathbb{A}$. By the second item, we have that d' and $1 + kd$ are coprime, for all $k \in \mathbb{Z}$. Let ℓ be such that $\ell d = -1 \bmod \frac{d'}{\gcd(d,d')}$. If d' and $1 + \ell d$ are coprime, there exist $u, v \in \mathbb{Z}$ such that $ud' + v(1 + \ell d) = 1$. Taking this equation modulo $\frac{d'}{\gcd(d,d')}$ we obtain $0 = 1 \bmod \frac{d'}{\gcd(d,d')}$, so that $\gcd(d, d') = d'$ and d' divides d.   ◀

4   Hardness

Our sources of hardness come from *pp-interpretations*, that we define now. A structure $\mathbb{B}$ is said to be *one-dimensional pp-interpretable* in $\mathbb{A}$ if there exists a partial surjective map $h \colon A \to B$, called the *coordinate map*, such that the inverse image of every relation of $\mathbb{B}$ (including the equality relation and the unary relation B) under h has a pp-definition in A. Formally, we require that for every k-ary relation R of $\mathbb{B}$, there exists a pp-formula $\varphi_R(x_1, \ldots, x_k)$ in the language of $\mathbb{A}$ such that

$$\mathbb{A} \models \varphi_R(a_1, \ldots, a_k) \Leftrightarrow \mathbb{B} \models R(h(a_1), \ldots, h(a_k))$$

holds for all $a_1, \ldots, a_k \in A$. This requirement for the equality relation of $\mathbb{B}$ and the unary relation B implies that the kernel of h and its domain have a pp-definition in $\mathbb{A}$. It is well-known that if $\mathbb{B}$ is pp-interpretable in $\mathbb{A}$, then $\mathrm{CSP}(\mathbb{B})$ reduces in polynomial time to $\mathrm{CSP}(\mathbb{A})$.

4.1   The fully modular case

One of the sources of hardness for our problems are expansions of the *general subgroup problem* from [14]. The general subgroup problem of a finite abelian group G is the CSP of $(G; +)$ expanded with a k-ary relation for every coset $\mathbf{a} + H$, where H is a subgroup of G^k. It is known that this problem is solvable in polynomial time (under some reasonable encoding of the input); in modern parlance, this follows from the fact that the operation $(x, y, z) \mapsto x - y + z$ is a Maltsev polymorphism of the template. Feder and Vardi [14, Theorem 34] proved that the problem becomes NP-hard if the template is further expanded by any other relation.

The general subgroup problem of $\mathbb{Z}/d\mathbb{Z}$ can be viewed as a CSP of a first-order reduct of $(\mathbb{Z}; +, 1)$ whose relations are defined by quantifier-free formulas only containing modular linear equations. This motivates the following definition.

▶ **Definition 7.** A relation $R \subseteq \mathbb{Z}^k$ is called *fully modular* if it is definable by a conjunction of disjunctions of modular linear equations, in which case we can even assume that all the modular linear equations involved in such a definition of R have the same modulus $d \geq 1$.

▶ **Proposition 8.** *Let $\mathbb{A}$ be a finite-signature core which is first-order definable in $(\mathbb{Z}; +, 1)$ and contains $+$. Suppose that $\mathbb{A}$ has a fully modular relation that is not Horn-definable. Then* $\mathrm{CSP}(\mathbb{A})$ *is NP-complete.*

Proof. Let R be a relation of $\mathbb{A}$ that is not Horn-definable and fully modular, and let $d \geq 1$ be such that R can be defined with only linear equalities modulo d. Let $\mathbb{A}/d\mathbb{A}$ be the structure with domain $\mathbb{Z}/d\mathbb{Z}$ containing the ternary relation $+$ as well as a relation S' for every relation S of arity k of $\mathbb{A}$, defined by

$$S' = \{(a_1, \ldots, a_k) \mid \exists q \in \mathbb{Z} : (qd + a_1, \ldots, qd + a_k) \in S\}.$$

Note that $\mathbb{A}/d\mathbb{A}$ is pp-interpretable in $\mathbb{A}$: the coordinate map is the canonical projection $x \mapsto x \bmod d$, whose kernel is pp-definable by the formula $\varphi_=(x, y) := \exists z(x - y = dz)$. As a consequence, $\mathrm{CSP}(\mathbb{A}/d\mathbb{A})$ reduces in logarithmic space to $\mathrm{CSP}(\mathbb{A})$. Moreover, if $\mathbb{A}$ is a core then $\mathbb{A}/d\mathbb{A}$ is also a core. It follows from general principles [2, Proposition 3.3] that $\mathrm{CSP}(\mathbb{A}/d\mathbb{A}, 1)$ reduces to $\mathrm{CSP}(\mathbb{A}/d\mathbb{A})$ and so to $\mathrm{CSP}(\mathbb{A})$. Note that every coset of a subgroup of $(\mathbb{Z}/d\mathbb{Z})^k$ is pp-definable in $(\mathbb{A}/d\mathbb{A}, 1)$ and that if R is not Horn-definable then R' is not a coset of a subgroup. It follows from Theorem 34 in the bible [14] that $\mathrm{CSP}(\mathbb{A})$ is NP-complete.   ◀

4.2   The unary case

In order to prove Theorem 1, we now focus on the case of *parametrised unary relations*.

▶ **Definition 9** (Compatibility). Let $\Lambda \subseteq \mathbb{Z} \setminus \{0\}$ be a set containing 1. We say that a set $\{S_\lambda\}_{\lambda \in \Lambda}$ of subsets of $\mathbb{Z}$ that are definable in $(\mathbb{Z}; +, 1)$ is *compatible* if there exist disjoint finite sets $A, B \subseteq \mathbb{Z}$ such that

- $S_\lambda = (S_\lambda^\circ \cup \lambda \cdot A) \setminus \lambda \cdot B$ for all $\lambda \in \Lambda$ and
- for all $d \geq 1$ and $c \in \{0, \ldots, d-1\}$, we have $c + d\mathbb{Z} \subseteq S_1^\circ \Leftrightarrow \lambda c + d\mathbb{Z} \subseteq S_\lambda^\circ$.

▶ **Definition 10** (Uniform pp-definability). Let $\mathbb{A}$ be a first-order reduct of $(\mathbb{Z}; +, 1)$. We say that $\{S_\lambda\}_{\lambda \in \Lambda}$ is *uniformly pp-definable in* $\mathbb{A}$ if there exists a pp-formula $\theta(x, y)$ such that $a \in S_\lambda$ if, and only if, $\mathbb{A} \models \theta(\lambda, a)$.

Note that the definition of being uniformly pp-definable implies that Λ has a pp-definition in $\mathbb{A}$, for $\exists y. \theta(x, y)$ is a pp-definition. Let $S \subseteq \mathbb{Z}^2$ be a binary relation that is pp-definable in $\mathbb{A}$. Then the family $\{S_\lambda\}_{\lambda \in \Lambda}$ where $\Lambda := \{a \in \mathbb{Z} \mid (a, b) \in S \text{ for some } b \in \mathbb{Z}\} \subseteq \mathbb{Z} \setminus \{0\}$ and $S_\lambda := \{a \in \mathbb{Z} \mid (\lambda, a) \in S\}$ is uniformly pp-definable in $\mathbb{A}$. But even if S contains a tuple of the form $(1, b)$ and no tuple of the form $(0, b)$, it might not necessarily satisfy the compatibility condition, as illustrated in the following example.

▶ **Example 11.** Let $S = \{(a, b) \in \mathbb{Z}^2 \mid a \neq 0 \wedge (a = b \vee a = 2b)\}$. Then $\Lambda = \mathbb{Z} \setminus \{0\}$, and for $\lambda \in \Lambda$ we have $S_\lambda = \{\lambda\}$ if $\lambda = 1 \bmod 2$ and $S_\lambda = \{\lambda, \frac{\lambda}{2}\}$ if $\lambda = 0 \bmod 2$. Therefore, the compatibility condition is not satisfied by $\{S_\lambda\}_{\lambda \in \Lambda}$.

In the following proof, we write 1-in-3-SAT for $\mathrm{CSP}(\{0, 1\}; \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\})$. It is well-known that this problem is NP-complete ([22]; for a proof see [20]).

▶ **Lemma 12.** *Let $\mathbb{A}$ be a finite-signature first-order reduct of $(\mathbb{Z}; +, 1)$ containing $+$. If $\{S_\lambda\}_{\lambda \in \Lambda}$ is a compatible set of unary relations that is uniformly pp-definable in $\mathbb{A}$ and if $1 < |S_\lambda| < \infty$ for all $\lambda \in \Lambda$, then $\mathrm{CSP}(\mathbb{A})$ is NP-hard.*

Proof. Since every S_λ is finite, one sees that $S_\lambda = \lambda \cdot A$ for the finite set A coming from the compatibility condition. Let $m_1 := \min(A)$ and $m_2 := \min(A \setminus \{m_1\})$. The formula

$$\exists \lambda (x + y + z = (m_2 - m_1)\lambda \wedge x + m_1 \lambda \in S_\lambda \wedge y + m_1 \lambda \in S_\lambda \wedge z + m_1 \lambda \in S_\lambda \wedge \lambda \in \Lambda)$$

defines the ternary relation consisting of $(a, b, c) \in \mathbb{Z}^3$ such that $a, b, c \in \{0, m_2 - m_1\}$ and exactly one of a, b, c is equal to $m_2 - m_1$. Note that this formula is in the language of $\mathbb{A}$, since $\{S_\lambda\}_{\lambda \in \Lambda}$ is uniformly pp-definable and in particular Λ is pp-definable in $\mathbb{A}$. This gives an interpretation of 1-in-3-SAT in $\mathbb{A}$, using the map $h: \{0, m_2 - m_1\} \to \{0, 1\}$ such that $h(0) = 0$ and $h(m_2 - m_1) = 1$. Therefore, $\mathrm{CSP}(\mathbb{A})$ is NP-hard.   ◀

▶ **Proposition 13.** *Let $\mathbb{A}$ be a finite-signature first-order reduct of $(\mathbb{Z}; +, 1)$ that contains $+$ and is a core. Let $\{S_\lambda\}_{\lambda \in \Lambda}$ be a compatible family that is uniformly pp-definable in $\mathbb{A}$ such that for every $\lambda \in \Lambda$ the set S_λ is not Horn-definable. Then $\mathrm{CSP}(\mathbb{A})$ is NP-hard.*

Proof. Let $A, B \subset \mathbb{Z}$ be finite such that $S_\lambda = (S_\lambda^\circ \cup \lambda \cdot A) \setminus (\lambda \cdot B)$ for all $\lambda \in \Lambda$. Since S_λ is not Horn-definable, we have $|S_\lambda| > 1$ for all $\lambda \in \Lambda$. If S_λ is finite for every $\lambda \in \Lambda$, then $\mathrm{CSP}(\mathbb{A})$ is NP-hard by Lemma 12. Therefore, we can assume that $S_\lambda^\circ \neq \emptyset$ for some $\lambda \in \Lambda$, and the second compatibility condition implies that S_λ° is infinite for all $\lambda \in \Lambda$. Let $d \geq 1$ be such that S_λ° is a union of cosets of $d\mathbb{Z}$ for all $\lambda \in \Lambda$. Write $S_1^\circ = \bigcup_{i=1}^n c_i + d\mathbb{Z}$, with $c_i \in \{0, \ldots, d-1\}$.

If $n \in \{2, \ldots, d-1\}$, we claim that we can pp-define a fully modular relation that is not Horn-definable. Indeed, let $\theta(x, y)$ be a formula that defines $\{S_\lambda\}_{\lambda \in \Lambda}$. Note that

$$\chi(x, y) := \theta(x, y) \wedge \theta(x, y + dx) \wedge \cdots \wedge \theta(x, y + \max(A \cup B)dx)$$

holds precisely on the pairs (λ, a) such that $a \in S_\lambda^\circ$: since x is forced to be in Λ by θ, a satisfying assignment gives a nonzero value λ to x. Thus, if all of $y, y + d\lambda, \ldots, y + \max(A \cup B)d\lambda$ are in S_λ, then they all must be in the modular part S_λ°. The relation T that χ defines is fully modular and is such that $T_\lambda = S_\lambda^\circ$ and in particular T is not Horn-definable. It follows from Proposition 8 that $\mathrm{CSP}(\mathbb{A})$ is NP-hard.

Otherwise, the set S_λ° consists of a single coset of $d\mathbb{Z}$ for all $\lambda \in \Lambda$, and this coset is $\lambda c_1 + d\mathbb{Z}$ by the compatibility condition on $\{S_\lambda\}_{\lambda \in \Lambda}$. Since S_λ° is assumed to not be Horn-definable, A must contain an element a. We claim that we can define another family of unary relations where the unary relations are finite and not singletons. Indeed, consider the formula

$$\psi(x, y) := \exists z \, (\theta(x, y) \wedge \theta(x, z) \wedge y + z = (c_1 + a)x)$$

and let $T \subseteq \mathbb{Z}^2$ be the relation that it defines. First, note that $\psi(\lambda, c_1)$ and $\psi(\lambda, a)$ hold for all $\lambda \in \Lambda$, so that $|T_\lambda| > 1$. We claim that T_λ is finite. Since $A \cap S_1^\circ = \emptyset$, one has $a \neq c_1 \bmod d$. Consequently, $c_1 + a \neq 2c_1 \bmod d$ and $(c_1 + a)\lambda \neq 2c_1\lambda \bmod d$. The equation $y + z = (c_1 + a)\lambda$ therefore forces that one of y and z is in $\lambda \cdot A$. Since A is finite, there are only finitely many pairs satisfying this condition, thus showing that $1 < |T_\lambda| < \infty$. It follows from Lemma 12 that $\mathrm{CSP}(\mathbb{A})$ is NP-hard. ◀

As a corollary we obtain a simple-to-state condition implying that $\mathrm{CSP}(\mathbb{A})$ is NP-hard (Corollary 16). The corollary relies on the fact that $\mathrm{End}(\mathbb{A})$, being identified with a subset of $\mathbb{Z}$, can be pp-defined in $\mathbb{A}$. We prove this in the next lemma.

▶ **Lemma 14.** *Let $\mathbb{A}$ be a finite-signature first-order reduct of $(\mathbb{Z}; +, 1)$ that contains $+$. Then the set $\mathrm{End}(\mathbb{A})$ has a pp-definition in $\mathbb{A}$ that is additionally quantifier-free.*

Proof. Let E be the set of all the formulas $R(a_1 \cdot x, \ldots, a_r \cdot x)$ for R in the language of $\mathbb{A}$ and $(a_1, \ldots, a_r) \in R$. We then have that $\mathbb{A} \models E(\lambda)$ iff $\lambda \in \mathrm{End}(\mathbb{A})$. We now show that there exists a finite subset $F \subseteq E$ that defines the same set of integers.

For each relation R of $\mathbb{A}$, fix a standard definition φ_R in conjunctive normal form of R in $(\mathbb{Z}; +, 1)$. Let M be the largest absolute value of a constant appearing in φ_R. Consider the finite family $\mathcal{F}$ of equations $\sum \mu_i x_i = m$, where $\sum \mu_i x_i = m'$ is some equation appearing in φ_R and $|m| \leq M$, together with all the equations $\sum \mu_i x_i = c \bmod d$ where $\sum \mu_i x_i = c' \bmod d$ is a modular equation appearing in φ_R and $c \in \{0, \ldots, d-1\}$. For each subset of $\mathcal{F}$ that is satisfiable by a tuple in R, pick a tuple $\mathbf{b} \in R$ satisfying the formulas in this subset and add this tuple to a set $\mathcal{S}$. Repeat this operation for every relation of $\mathbb{A}$, and let $\mathcal{S}$ be the finite set of tuples (of possibly different arities) that we obtain. Finally, let F be the subset of E where only the formulas associated with tuples from $\mathcal{S}$ are kept.

We claim that F defines $\text{End}(\mathbb{A})$. Since $F \subseteq E$, it suffices to show that every λ satisfying F is an endomorphism of $\mathbb{A}$. Let $\lambda \in \mathbb{Z}$ satisfy F, and let $\mathbf{a} \in R$ be a tuple in some relation of $\mathbb{A}$. Let $\mathbf{b} \in \mathcal{S}$ be such that $\mathbf{b}$ satisfies exactly the same equations in $\mathcal{F}$ as $\mathbf{a}$. By construction, $\lambda\mathbf{b} \in R$ so that in each clause of φ_R, some equation is satisfied by $\lambda\mathbf{b}$. We show that $\lambda\mathbf{a}$ satisfies the same equations, so that $\lambda\mathbf{a} \in R$. If $\lambda = 0$, then $\lambda\mathbf{b} = \lambda\mathbf{a}$ so that $\lambda\mathbf{a} \in R$. Suppose now that $\lambda \neq 0$. Let $\sum \mu_i x_i = c$ be a linear equation that is satisfied by $\lambda\mathbf{b}$. Then necessarily λ divides c, so that $\mathbf{b}$ satisfies $\sum \mu_i x_i = \frac{c}{\lambda}$ and $|\frac{c}{\lambda}| \leq |c| \leq M$, so that $\sum \mu_i x_i = \frac{c}{\lambda}$ is an equation in $\mathcal{F}$. Consequently, $\mathbf{a}$ also satisfies this equation and $\lambda\mathbf{a}$ satisfies $\sum \mu_i x_i = c$. The proof for modular linear equations is similar. This proves that λ is an endomorphism of $\mathbb{A}$ and concludes the proof. ◀

▶ **Lemma 15.** *Let $R \subseteq \mathbb{Z}$ be first-order definable over $(\mathbb{Z}; +, 1)$ such that $(\mathbb{Z}; +, R)$ is a core.*
- *If $R^+ \neq \emptyset$, then $\{1\}$ or $\{1, -1\}$ is pp-definable in $(\mathbb{Z}; +, R)$.*
- *If $R^+ = \emptyset$, then $R^- = \emptyset$ or $R = \mathbb{Z} \setminus \{0\}$.*

Proof. Let n be such that R° is a union of n cosets of $d\mathbb{Z}$, i.e.,

$$R^\circ = \bigcup_{i=1}^{n} c_i + d\mathbb{Z}.$$

Let us prove the first item. By Lemma 14, it suffices to prove that the only possible endomorphisms of the structure $(\mathbb{Z}; +, R)$ are $x \mapsto \lambda \cdot x$ with $\lambda \in \{1, -1\}$. Suppose that $x \mapsto \lambda \cdot x$ is an endomorphism. Then $\lambda \neq 0$ since the structure is a core, so suppose that $|\lambda| > 1$. Let a be the maximal element of R^+, and note that in particular $a + d \notin R$ (it cannot be in R^+ because of the maximality assumption, and cannot be equal to any c_i modulo d). Then $a \in R$, so $\lambda^q a \in R$ for all $q \in \mathbb{N}$. In particular, if q is such that $\lambda^q > \max(R^+ \cup R^-)$ we obtain $\lambda^q a \in R^\circ$. This means that $\lambda^q a = c_i \bmod d$ for some $i \in \{1, \ldots, n\}$. Finally, $\lambda^q(a + d) = \lambda^q a + \lambda^q d = c_i \bmod d$, so that $\lambda^q(a + d) \in R$. This implies that $x \mapsto \lambda^q \cdot x$ is not an embedding, contradicting the core assumption on $(\mathbb{Z}; +, R)$.

Let us now prove the second item. Let b be some element of R^-. We must have $b = c_i \bmod d$ for some $i \in \{1, \ldots, n\}$ since $R^- \subset R^\circ$. Note that the map $x \mapsto (d + 1)x$ is an endomorphism of $(\mathbb{Z}; +, R)$, so it has to be an embedding. It follows that $(d + 1)^m \cdot b \notin R$ for any m. Suppose that b is not 0. Choose m so that $(d + 1)^m \cdot |b| > \max_{e \in R^-} |e|$ so that $(d + 1)^m \cdot b \notin R^-$. But $(d + 1)^m b = c_i \bmod d$, a contradiction. It follows that $R^- \subseteq \{0\}$, which concludes the proof. ◀

▶ **Corollary 16.** *Let $\mathbb{A}$ be a finite-signature first-order reduct of $(\mathbb{Z}; +, 1)$ which contains $+$ and is a core. If $\text{End}(\mathbb{A})$ is not Horn-definable, then $\text{CSP}(\mathbb{A})$ is NP-hard. Moreover, if $\text{End}(\mathbb{A})$ is Horn-definable, then it is either $\{1\}$, $\mathbb{Z} \setminus \{0\}$, or $1 + d\mathbb{Z}$ for some $d \geq 2$.*

Proof. Lemma 14 implies that $R := \text{End}(\mathbb{A})$ has a quantifier-free pp-definition in $\mathbb{A}$. We first prove that $(\mathbb{Z}; +, R)$ is a core. Indeed, let λ be an endomorphism of $(\mathbb{Z}; +, R)$. Since $1 \in R$, we obtain that $\lambda \in R$, so that $x \mapsto \lambda x$ is a self-embedding of $\mathbb{A}$ by the fact that $\mathbb{A}$ is a core. Since R is has a quantifier-free definition over $\mathbb{A}$, it follows that $x \mapsto \lambda x$ is also a self-embedding of $(\mathbb{Z}; +, R)$.

First consider the case that R is not Horn-definable. If $R^+ \neq \emptyset$ then Lemma 15 implies that $\{1\}$ or $\{1, -1\}$ are pp-definable in $(\mathbb{Z}; +, R)$. All endomorphisms of $\mathbb{A}$ must preserve this set, so $\text{End}(\mathbb{A}) = R = \{1, -1\}$ or $\text{End}(\mathbb{A}) = R = \{1\}$; since R is no Horn-definable, we must even have $R = \{1, -1\}$. But then the family $\{S_\lambda\}_{\lambda \in \{-1,1\}}$ with $S_{-1} = S_1 = \{1, -1\}$ is uniformly definable and compatible, the conditions being satisfied for $A = \{-1, 1\}$ and $B = \emptyset$:

- $S_1 = A = S_{-1}$;
- $S_1^\circ = S_{-1}^\circ = \emptyset$.

Then Proposition 13 applied to S_λ implies that CSP($\mathbb{A}$) is NP-hard.

If $R^+ = \emptyset$ then Lemma 15 implies that $R^- = \emptyset$ or that $R = \mathbb{Z} \setminus \{0\}$. In the latter case, R would be Horn, contrary to the assumptions, so $R^- = \emptyset$. In this case, R is fully modular, but not Horn definable, so NP-hardness of CSP($\mathbb{A}$) follows from Proposition 8. This shows the first part of the statement.

Finally, consider the case that R is Horn-definable. If $R^+ = \emptyset$ then Lemma 15 implies that $R = \mathbb{Z} \setminus \{0\}$, and we are done, or $R^- = \emptyset$, in which case $R = 1 + d\mathbb{Z}$ for some $d \geq 2$ and we are also done. Otherwise, $R^+ \neq \emptyset$ and Lemma 15 implies that $\{1\}$ or $\{1, -1\}$ is pp-definable in $(\mathbb{Z}; +, R)$. ◄

4.3 Arbitrary arities

We finally present the hardness proof in the general case where the structure contains a relation that is not Horn-definable. The strategy is to cut from a non-Horn relation R a uniformly definable family $\{S_\lambda\}_{\lambda \in \Lambda}$ of lines for which each S_λ is not Horn-definable. In a second step, we ensure that we get a family satisfying the compatibility condition, and we conclude using Proposition 13. Call a formula φ in conjunctive normal form *reduced* if removing any literal or clause from φ yields a formula that is not equivalent to φ. Note that minimal formulas are necessarily reduced.

▶ **Lemma 17.** *Let $\mathbb{A}$ be a finite-signature first-order reduct of $(\mathbb{Z}; +, 1)$ which contains $+$ and is a core. Suppose that $\mathbb{A}$ contains a relation R that is not Horn-definable. Then* CSP($\mathbb{A}$) *is NP-hard, or $\mathbb{A}$ pp-defines a relation that is not Horn-definable and that has a minimal definition containing a non-Horn clause ψ such that:*
- *no negated linear equation is in ψ,*
- *at least one linear equation is in ψ.*

Proof. Let φ be a standard minimal definition of R in conjunctive normal form, and let ψ be a clause of φ that is not Horn. From Corollary 16, we can suppose that End($\mathbb{A}$) is $\{1\}$, $\mathbb{Z} \setminus \{0\}$, or $1 + d\mathbb{Z}$ for $d \geq 2$. This implies that either $\{1\}$ is pp-definable or, by Proposition 6, all the linear equations appearing in φ are homogeneous.

We can assume that ψ does not contain any negative literal, per the assumption that φ is minimal: indeed, consider the relation R' defined by the formula

$$\varphi' := \varphi \wedge \sum \lambda_i x_i = c \tag{†}$$

where $\sum \lambda_i x_i \neq c$ is in ψ. Either $c = 0$, in which case the relation R' defined by (†) is pp-definable in $\mathbb{A}$, or $\{1\}$ is pp-definable in $\mathbb{A}$ and R' is pp-definable in $\mathbb{A}$, too. The relation R' is not Horn-definable, and when we reduce the definition φ' of R' we obtain a formula that has fewer literals in clauses that contain more than one literal, in contradiction to the minimality of φ.

If ψ contains a linear equation then we are done. Otherwise, ψ only contains modular linear equations. If End($\mathbb{A}$) = $\mathbb{Z} \setminus \{0\}$ then by Proposition 6 any modulus of a modular linear equation appearing in ψ would have to be coprime with every nonzero integer, which is impossible. Therefore, End($\mathbb{A}$) is $\{1\}$ or $1 + d\mathbb{Z}$ for $d \geq 2$. In the latter case, we can assume by Proposition 6 that all the modular linear equations in ψ are modulo a divisor of d. In the former case, let d be a common multiple of all the moduli appearing in a modular linear equation in ψ. Consider the structure $\mathbb{A}/d\mathbb{A}$ defined in Proposition 8. The relation T

obtained from R in this structure is not a coset of a subgroup H of $(\mathbb{Z}/d\mathbb{Z})^k$ (where k is the arity of R): otherwise this coset is definable by a conjunction θ of modular linear equations modulo a divisor of d. Replacing ψ by θ in φ would produce a smaller definition of R, a contradiction to the minimality of φ. Moreover, $(\mathbb{A}/d\mathbb{A}, 1)$ is pp-interpretable in $\mathbb{A}$: in the two cases that $\mathrm{End}(\mathbb{A}) = \{1\}$ and $\mathrm{End}(\mathbb{A}) = 1 + d\mathbb{Z}$, the preimage of $\{1\}$ under the canonical projection $x \mapsto x \bmod d$ is pp-definable in $\mathbb{A}$. We conclude as in Proposition 8 that $\mathrm{CSP}(\mathbb{A})$ is NP-hard.   ◀

▶ **Theorem 18.** *Let $\mathbb{A}$ be a finite-signature first-order reduct of $(\mathbb{Z}; +, 1)$ which contains $+$ and is a core. Suppose that $\mathbb{A}$ contains a relation that is not Horn-definable. Then $\mathrm{CSP}(\mathbb{A})$ is NP-hard.*

Proof. From Lemma 17, we can suppose that $\mathbb{A}$ pp-defines a relation R that is not Horn-definable, that has a reduced standard definition φ containing a non-Horn clause ψ with at least one linear equation (L) and no negated linear equation. Since ψ is not Horn, it contains at least another equation (L'), possibly modular. Let $(a_1, \ldots, a_n)$ satisfy φ and only (L) in ψ. Such a tuple exists by the assumption that φ is reduced. Similarly, let $(b_1, \ldots, b_n)$ satisfy φ and only (L') in ψ. Let S be the binary relation such that $(\lambda, t) \in S$ if, and only if, $\lambda \in \mathrm{End}(\mathbb{A})$ and $t(\mathbf{a} - \mathbf{b}) + \lambda\mathbf{b}$ is in R. Note that $\mathbf{a}$ and $\mathbf{b}$ being fixed, S is pp-definable over $\mathbb{A}$. Therefore, we obtain a family $\{S_\lambda\}_{\lambda \in \Lambda}$ that is uniformly definable in $\mathbb{A}$, where $\Lambda = \mathrm{End}(\mathbb{A})$. Clearly, $1 \in \Lambda$, and $\Lambda \subseteq \mathbb{Z} \setminus \{0\}$ and $0, \lambda \in S_\lambda$. Moreover, note that

$$S_\lambda \cap \lambda \cdot \mathbb{Z} = \lambda \cdot S_1 \tag{‡}$$

holds for all $\lambda \in \mathrm{End}(\mathbb{A})$. Indeed:

$$\begin{aligned} t \in S_1 &\Leftrightarrow t(\mathbf{a} - \mathbf{b}) + \mathbf{b} \in R \\ &\Leftrightarrow \lambda t(\mathbf{a} - \mathbf{b}) + \lambda\mathbf{b} \in R &&\text{because } \mathbb{A} \text{ is a core} \\ &\Leftrightarrow \lambda t \in S_\lambda. \end{aligned}$$

We prove that for all $\lambda \in \mathrm{End}(\mathbb{A})$ the relation S_λ is not Horn-definable. Since $0, \lambda \in S_\lambda$, it suffices to prove that S_λ omits infinitely many multiples of λ, and by (‡) it suffices to prove that $\ell \notin S_1$ for infinitely many ℓ. Let ℓ be such that $\ell = 1 \bmod d'$, for every modulus d' appearing in ψ. We claim that $\ell(\mathbf{a} - \mathbf{b}) + \mathbf{b}$ does not satisfy any modular linear equation in ψ. Indeed, let $\sum_i \sigma_i x_i = c \bmod d'$ be such a modular linear equation. Then we have

$$\sum_i \sigma_i(\ell(a_i - b_i) + b_i) = c \bmod d' \Leftrightarrow \sum_i \sigma_i a_i = c \bmod d',$$

which is a contradiction to the choice of $\mathbf{a}$ since $\mathbf{a}$ only satisfies (L) in ψ and (L) is assumed to be non-modular. Consider now a linear equation $\sum \sigma_i x_i = c$ in ψ. This equation is satisfied by $\ell(\mathbf{a} - \mathbf{b}) + \mathbf{b}$ if, and only if

$$\ell \cdot \sum_i \sigma_i(a_i - b_i) = c - \sum_i \sigma_i b_i. \tag{⋆}$$

Suppose first that $\sum \sigma_i(a_i - b_i) = 0$. Then (⋆) is satisfied if, and only if, we have $\sum \sigma_i b_i = c = \sum \sigma_i a_i$. This implies that both $\mathbf{a}$ and $\mathbf{b}$ satisfy the equation; this is a contradiction to our choice of the vectors $\mathbf{a}$ and $\mathbf{b}$, so that $\ell(\mathbf{a} - \mathbf{b}) + \mathbf{b}$ does not satisfy (⋆). Suppose now that $\sum \sigma_i(a_i - b_i) \neq 0$. If $\ell > |c - \sum \sigma_i b_i|$, it is then clear that (⋆) is not satisfied. Therefore, for infinitely many ℓ, the tuple $\ell(\mathbf{a} - \mathbf{b}) + \mathbf{b}$ does not satisfy any literal in ψ and $\ell \notin S_1$.

Let $\theta(x, y)$ be a minimal reduced standard definition of S. By inspection of the formula θ, one finds that S_λ^+ and S_λ^- consist of points of the form $-\frac{a\lambda}{b}$, where $ax + by = 0$ is an equation in θ. Note that since a and b are taken coprime by the minimality of θ, if b divides $a\lambda$ then b divides λ. Let $m := \mathrm{lcm}\{|b| : ax + by = 0 \text{ is an equation in } \theta\}$, and note that $m\mathbb{Z} \cap \Lambda$ is not empty by the previous remark. Let $T \neq \emptyset$ be the binary relation defined by $\theta(m \cdot x, y)$. For all $\lambda \in \mathrm{End}(\mathbb{A})$, the set T_λ is not Horn-definable and of the form $(T_\lambda^\circ \cup (\lambda \cdot P)) \setminus (\lambda \cdot Q)$, where P and Q are finite sets that are independent of λ and $T_\lambda^\circ = S_{m\lambda}^\circ$. For the family of relations $\{T_\lambda\}_{\lambda \in \mathrm{End}(\mathbb{A})}$ to satisfy the compatibility condition, it remains to prove that $c + d\mathbb{Z} \subseteq T_1^\circ$ if, and only if, $\lambda c + d\mathbb{Z} \subseteq T_\lambda^\circ$, for all $d \geq 1$ and $c \in \{0, \dots, d-1\}$. Suppose that $c + d\mathbb{Z} \subseteq T_1^\circ$ for some $d \geq 1$ and suppose that the cosets of T_λ° are cosets of $d'\mathbb{Z}$. By Proposition 6, λ and d' are coprime. Therefore, there exists $\mu \in \mathbb{Z}$ such that $\lambda\mu = 1 \bmod d'$. We have $c + d\mu\mathbb{Z} \subseteq T_1^\circ$, because d divides $d\mu$. It follows that $\lambda c + \lambda d\mu\mathbb{Z} \subseteq T_\lambda^\circ$. Now, let $x \in \lambda c + d\mathbb{Z}$, say $x = \lambda c + qd$. Then we have $x = \lambda c + q\lambda\mu d \bmod d'$. Note that $\lambda c + q\lambda\mu d \in \lambda c + \lambda d\mu\mathbb{Z} \subseteq T_\lambda^\circ$. Since the cosets in T_λ° are cosets of $d'\mathbb{Z}$, we obtain that $x \in T_\lambda^\circ$ and consequently that $\lambda c + d\mathbb{Z} \subseteq T_\lambda^\circ$. Conversely, if $\lambda c + d\mathbb{Z} \subseteq T_\lambda^\circ$ then $\lambda c + d\lambda\mathbb{Z} \subseteq T_\lambda^\circ$, because d divides $d\lambda$. Since $\mathbb{A}$ is a core, $x \mapsto \lambda \cdot x$ is a self-embedding of $\mathbb{A}$, so that $c + d\mathbb{Z} \subseteq T_1^\circ$.

To conclude, the family of compatible relations $\{T_\lambda\}_{\lambda \in \mathrm{End}(\mathbb{A})}$ is uniformly pp-definable in $\mathbb{A}$ and consists of non-Horn relations. By Proposition 13, we obtain that $\mathrm{CSP}(\mathbb{A})$ is NP-hard. ◀

We illustrate our proofs in some examples below.

▶ **Example 19.** Consider the binary relation

$$S = \{(\lambda, t) \mid (\lambda = 1 \bmod 4 \wedge t = 1 \bmod 4) \vee (\lambda = 3 \bmod 4 \wedge t = 3 \bmod 4) \vee t = 0\}.$$

One sees that the set of endomorphisms of $\mathbb{A} := (\mathbb{Z}; +, S)$ is equal to $\mathrm{End}(\mathbb{A}) := 1 + 2\mathbb{Z}$. Moreover, S is not Horn-definable. For every $\lambda \in \mathrm{End}(\mathbb{A})$, one has $S_\lambda = \{0\} \cup (\lambda + 4\mathbb{Z})$. When λ is fixed, one can define a finite set by $\exists y(x \in S_\lambda \wedge y \in S_\lambda \wedge x + y = \lambda)$, which defines $\{0, \lambda\}$. One then obtains a reduction from 1-in-3-SAT by $\exists x, y, z \in \{0, \lambda\} : x + y + z \in \{0, \lambda\}$. Finally, by existentially quantifying over $\lambda \in \mathrm{End}(\mathbb{A})$ we obtain a reduction from 1-in-3-SAT to $\mathrm{CSP}(\mathbb{Z}; +, S)$.

▶ **Example 20.** Let $R := \{0\} \cup (1 + 3\mathbb{Z}) \cup (2 + 3\mathbb{Z})$ and $K = 1 + 3\mathbb{Z}$. Note that Proposition 13 does not apply to $\mathrm{CSP}(\mathbb{Z}; +, R)$ since $(\mathbb{Z}; +, R)$ is not a core (we have $0 \in \mathrm{End}(\mathbb{Z}; +, R)$). Neither does Corollary 16 apply to $\mathrm{CSP}(\mathbb{Z}; +, R, K)$ since $\mathrm{End}(\mathbb{Z}; +, R, K) = K$, which is clearly Horn-definable in $(\mathbb{Z}; +, R, K)$. But one obtains hardness of $\mathrm{CSP}(\mathbb{Z}; +, R, K)$ by Theorem 18. Indeed, pick $a = 0$ (satisfying the linear equation $x = 0$ in the definition of R) and $b = 1$ (satisfying the modular linear equation $x = 1 \bmod 3$ in the definition of R), and define the relation $S = \{(\lambda, t) \mid \lambda \in K \wedge \lambda - t \in S\}$. Note that for all $\lambda \in K$, we have $S_\lambda = \{\lambda\} \cup 3\mathbb{Z} \cup (2 + 3\mathbb{Z})$. The formula $\exists w(w \in K \wedge S(\lambda, t) \wedge S(\lambda, t + 3w))$ defines the relation $T = \{(\lambda, t) \mid \lambda = 1 \bmod 3 \wedge (t = 0 \bmod 3 \vee t = 2 \bmod 3)\}$, which is fully modular and not Horn-definable. Proposition 8 implies that $\mathrm{CSP}(\mathbb{Z}; +, R, S)$ is NP-hard.

5   Tractability

In this section we show the following.

▶ **Proposition 21.** *Let $\mathbb{A}$ be a structure with finite relational signature, domain $\mathbb{Z}$, and whose relations have quantifier-free Horn definitions over $(\mathbb{Z}; +, 1)$. Then there is an algorithm that solves $\mathrm{CSP}(\mathbb{A})$ in polynomial time.*

This result follows from the following more general result.

▶ **Theorem 22.** *Let φ be a quantifier-free Horn formula over $(\mathbb{Z}; +)$, allowing parameters from $\mathbb{Z}$ represented in binary. Then there exists a polynomial-time algorithm to decide whether φ is satisfiable over $(\mathbb{Z}; +)$.*

The proof of Theorem 22 can be found at the end of this section. We first show how to derive Proposition 21.

Proof of Proposition 21. The input of CSP($\mathbb{A}$) consists of a primitive positive sentence whose atomic formulas are of the form $R(x_1, \ldots, x_k)$ where R is quantifier-free Horn definable over $\mathcal{L}_{(\mathbb{Z};+,1)}$. Since $\sum_{i=1}^n a_i x_i = b \bmod c$ is equivalent to $\sum_{i=1}^n a_i x_i = b + ck$, where k is a new integer variable, we can as well assume that the input to our problem consists of a set of Horn clauses over $(\mathbb{Z}; +, 1)$. This is tacitly the process of quantifier introduction, the converse of quantifier elimination. Then apply Theorem 22. ◀

Our algorithm for the proof of Theorem 22 uses two other well-known algorithms:
1. a polynomial-time algorithm for satisfiability of linear diophantine equations, i.e., the subproblem of the computational problem from Theorem 22 where the input only contains atomic formulas (see, e.g., [23]).
2. a polynomial-time algorithm to compute the rank of a matrix over $\mathbb{Q}$; this allows us in particular to decide whether a given linear system of equalities implies another equality over the rationals (this is standard, using Gaussian elimination; again, see [23] for a discussion of the complexity).

These two algorithms can be combined to obtain the following.

▶ **Lemma 23.** *There is a polynomial-time algorithm that decides whether a given system Φ of linear diophantine equations implies another given diophantine equation ψ over $\mathbb{Z}$.*

Proof. First, use the first algorithm above to test whether Φ has a solution over $\mathbb{Z}$. If no, return yes (false implies everything). If yes, we claim that Φ implies ψ over $\mathbb{Q}$ (which can be tested by the second algorithm above) if and only if Φ implies ψ over the integers. Clearly, if every rational solution of Φ satisfies ψ, then so does every integer solution. Suppose now that there exists a rational solution α to Φ which does not satisfy ψ. Also take an integer solution β to Φ. Then on the line L that goes through α and β there are infinitely many integer points. If infinitely many points on a line satisfy ψ, then all points of the line must satisfy ψ. Since $\alpha \in L$ does not satisfy ψ it follows that an integer point on L does not satisfy ψ, i.e., Φ does not imply ψ over the integers. ◀

Given the two mentioned algorithms, our procedure for the proof of Theorem 22 is basically an implementation of positive unit clause resolution. It takes the same form as the algorithm presented in [7] for satisfiability over the rationals.

Proof. We follow the proof of Proposition 3.1 from [7]. We first discuss the correctness of the algorithm.

When $\mathcal{U}$ logically implies φ (which can be tested with the algorithm from Lemma 23) then the negative literal $\neg\varphi$ is never satisfied and can be deleted from all clauses without affecting the set of solutions. Since this is the only way in which literals can be deleted from clauses, it is clear that if one clause becomes empty the instance is unsatisfiable.

If the algorithm terminates with *satisfiable*, then no negation of an inequality is implied by $\mathcal{U}$. If r is the rank of the linear equation system defined by $\mathcal{U}$, we can use Gaussian

---

// Input: a set of Horn-clauses $\mathcal{C}$ over $(\mathbb{Z}; +)$ with parameters.

// Output: *satisfiable* if $\mathcal{C}$ is satisfiable in $(\mathbb{Z}; +)$, *unsatisfiable* otherwise

Let $\mathcal{U}$ be clauses from $\mathcal{C}$ that only contain a single positive literal.

If $\mathcal{U}$ is unsatisfiable then return *unsatisfiable*.

Do

    For all negative literals $\neg\varphi$ in clauses from $\mathcal{C}$

        If $\mathcal{U}$ implies φ, then delete the negative literal $\neg\varphi$ from all clauses in $\mathcal{C}$.

    If $\mathcal{C}$ contains an empty clause, then return *unsatisfiable*.

    If $\mathcal{C}$ contains a clause with a single positive literal ψ, then add $\{\psi\}$ to $\mathcal{U}$.

Loop until no literal has been deleted

Return *satisfiable*.

Figure 1 An algorithm for satisfiability of Horn formulas with parameters over $(\mathbb{Z}; +)$.

elimination to eliminate r of the variables from all literals in the remaining clauses. For each of the remaining inequalities, consider the sum of absolute values of all coefficients. Let S be one plus the maximum of this sum over all the remaining inequalities. Then setting the i-th variable to S^i satisfies all clauses. To see this, take any inequality, and assume that i is the highest variable index in this inequality. Order the inequality in such a way that the variable with highest index is on one side and all other variables on the other side of the $\neq$ sign. The absolute value on the side with the i-th variable is at least S^i. The absolute value on the other side is less than $S^i - S$, since all variables have absolute value less than S^{i-1} and the sum of all coefficients is less than $S - 1$ in absolute value. Hence, both sides of the inequality have different absolute value, and the inequality is satisfied. Since all remaining clauses have at least one inequality, all constraints are satisfied.

Now let us address the complexity of the algorithm. With appropriate data structures, the time needed for removing negated literals $\neg\varphi$ from all clauses when φ is implied by $\mathcal{U}$ is linearly bounded in the input size since each literal can be removed at most once. ◀

6 Conclusion

We are finally in position to prove the main result.

Proof of Theorem 1. Let $\mathbb{A}$ be a finite-signature first-order reduct of $(\mathbb{Z}; +, 1)$ that $\mathbb{A}$ contains $+$. By Lemma 5 there exists a core $\mathbb{B}$ of $\mathbb{A}$. If $\mathbb{B}$ has only one element then $\mathrm{CSP}(\mathbb{B})$ and $\mathrm{CSP}(\mathbb{A})$ are trivially in P. Otherwise, $\mathbb{B}$ is itself first-order definable in $(\mathbb{Z}; +, 1)$ and contains $+$ by Lemma 5, and the statement follows from Theorem 18 and Proposition 21. ◀

References

1 Libor Barto, Michael Kompatscher, Miroslav Olšák, Trung Van Pham, and Michael Pinsker. The equivalence of two dichotomy conjectures for infinite domain constraint satisfaction problems. In *32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017, Reykjavik, Iceland, June 20-23, 2017*, pages 1–12, 2017. `doi:10.1109/LICS.2017.8005128`.

2 Libor Barto, Jakub Opršal, and Michael Pinsker. The wonderland of reflections. *Israel Journal of Mathematics*, 2017. To appear. Preprint arXiv:1510.04521.

3 Libor Barto and Michael Pinsker. The algebraic dichotomy conjecture for infinite domain constraint satisfaction problems. In *Proceedings of the 31st Annual ACM/IEEE Symposium*

on Logic in Computer Science, LICS '16, New York, NY, USA, July 5-8, 2016, pages 615–622, 2016. `doi:10.1145/2933575.2934544`.

4   M. Bodirsky, B. Martin, and A. Mottet. Discrete temporal constraint satisfaction problems. *Journal of the ACM*, 65(2), 2018. preprint available at https://arxiv.org/abs/1503.08572. `doi:10.1145/3154832`.

5   Manuel Bodirsky, Víctor Dalmau, Barnaby Martin, Antoine Mottet, and Michael Pinsker. Distance constraint satisfaction problems. *Information and Computation*, 247:87–105, 2016.

6   Manuel Bodirsky, Peter Jonsson, and Timo von Oertzen. Essential convexity and complexity of semi-algebraic constraints. *Logical Methods in Computer Science*, 8(4), 2012. An extended abstract about a subset of the results has been published under the title *Semilinear Program Feasibility* at ICALP'10.

7   Manuel Bodirsky, Peter Jonsson, and Timo von Oertzen. Horn versus full first-order: Complexity dichotomies in algebraic constraint satisfaction. *J. Log. Comput.*, 22(3):643–660, 2012. `doi:10.1093/logcom/exr011`.

8   Manuel Bodirsky and Jan Kára. The complexity of temporal constraint satisfaction problems. *J. ACM*, 57(2), 2010. `doi:10.1145/1667053.1667058`.

9   Manuel Bodirsky and Marcello Mamino. Constraint Satisfaction Problems over Numeric Domains. In Andrei Krokhin and Stanislav Zivny, editors, *The Constraint Satisfaction Problem: Complexity and Approximability*, volume 7 of *Dagstuhl Follow-Ups*, pages 79–111. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2017. `doi: 10.4230/DFU.Vol7.15301.79`.

10   Manuel Bodirsky, Barnaby Martin, and Antoine Mottet. Constraint satisfaction problems over the integers with successor. In *Proceedings of ICALP'15*, 2015.

11   Manuel Bodirsky and Michael Pinsker. Schaefer's theorem for graphs. *Journal of the ACM*, 62(3):Article no. 19, 1–52, 2015. A conference version appeared in the Proceedings of STOC 2011, pages 655–664.

12   Andrei A. Bulatov. A dichotomy theorem for nonuniform CSPs. In *Proceedings of FOCS'17*, 2017. arXiv:1703.03021.

13   T.-W. J. Chou and G. E. Collins. Algorithms for the solution of systems of linear diophantine equations. *SIAM J. Computing*, 11:687–708, 1982.

14   T. Feder and M. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory. *SIAM Journal on Computing*, 28:57–104, 1998.

15   Peter Jonsson and Tomas Lööw. Computational complexity of linear constraints over the integers. *Artificial Intelligence*, 195:44–62, 2013. An extended abstract appeared at IJCAI 2011.

16   Peter Jonsson and Johan Thapper. Constraint satisfaction and semilinear expansions of addition over the rationals and the reals. *J. Comput. Syst. Sci.*, 82(5):912–928, 2016. `doi:10.1016/j.jcss.2016.03.002`.

17   Ravindran Kannan and Achim Bachem. Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix. *SIAM J. Comput.*, 8(4):499–507, 1979.

18   David Marker. *Model Theory: An Introduction.* Springer, 2002.

19   Daniele Micciancio and Bogdan Warinschi. *A Linear Space Algorithm for Computing the Hermite Normal Form*, pages 231–236. Association for Computing Machinery (ACM), United States, 2001.

20   Christos H. Papadimitriou. *Computational Complexity.* Addison-Wesley, 1994.

21   M. Presburger. über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. *Comptes Rendus du I congres de Mathématiciens des Pays Slaves*, pages 92–101, 1929.

22 T. J. Schaefer. The complexity of satisfiability problems. In *Proceedings of STOC'78*, pages 216–226, 1978.

23 Alexander Schrijver. *Theory of Linear and Integer Programming*. Wiley - Interscience Series in Discrete Mathematics and Optimization, 1998.

24 Arne Storjohann. Computing hermite and smith normal forms of triangular integer matrices. *Linear Algebra and its Applications*, 282:25–45, 1998.

25 Dmitriy Zhuk. The Proof of CSP Dichotomy Conjecture. In *Proceedings of FOCS'17*, 2017. arXiv:1704.01914.