Discrete Temporal Constraint Satisfaction Problems*

Manuel Bodirsky, Barnaby Martin, Antoine Mottet

November 9, 2017

Abstract

A *discrete temporal constraint satisfaction problem* is a constraint satisfaction problem (CSP) over the set of integers whose constraint language consists of relations that are first-order definable over the order of the integers. We prove that every discrete temporal CSP is in P or NP-complete, unless it can be formulated as a finite domain CSP in which case the computational complexity is not known in general.

Contents

*A subset of the results was announced in the proceedings of ICALP'15 under the title *Constraint Satisfaction Problems over the Integers with Successor*. Manuel Bodirsky and Antoine Mottet have received funding from the European Research Council (Grant Agreement no. 681988, CSP-Infinity), and the DFG (Project number 622397). Barnaby Martin was supported by EPSRC grant EP/L005654/1. Antoine Mottet is supported by DFG Graduiertenkolleg 1763 (QuantLA).

1 Introduction

"Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk."[1]
Leopold Kronecker

A *constraint satisfaction problem* 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 which constraint languages give rise to polynomial-time solvable CSPs has been the topic of intensive research over the past years. It has been conjectured by Feder and Vardi [13] that CSPs for constraint languages over finite domains have a complexity dichotomy: they are in P or NP-complete.

A famous CSP over the integers is *feasibility of systems of linear inequalities*. 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 computational problem, Jonsson and Lööw [18] proposed to study the class of CSPs where the constraint language Γ is definable in *Presburger arithmetic*; that is, it consists of relations that have a first-order definition over $(\mathbb{Z}; \leq, +)$. Equivalently, each relation $R(x_1, \ldots, x_n)$ in Γ 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 $\{x_1, \ldots, x_n\}$. Several constraint languages in this class are known where the CSP can be solved in polynomial time; an example of such a CSP is the problem of deciding whether a system of linear diophantine equations has a solution (a polynomial-time algorithm is given in [11]). However, a complete complexity classification for the CSPs of Jonsson-Lööw languages appears to be a very ambitious goal.

One of the most basic classes of constraint languages that falls into the framework of Jonsson and Lööw is the class of *distance constraint satisfaction problems* [3]. A distance constraint satisfaction problem is a CSP for a constraint language over the integers whose relations have a first-order definition over $(\mathbb{Z}; succ)$ where $succ$ is the successor relation. It has been shown previously that distance CSPs for *locally finite* constraint languages, that is, constraint languages whose relations have *bounded Gaifman degree*, are NP-complete, in P, or can be formulated with a constraint language over a finite domain [3]. Another class of problems which can be expressed as Jonsson-Lööw constraint satisfaction problems is the class of *temporal CSPs* [5]. This is the class of problems

[1] *"God made the integers, all the rest is the work of man."* Quoted in *Philosophies of Mathematics*, page 13, by Alexander George, Daniel J. Velleman, Philosophy, 2002.

whose constraint languages are over the rational numbers with relations definable over $(\mathbb{Q}; <)$. While the order of the rationals is not isomorphic to the order of the integers because of its density, this density is not witnessed by *finite structures*: for any finite substructure of $(\mathbb{Q}; <)$, one can find a substructure of $(\mathbb{Z}; <)$ that is order-isomorphic to it. It follows that for every structure Γ whose relations are first-order definable in $(\mathbb{Q}; <)$, there exists a structure Δ that is definable in $(\mathbb{Z}; <)$ and such that Γ and Δ have the same CSP. The converse is not true, since the structure $(\mathbb{Z}; succ)$ is a first-order reduct of $(\mathbb{Z}; <)$ that does not have the same CSP as any first-order reduct of $(\mathbb{Q}; <)$.

In the present article, we study the class of *discrete temporal constraint satisfaction problems*, that is, the constraint satisfaction problems whose constraint language is first-order definable over $(\mathbb{Z}; <)$. Reasoning about discrete time appears in many areas of theoretical computer science. For example for scheduling problems or for temporal logics in verification, time is often taken to be discrete. So we are interested in the computational complexity of basic temporal reasoning tasks for the discrete instead of the continuous setting. Our goal is two-fold: on the one hand, we want to find restricted classes of constraints where the corresponding CSP can be solved in polynomial time. Such polynomial-time tractable fragments can be valuable in many respects:

- Polynomial-time fragments might explain why a seemingly hard computational problem can be solved well in practice by a constraint solver. This might be the case because all the constraints that appear in an instance come from a large polynomial-time tractable class on which the solver runs in polynomial time (see, e.g., the tractable class presented in Section 6.1).

- Polynomial-time fragments might also form the basis for systems that follow the SAT modulo theories paradigm. The idea here is that the highly successful technology in modern SAT solvers can be successfully applied also to more expressive reasoning tasks where we replace the propositional variables by atomic formulas that are interpreted with respect to some fixed theory. If there is an efficient solver for satisfiability of sets of atomic formulas over this theory, then this might help the propagation mechanism of SAT modulo theories solvers. Indeed, deciding satisfiability of sets of atomic formulas can be modelled as a CSP. In Section 6.2 we will encounter such a fragment that has been discovered by the SAT modulo theories community, namely the max-atom problem [1] (this problem can be solved in polynomial time when the involved constants are represented in unary).

On the other hand, our second goal is to obtain a full complexity classification. Knowing all the easy and the hard cases for a large class of computational problems for discrete time is a powerful tool when studying temporal reasoning problems. Since such problems appear as subproblems in many different areas of theoretical computer science, we expect that a complete complexity classification will be useful in many different contexts.

Our main result shows that the class of discrete temporal CSPs exhibits a P/NP-complete dichotomy (modulo the Feder-Vardi conjecture for finite-domain CSPs; several authors claimed recently to have proved this conjecture [22, 10, 23]). This result properly extends the results mentioned above for locally finite distance CSPs, since *succ* is first-order definable over $(\mathbb{Z}; <)$. By the comments of the previous paragraph it also extends the classification of temporal CSPs. A cornerstone of our proof is the characterization of those problems that are discrete temporal CSPs but that are not temporal CSPs; the corresponding constraint languages have an interesting notion of *rank* which we use in the following to obtain a strong pre-classification of those languages up to homomorphic equivalence. The notion of rank is central to reduce the classification to the natural special case where the binary successor relation is part of the language.

Our proof relies on the so-called universal-algebraic approach; this is the first time that this approach has been used for constraint languages that are not finite or countably infinite *ω-categorical* (a countable structure is by definition ω-categorical if and only if it is the unique countable model of its first-order theory). The central insight of the universal-algebraic approach to constraint satisfaction is that the computational complexity of a CSP is captured by the set of *polymorphisms* of the constraint language. One of the ideas of the present article is that in order to use polymorphisms when the constraint language is not ω-categorical, we have to pass to the countable saturated model of the first-order theory of $(\mathbb{Z}; <)$. The relevance of saturated models for the universal-algebraic approach has already been pointed out in joint work of the first two authors with Martin Hils [4], but this is the first time that this perspective has been used to perform complexity classification for a large class of concrete computational problems. Our classification has a particularly simple form when the constraint language Γ not only contains the binary successor relation, but also the relation $<$: if Γ has the polymorphism $(x, y) \mapsto \max(x, y)$ or $(x, y) \mapsto \min(x, y)$, then CSP($\Gamma$) is in P, and is NP-hard otherwise.

The formal definitions of CSPs and discrete temporal CSPs can be found in Section 2. The border between discrete temporal CSPs in P and NP-complete discrete temporal CSPs can be most elegantly stated using the terminology of the mentioned universal-algebraic approach to constraint satisfaction. This is why we first give a brief introduction to this approach in Section 3, and only then give the technical description of our main result in Section 4. Section 5 gives a classification of the structures over the integers with finitely many relations definable over $(\mathbb{Z}; <)$ that might be of independent interest; this classification is the basis of our classification of the complexity of discrete temporal CSPs. Our algorithmic results can be found in Section 6. Finally, we put all the results together to prove our main result in Section 7. We discuss our result and promising future research questions in Section 8.

2 Discrete Temporal Constraint Satisfaction Problems

A *relational signature* is a set τ of relation symbols, where each symbol $R \in \tau$ has an *arity* $ar(R) \in \mathbb{N}$. Let $\tau = \{R_1, R_2, \dots\}$ be a relational signature. A *τ-structure* Γ is a tuple $(D; R_1^\Gamma, R_2^\Gamma, \dots)$ where D is a set – called the *domain* – and $R_i^\Gamma \subseteq D^{ar(R_i)}$ are relations on D. A *τ-formula* is a first-order formula built from the relations from τ, and equality. A τ-formula is *primitive positive (pp)* if it is of the form $\exists x_1, \dots, x_k(\psi_1 \wedge \dots \wedge \psi_m)$ where each ψ_i is an atomic τ-formula. *Sentences* are formulas without free variables.

Definition 1 (CSP(Γ))**.** Let Γ be a structure with a finite relational signature (also called the *constraint language*). Then the *constraint satisfaction problem for* Γ is the following computational problem.

Input: A primitive positive τ-sentence Φ.
Question: $\Gamma \models \Phi$?

A relational structure Γ is a *first-order reduct* of a structure Δ if it has the same domain as Δ and every relation R^Γ is *first-order definable* over Δ. That is, if R^Γ has arity k, there exists a first-order formula φ in the signature of Δ with k free variables such that for all elements $a_1, \dots, a_k$ of Γ we have $R^\Gamma(a_1, \dots, a_k) \Leftrightarrow \Delta \models \varphi(a_1, \dots, a_k)$.

Definition 2 (Discrete Temporal CSP)**.** A *discrete temporal CSP* is a constraint satisfaction problem where the constraint language is a first-order reduct of $(\mathbb{Z}; <)$ with finite signature.

Example 1. We present some concrete examples first-order reducts of $(\mathbb{Z}; <)$; some of the relations from these examples will re-appear in later sections to illustrate important phenomena for reducts of $(\mathbb{Z}; <)$.

1. $(\mathbb{Z}; succ^p)$, where $succ^p = \{(x, y) \in \mathbb{Z}^2 \mid y = x + p\}$ for $p \in \mathbb{Z}$. Note that this structure is not connected, and that it has the same CSP as $(\mathbb{Z}; succ)$. This example and example (3) will be considered again in Example 4.

2. $(\mathbb{Z}; \mathrm{Diff}_S)$, where $\mathrm{Diff}_S := \{(x, y) \in \mathbb{Z}^2 \mid y - x \in S\}$ for a finite set $S \subset \mathbb{Z}$.

3. $(\mathbb{Z}; succ^2, \mathrm{Diff}_{\{-2,-1,0,1,2\}})$.

4. $(\mathbb{Z}; F)$ where F is the 4-ary relation $\{(x, y, u, v) : y = x + 1 \Leftrightarrow v = u + 1\}$. This example and the following example have unbounded Gaifman degree (see Section 5.1), so they do not fall into the scope of [3].

5. $(\mathbb{Z}; \neq, \mathrm{Dist}_i)$ where $i \in \mathbb{N}$ and $\mathrm{Dist}_i := \{(x, y) : |x - y| = i\}$.

6. $(\mathbb{Z}; \{(x, y, z) \in \mathbb{Z}^3 \mid z+1 \leq \max(x, y)\})$. This structure is not as first-order reduct of $(\mathbb{Z}; succ)$. Neither does it have the same CSP as a first-order reduct of $(\mathbb{Q}; <)$, so we have a discrete temporal CSP that is not a temporal CSP and does not fall into the scope of [5]. The CSP for this structure is closely related to the so-called Max-Atom problem; the connection is explained in Section 8.

For a subset S of the domain of a structure Γ, we write $\Gamma[S]$ for the structure induced on S by Γ. The structure $(\mathbb{Z}; <)$ admits quantifier elimination in the language consisting of the binary relations $R_c = \{(x, y) \in \mathbb{Z}^2 \mid y \leq x + c\}$ for $c \in \mathbb{Z}$. This means that every first-order formula $\varphi(x_1, \dots, x_n)$ in the expanded language $\{R_c \mid c \in \mathbb{Z}\}$ is equivalent to a quantifier-free formula in the same language. To see this, note that it suffices to prove that one can eliminate the quantifiers in existential formulas rather than in general first-order formulas; in fact, by de Morgan and the equivalence between $\neg y < x + c$ and $x < y + (1 - c)$ it suffices to prove that one can eliminate the quantifiers in primitive positive formulas. Seeing a primitive positive formula as a system of inequalities, one then performs Gaussian elimination to remove the variables that are existentially quantified. The result of this is a system of inequalities that can be translated back into a quantifier-free formula. Similarly, $(\mathbb{Z}; succ)$ admits quantifier elimination in the language consisting of the binary relations given by $y = x + c$ for $c \in \mathbb{Z}$. Whenever we write that φ is a quantifier-free formula, we mean that φ is written in one of those two languages; which one will always be clear from the context. The empty relation, $\mathbb{Z}^2$, and the binary relations defined by $y = x+c$ for $c \in \mathbb{Z}$ are called *basic* relations. The following is easy to see.

Proposition 1. *All discrete temporal CSPs are in NP.*

Proof. Let q be the size of the biggest integer that appears in the quantifier-free formulas that define the relations in Γ over $(\mathbb{Z}; <)$; that is, for any atomic formula $x \leq y + k$ in those formulas, $k \in \mathbb{Z}$, we have $|k| \leq q$. For an instance Φ of $\mathrm{CSP}(\Gamma)$ with n variables, it is clear that $\Gamma \models \Phi$ if and only if Φ is true on $\Gamma[\{1, \dots, (q + 1)n\}]$. We may guess a satisfying assignment of values from $\{1, \dots, (q + 1)n\}$ to the variables of Φ, and verify in polynomial time that all the constraints are satisfied. $\square$

The main result of our article (Theorem 3) immediately implies the following.

Theorem 1. *Every discrete temporal CSP is NP-complete, in P, or polynomial-time equivalent to a finite-domain CSP.*

3 The Algebraic Approach

The starting point of the universal algebraic approach to analyze the complexity of CSPs is the observation that when a relation R can be defined by a primitive positive formula over Γ, then $\mathrm{CSP}(\Gamma)$ can simulate the 'richer' problem $\mathrm{CSP}(\Delta)$ where $\Delta = (\Gamma, R)$ has been obtained from Γ by adding R as another relation. The proof of this fact given by Jeavons, Cohen, and Gyssens [17] works for all structures Γ over finite or over infinite domains. Since we will use this fact very frequently, we will not explicitly refer back to it from now on.

Polymorphisms are an important tool to study the question which relations are primitive positive definable in Γ. We say that a function $f\colon D^n \to D$ *preserves* a relation $R \subseteq D^m$ if for all $t_1, \dots, t_n \in R$ the tuple $f(t_1, \dots, t_n)$ obtained by applying f componentwise to the tuples $t_1, \dots, t_n$ is also in R; otherwise, f *violates* R. A *polymorphism* of a relational structure Γ with domain D is a function from D^n to D, for some finite n, which preserves all relations of Γ. An *endomorphism* is a unary polymorphism. An *embedding* of a structure Γ is an injective endomorphism of Γ that also preserves the complement of the relations of Γ. An *automorphism* is a surjective embedding. We write $\mathrm{Pol}(\Gamma)$, $\mathrm{End}(\Gamma)$, and $\mathrm{Aut}(\Gamma)$ for the set of all polymorphisms, endomorphisms, and automorphisms of Γ.

We write $\mathcal{O}$ for $\bigcup_{k\in\mathbb{N}}(D^k \to D)$. A subset $\mathcal{F}$ of $\mathcal{O}$ *generates* $f \in \mathcal{O}$ if f can be obtained from projections and functions in $\mathcal{F}$ by composition. Note that every function generated by polymorphisms of Γ is again a polymorphism. We will need the fact that the set of all polymorphisms of Γ is furthermore *locally closed*, that is, when $f\colon D^k \to D$ is such that for all finite $S \subseteq D^k$ there exists an $e \in \mathrm{Pol}(\Gamma)$ such that $e(x) = f(x)$ for all $x \in S$, then f is also a polymorphism of Γ. A subset $\mathcal{F}$ of $\mathcal{O}$ *locally generates* $f \in \mathcal{O}$ if for every finite subset S of D, there exists a function g that is generated by $\mathcal{F}$ and such that the restrictions of g and f to S coincide.

The polymorphisms of a structure Γ also preserve all relations that are primitive positive definable in Γ; this holds for arbitrary finite and infinite structures Γ. If Γ is finite [9, 14] or ω-categorical [7], then a relation is preserved by all polymorphisms *if and only if* it is primitively positively definable in Γ.

The structures that we consider in this article will in general not be ω-categorical; however, following the philosophy in [4], one can refine these universal-algebraic methods to apply them also in our situation. We will describe these refinements in the rest of this section.

The *(first-order) theory* of a structure Γ, denoted by $\mathrm{Th}(\Gamma)$, is the set of all first-order sentences that are true in Γ. We define some notation to conveniently work with models of $\mathrm{Th}(\Gamma)$ and their first-order reducts.

Definition 3 ($\kappa.\mathbb{Z}$)**.** Let κ be a linearly ordered set. We write $\kappa.\mathbb{Z}$ for κ copies of $\mathbb{Z}$ indexed by the elements of κ; formally, $\kappa.\mathbb{Z}$ is the set $\{(p, z) : p \in \kappa, z \in \mathbb{Z}\}$. Then $(\kappa.\mathbb{Z}; <)$ is the structure where $<$ denotes the lexicographic order on $\kappa.\mathbb{Z}$, that is, we define $(p, z) < (p', z')$ if $p < p'$ holds or if $p = p'$ and $z < z'$. If $p \in \kappa$, we write $p.\mathbb{Z}$ to denote the copy of $\mathbb{Z}$ indexed by p, instead of $\{p\} \times \mathbb{Z}$.

It is well known and easy to see that the models of $\mathrm{Th}(\mathbb{Z}; <)$ are precisely the structures isomorphic to $(\kappa.\mathbb{Z}; <)$, for some linear order κ. When $k \in \mathbb{Z}$ and $u = (p, z) \in \kappa.\mathbb{Z}$, we write $u + k$ for $(p, z + k)$.

Definition 4 ($\kappa.\Gamma$)**.** Let Γ be a first-order reduct of $(\mathbb{Z};<)$ with signature τ. Then $\kappa.\Gamma$ denotes the 'corresponding' first-order reduct of $(\kappa.\mathbb{Z};<)$ with signature τ. Formally, when $R \in \tau$ and φ_R is a formula that defines R^Γ, then $R^{\kappa.\Gamma}$ is the relation defined by φ_R over $(\kappa.\mathbb{Z};<)$.

In the following, we identify $\mathbb{Z}$ with the copy of $\mathbb{Z}$ induced by $0.\mathbb{Z}$ in $\mathbb{Q}.\mathbb{Z}$. That is, we view $(\mathbb{Z};<)$ as a substructure of $(\mathbb{Q}.\mathbb{Z};<)$, and consequently Γ as a substructure of $\mathbb{Q}.\Gamma$ for each first-order reduct Γ of $(\mathbb{Z};<)$. The structures Γ and $\mathbb{Q}.\Gamma$ have the same first-order theory; in particular, they satisfy the same primitive positive sentences. It follows that Γ and $\mathbb{Q}.\Gamma$ have the same CSP. Let $\varphi(x_1,\ldots,x_k)$ be a first-order formula in the language of Γ. This formula defines a relation $R \subseteq \mathbb{Z}^k$ in Γ and a relation $R' \subseteq (\mathbb{Q}.\mathbb{Z})^k$. One sees (for example using quantifier elimination) that $R = R' \cap \mathbb{Z}^k$, i.e., the relations definable in Γ are precisely the intersections of $\mathbb{Z}$ with relations defined in $\mathbb{Q}.\Gamma$. The link between endomorphisms of Γ and of $\mathbb{Q}.\Gamma$ is more complicated, and is covered in Section 5.

A *type* of a structure Δ is a set p of formulas with free variables $x_1,\ldots,x_n$ such that $p \cup \mathrm{Th}(\Delta)$ is satisfiable (that is, $\{\varphi(c_1,\ldots,c_n) : \varphi \in p\} \cup \mathrm{Th}(\Delta)$, for new constant symbols $c_1,\ldots,c_n$, has a model). The type of a tuple $(a_1,\ldots,a_n)$ in a structure Δ is the set of first-order formulas $\varphi(x_1,\ldots,x_n)$ such that $\Delta \models \varphi(a_1,\ldots,a_n)$. A countable τ-structure Γ is *saturated* if for all choices of finitely many elements $a_1,\ldots,a_n$ in Γ, and every unary type p of $(\Gamma,a_1,\ldots,a_n)$, there exists an element b of Γ such that $(\Gamma,a_1,\ldots,a_n) \models \varphi(b)$ for all $\varphi \in p$. When Γ and Δ are two countable saturated structures with the same first-order theory, then Γ and Δ are isomorphic [16, Theorem 8.1.8]. Note that $(\mathbb{Q}.\mathbb{Z};<)$ is saturated. More generally, $\mathbb{Q}.\Gamma$ is saturated for every first-order reduct Γ of $(\mathbb{Z};<)$.

We define the function $-\colon (\kappa.\mathbb{Z})^2 \to (\mathbb{Z} \cup \{\infty\})$ for $x,y \in \kappa.\mathbb{Z}$ by

$$\begin{aligned} x - y &:= k \in \mathbb{Z} && \text{if } x = y + k \\ x - y &:= \infty && \text{otherwise.} \end{aligned}$$

When Γ and Δ are two structures with the same relational signature τ, then a *homomorphism* from Γ to Δ is a function f from the domain of Γ to the domain of Δ such that for every $R \in \tau$ of arity k we have $R^\Gamma(u_1,\ldots,u_k) \Rightarrow R^\Delta(f(u_1),\ldots,f(u_k))$. If there is a homomorphism from Γ to Δ, and vice versa, then $\mathrm{CSP}(\Gamma)$ and $\mathrm{CSP}(\Delta)$ are the same computational problem.

Lemma 1 (See Lemma 2.1 in [4])**.** *Let Γ be a countable saturated structure, let Δ be countable, let $d_1,\ldots,d_k$ be elements of Δ, and let $c_1,\ldots,c_k$ be elements of Γ. Suppose that for all primitive positive formulas φ such that $\Delta \models \varphi(d_1,\ldots,d_k)$ we have $\Gamma \models \varphi(c_1,\ldots,c_k)$. Then there exists a homomorphism from Δ to Γ that maps d_i to c_i for all $i \leq k$.*

To classify the computational complexity of the CSP for all first-order reducts of a structure Γ, it often turns out to be important to study the possible endomorphisms of those reducts first, before studying the polymorphisms. This has for instance been the case for the first-order reducts of $(\mathbb{Q};<)$ in [5] and the first-order reducts of the countably infinite random graph in [8].

We are now in the position to state a general fact, Theorem 2, whose proof might explain the importance of saturated models for the universal-algebraic approach. Let Γ be a structure with domain D. A relation $R \subseteq D^k$ is said to be *n-generated* under $\mathrm{End}(\Gamma)$ if there exist tuples $t_1,\ldots,t_n \in R$ such that for every $t \in R$, there exist $e \in \mathrm{End}(\Gamma)$ and $i \in \{1,\ldots,n\}$ such that $e(t_i) = t$. An *existential positive formula* is a first-order formula without universal quantifiers and without negations. A *universal negative formula* is a first-order formula without existential quantifiers where the negation symbol only appears before an atom, and where all the atoms are negated.

Theorem 2. *Let Γ be a countable saturated structure, let Δ be a first-order reduct of Γ, and R a relation with a first-order definition in Γ. Then*

- *R has a first-order definition in Δ if and only if R is preserved by the automorphisms of Δ;*
- *R has an existential positive definition in Δ if and only if R is preserved by all the endomorphisms of Δ;*
- *if R is n-generated under $\mathrm{End}(\Delta)$, then R has a primitive positive definition in Δ if and only if R is preserved by all polymorphisms of Δ of arity n.*

Proof. Suppose that R is k-ary, and let φ be the first-order definition of R in Γ. It is well-known that first-order formulas are preserved by automorphisms of Δ, that existential positive formulas are preserved by endomorphisms of Δ, and that primitive positive formulas are preserved by polymorphisms of Δ.

Suppose first that R is preserved by all automorphisms of Δ. Let φ be a first-order definition of R in Γ. Let Ψ be the set of all first-order formulas in the language of Δ that are consequences of R. Formally,

$$\Psi = \{\psi(x_1, \ldots, x_k) \mid \forall (a_1, \ldots, a_k) \in R, \Delta \models \psi(a_1, \ldots, a_k)\}.$$

We prove that if a tuple $\overline{a}$ satisfies every formula in Ψ then $\overline{a}$ is in R. Let $\overline{a}$ be such a tuple. Let p be the type of $\overline{a}$ in Δ. By replacing in p every relation symbol of the signature of Δ by a first-order definition of the corresponding relation in Γ, we obtain a set q of formulas in the language of Γ. If we can find some tuple $\overline{b}$ that satisfies $\{\varphi\} \cup q$ in Γ, then we are done. Indeed, we have that $\overline{b}$ is in R, and $\overline{b}$ has the same type as $\overline{a}$ in Δ. The fact that $\overline{a}$ and $\overline{b}$ have the same type is equivalent to the fact that the structures $(\Delta, \overline{a})$ and $(\Delta, \overline{b})$ have the same first-order theory. We stated above that two countable saturated structures with the same first-order theory are isomorphic. Therefore, there exists an isomorphism $\alpha\colon (\Delta, \overline{b}) \to (\Delta, \overline{a})$. This isomorphism is an automorphism of Δ that maps $\overline{b}$ to $\overline{a}$, so that $\overline{a}$ is in R. So let us assume that $\{\varphi\} \cup q$ is not satisfiable in Γ. Since Γ is saturated, the set $\{\varphi\} \cup q$ cannot possibly be a type. It follows that $\mathrm{Th}(\Gamma) \cup q \cup \{\varphi\}$ is not satisfiable. By the compactness theorem of first-order logic, there exists a finite subset q' of q such that $\mathrm{Th}(\Gamma) \cup q' \cup \{\varphi\}$ is not satisfiable. Note that q is closed under conjunctions of formulas, so that the conjunction of all the formulas of q' is a formula ψ in q. Therefore, $\mathrm{Th}(\Gamma) \cup \{\varphi, \psi\}$ is not satisfiable, i.e., we have $\mathrm{Th}(\Gamma) \models \forall x_1, \ldots, x_k(\varphi(x_1, \ldots, x_k) \Rightarrow \neg\psi(x_1, \ldots, x_k))$. By construction, the formula ψ corresponds to a formula θ in the language of Δ. We obtain that $\neg\theta$ is in Ψ, so $\neg\theta$ is in p. But $\theta \in p$, a contradiction.

Suppose now that R is preserved by all endomorphisms of Δ. In particular R is preserved by all the automorphisms of Δ, so that there exists a first-order definition φ of R in Δ. Let Ψ be the set of all universal negative consequences of R in Δ. Formally,

$$\Psi = \{\psi(x_1, \ldots, x_k) \text{ universal negative formula } \mid \forall (a_1, \ldots, a_k) \in R, \Delta \models \psi(a_1, \ldots, a_k)\}.$$

As above, we aim to prove that if $\overline{a}$ satisfies all the formulas in Ψ, then $\overline{a}$ is in R. Let $\overline{a}$ be such a tuple, and let now p be the ep-type of $\overline{a}$, that is, the set of all the existential positive formulas ψ such that $\Delta \models \psi(\overline{a})$. If $p \cup \{\varphi\}$ is satisfiable in Δ, then we are done: there exists a tuple $\overline{b} \in R$ that has the same existential positive type as $\overline{a}$. Lemma 1 implies that there exists an endomorphism of Δ that maps $\overline{b}$ to $\overline{a}$, so that $\overline{a}$ is in R. If $p \cup \{\varphi\}$ is not satisfiable in Δ, there exists a single formula $\psi \in p$ such that $\Gamma \models \forall x_1, \ldots, x_k(\varphi(x_1, \ldots, x_k) \Rightarrow \neg\psi(x_1, \ldots, x_k))$. To ψ corresponds an existential positive formula θ in the language of Δ. We obtain that $\neg\theta$ is equivalent to a formula in Ψ, so that $\overline{a}$ must satisfy $\neg\theta$, contradicting the fact that $\overline{a}$ already satisfies θ.

Finally, suppose that R is n-generated under $\mathrm{End}(\Delta)$, and that R is preserved by all polymorphisms of Δ of arity n. Let $(b_1^1, \dots, b_k^1), \dots, (b_1^n, \dots, b_k^n)$ be n tuples of length k generating the relation R under $\mathrm{End}(\Delta)$. Let Ψ be the set of all primitive positive formulas with free variables $x_1, \dots, x_k$ that hold on all these tuples, i.e.

$$\Psi = \{\psi(x_1, \dots, x_k) \text{ pp-formula} \mid \forall i \in \{1, \dots, n\}, \Delta \models \psi(\bar{b}^i)\}.$$

If $\overline{a}$ is in R, there exists by assumption an endomorphism e of Δ and an $i \in \{1, \dots, n\}$ such that $e(\bar{b}^i) = \overline{a}$. Since primitive positive formulas are preserved by endomorphisms, the tuple $\overline{a}$ satisfies every primitive positive formula that $\bar{b}^i$ satisfies, so that in particular $\overline{a}$ satisfies Ψ. We now prove the converse. If $\overline{a}$ satisfies Ψ, we have that every primitive positive formula that holds on $(b_1^1, \dots, b_k^1), \dots, (b_1^n, \dots, b_k^n)$ in Δ^n also holds on $\overline{a}$. By Lemma 1 and saturation of Δ, there exists a homomorphism from Δ^n to Δ that maps $(b_i^1, \dots, b_i^n)$ to a_i for all $i \in \{1, \dots, k\}$. This map is a polymorphism of Δ, and since R is preserved by polymorphisms of arity n, $(a_1, \dots, a_k) \in R$. Therefore, $\overline{a}$ satisfies Ψ if and only if $\overline{a} \in R$. Similarly as before, a compactness argument for first-order logic over Γ shows that Ψ is equivalent to a single primitive positive formula that is equivalent to φ. □

4 Detailed Statement of the Results

In this section, we describe the border between the NP-complete and the polynomial-time tractable discrete temporal CSPs, modulo the Feder-Vardi dichotomy conjecture.

Definition 5. Let d be a positive integer. The *d-modular max*, $\max_d \colon \mathbb{Z}^2 \to \mathbb{Z}$, is defined by $\max_d(x, y) := \max(x, y)$ if $x = y \mod d$ and $\max_d(x, y) := x$ otherwise. The *d-modular min* is defined analogously, with $\min_d(x, y) = \min(x, y)$ if $x = y \mod d$ and $\min_d(x, y) = x$ otherwise.

Note that $\max_d$ and $\min_d$ are not commutative when $d > 1$. Also note that $\max_1 = \max$ and $\min_1 = \min$ are the usual maximum and minimum operations. Examples of relations which are preserved by max and which are definable over $(\mathbb{Z}; <)$ are the relations appearing in the last item of Example 1. An example of a relation which is preserved by $\max_d$ is the ternary relation containing the triples of the form

$$(a + d, a, a), (a + d, a + d, a), (a, a + d, a)$$

for all $a \in \mathbb{Z}$. Note that for a fixed d, this relation is preserved by $\max_d$ but not by $\max_{d'}$ for any other d'.

Theorem 3. *Let Γ be a first-order reduct of $(\mathbb{Z}; <)$ with finite signature. Then there exists a structure Δ such that $\mathrm{CSP}(\Delta)$ equals $\mathrm{CSP}(\Gamma)$ and at least one of the following cases applies.*

1. *Δ has a finite domain, and the CSP for Γ is conjectured to be in P or NP-complete [13].*

2. *Δ is a reduct of $(\mathbb{Q}; <)$, and the complexity of $\mathrm{CSP}(\Delta)$ has been classified in [5].*

3. *Δ is a reduct of $(\mathbb{Z}; <)$ and preserved by max or by min. In this case, $\mathrm{CSP}(\Delta)$ is in P.*

4. *Δ is a reduct of $(\mathbb{Z}; succ)$ such that Δ is preserved by a modular max or a modular min, or such that $\mathbb{Q}.\Delta$ is preserved by a binary injective function preserving succ. In this case,* $\mathrm{CSP}(\Delta)$ *is in P.*

5. $\mathrm{CSP}(\Gamma)$ *is NP-complete.*

As an illustration of the algorithmic consequences of our main result, we give examples of computational problems that can be formulated as discrete temporal CSPs and are in P.

Example 2. Fix positive integers $d, C \geq 1$.

Input: A system of constraints of the form ($x = y \bmod d$ and $a \leq x - y \leq b$) where $a, b \in \mathbb{Z}$ are such that $|a|, |b| \leq C$.
Question: Is the system satisfiable in $\mathbb{Z}^n$?

This problem can be seen as $\mathrm{CSP}(\mathbb{Z}; \mathrm{Diff}_{S_1}, \dots, \mathrm{Diff}_{S_k})$ where $S_1, \dots, S_k$ are all the sets of the form $\{a, a+d, \dots, b\}$ for $a, b \in \mathbb{Z}$, $|a|, |b| \leq C$, and $d | (b-a)$. All the relations are preserved by the d-modular maximum function, and thus Theorem 3 implies that this CSP is in P.

Example 3. Consider the reduct $(\mathbb{Z}; R, succ)$ of $(\mathbb{Z}; <)$ where

$$R := \{(x, y, z) \in \mathbb{Z} \mid x \leq \max(y, z)\}$$

The relations R and $succ$ are preserved by the (regular) maximum function, and thus Theorem 3 implies that this CSP is in P. The problem $\mathrm{CSP}(\mathbb{Z}; R, succ)$ is easily seen to be equivalent to the so-called *Max-Atom problem* [1] where numbers are represented in unary, which is known to be in P; see Section 8.

5 Definability of Successor and Order

The goal of this section is a proof that the CSPs for first-order reducts of $(\mathbb{Z}; <)$ fall into five classes. This will allow us to focus in later sections on first-order reducts of $(\mathbb{Z}; <)$ where $succ$ is pp-definable.

Theorem 4. *Let Γ be a first-order reduct of $(\mathbb{Z}; <)$ with finite signature. Then* $\mathrm{CSP}(\Gamma)$ *equals* $\mathrm{CSP}(\Delta)$ *where Δ is one of the following:*

1. *a finite structure;*

2. *a first-order reduct of $(\mathbb{Q}; <)$;*

3. *a first-order reduct of $(\mathbb{Z}; <)$ where* Dist_k *is pp-definable for all $k \geq 1$;*

4. *a first-order reduct of $(\mathbb{Z}; <)$ where succ and $<$ are pp-definable;*

5. *a first-order reduct of $(\mathbb{Z}; succ)$ where succ is pp-definable.*

The proof of this result requires some effort and spreads over the following subsections. Before we go into this, we explain the significance of the five classes for the CSP.

For the first class, we mention that it is still open whether finite-domain CSPs exhibit a complexity dichotomy (the Feder-Vardi conjecture states that these problems are either in P or NP-complete), and to the best of our knowledge this is also open for those finite structures Δ that appear as cores of reducts of $(\mathbb{Z};<)$ (such structures Δ must have a transitive automorphism group). However, finite-domain CSPs have a well-developed theory with many partial results, and there is an algebraic condition for Γ that implies NP-hardness of $\mathrm{CSP}(\Gamma)$ which is believed to capture precisely those finite-domain CSPs that are not in P. The study of finite-domain CSPs is outside the scope of the present article, so we focus on the remaining four classes.

The CSPs for first-order reducts of $(\mathbb{Q};<)$ have been studied by Bodirsky and Kára [5]; they are either in P or NP-complete. Hence, we are done if there exists a first-order reduct Δ of $(\mathbb{Q};<)$ such that $\mathrm{CSP}(\Delta) = \mathrm{CSP}(\Gamma)$. Several equivalent characterisations of those first-order reducts Γ will be given in Section 5.4. This is essential for proving Theorem 4.

When Γ is a first-order reduct of $(\mathbb{Z};<)$ where for all $k \geq 1$ the relation Dist_k is pp-definable, then $\mathrm{CSP}(\Gamma)$ is NP-complete; this is a consequence of Proposition 27 from [3], restated here.

Proposition 2. *Suppose that the relations* Dist_1 *and* Dist_5 *are pp-definable in* Γ. *Then* $\mathrm{CSP}(\Gamma)$ *is NP-hard.*

The previous paragraphs explain why Theorem 4 indeed reduces the complexity classification of CSPs for finite-signature first-order reducts Γ of $(\mathbb{Z};<)$ to the case where *succ* is pp-definable in Γ, which corresponds to the classes (4) and (5) of Theorem 4.

5.1 Degrees

We consider three notions of *degree* for relations R that are first-order definable in $(\mathbb{Z};<)$:

- For $x \in \mathbb{Z}$, we consider the number of $y \in \mathbb{Z}$ that appear together with x in a tuple from R; this number is the same for all $x \in \mathbb{Z}$, and called the *Gaifman-degree* of R (it is the degree of the Gaifman graph of $(\mathbb{Z};R)$).
- The *distance degree* of R is the supremum of d such that there are $x, y \in \mathbb{Z}$ that occur together in a tuple of R and $|x-y| = d$.
- The *quantifier-elimination degree (qe-degree)* of R is the minimal q so that there is a quantifier-free definition φ of R, such that for every literal $x \leq y + c$ in φ, we have $|c| \leq q$.

The degree of a first-order reduct of $(\mathbb{Z};<)$ is the supremum of the degrees of its relations, for any of the three notions of degree. The article [3] considered first-order reducts of $(\mathbb{Z};succ)$ with finite Gaifman-degree. Note that the Gaifman-degree is finite if and only if the distance degree is finite. In this article, qe-degree will play the central role, as any first-order reduct of $(\mathbb{Z};<)$ with finite relational signature has finite qe-degree.

5.2 Compactness

In this section we present some results, based on applications of König's tree lemma, that show how properties of finite substructures of finite-signature first-order reducts Γ of $(\mathbb{Z};<)$ correspond to the existence of certain homomorphisms from Γ to $\mathbb{Q}.\Gamma$. We first recall the statement of König's tree lemma.

Lemma 2 (König's tree lemma)**.** *Let* $\mathcal{T}$ *be an infinite tree such that each vertex has finitely many neighbours. Then* $\mathcal{T}$ *contains an infinite path.*

Let $(\kappa.\mathbb{Z};<)$ be a model of $\mathrm{Th}(\mathbb{Z};<)$, let S be any set, let $s \in \mathbb{N}$, and $f\colon S \to \kappa.\mathbb{Z}$. We say that $x, y \in S$ are (f,s)-*connected* if there is a sequence $x = u_1, \dots, u_k = y \in S$ so that $0 \leq |f(u_i) - f(u_{i+1})| \leq s$ for all $i \in \{1, \dots, k-1\}$. Note that this notion of connectivity defines an equivalence relation on S whose equivalence classes are naturally ordered. We define an equivalence relation $\sim_s$ on functions $f, g\colon S \to \kappa.\mathbb{Z}$ as follows: $f \sim_s g$ when the following conditions are met:

- $x, y \in S$ are (f,s)-connected if and only if they are (g,s)-connected,
- if $x, y \in S$ are (f,s)-connected (and therefore (g,s)-connected) then $f(x) - f(y) = g(x) - g(y)$,
- if $x, y \in S$ are not (f,s)-connected then $f(x) < f(y) \Leftrightarrow g(x) < g(y)$.

In other words, $f \sim_s g$ iff the equivalence relations defined by (f,s)-connectivity and (g,s)-connectivity have the same equivalence classes, are such that within each equivalence class the pairwise distances are the same, and the order of the equivalence classes is the same. This implies that if S is a finite set, there are only finitely many $\sim_s$-equivalence classes of functions $S \to \kappa.\mathbb{Z}$. Note that if $f \sim_s g$ and $s' \leq s$ then we also have $f \sim_{s'} g$.

Lemma 3 (Substitution Lemma)**.** *Let Γ be a first-order reduct of $(\mathbb{Z};<)$ with qe-degree q, and let Δ be a structure with the same signature as Γ and domain D. Let κ be a linearly ordered set. Let $f, g\colon D \to \kappa.\mathbb{Z}$ be such that $f \sim_q g$. Then f is a homomorphism from Δ to $\kappa.\Gamma$ if and only if g is such a homomorphism.*

Proof. Suppose that f is a homomorphism from Δ to $\kappa.\Gamma$. To prove that g is a homomorphism, it suffices to prove that $g(a) < g(b) + c$ if and only if $f(a) \leq f(b) + c$ for all $a, b \in D$ and $|c| \leq q$. This follows from the fact that every relation of Γ can be defined from literals of the form $x \leq y + c$ with $|c| \leq q$ using conjunctions and disjunctions. Let $a, b \in D$ and suppose that $f(a) \leq f(b) + c$. If a, b are (f,q)-connected, we have $g(b) - g(a) = f(b) - f(a) \geq c$ whence $g(a) \leq g(b) + c$. If a, b are not (f,q)-connected, we have in particular $|f(a) - f(b)| > q$ and $|g(a) - g(b)| > q$. This implies that if $f(a) < f(b)$ then $g(a) < g(b) - q \leq g(b) - |c| \leq g(b) + c$, so $g(a) \leq g(b) + c$. On the other hand, if $f(b) < f(a)$ then $f(b) + q < f(a)$. This gives $q < f(a) - f(b) \leq c$, a contradiction to $|c| \leq q$. ☐

Lemma 4. *Let S be a subset of $\mathbb{Q}.\mathbb{Z}$ and let $(a_i)_{i\in\mathbb{N}}$ be an enumeration of S. Let $(F_i)_{i\in\mathbb{N}}$ be a sequence of $\sim_s$-equivalence classes of functions from $\{a_0, \dots, a_i\} \to \mathbb{Q}.\mathbb{Z}$, for some $s \in \mathbb{N}$, such that $g \in F_j$ and $i < j$ imply that $g|_{\{a_0,\dots,a_i\}} \in F_i$. Then there exists a function $h\colon S \to \mathbb{Q}.\mathbb{Z}$ such that $h|_{\{a_0,\dots,a_i\}} \in F_i$ for all i and if $x, y \in S$ are not (g,s)-connected for any $g \in \bigcup_i F_i$, then $h(x) - h(y) = \infty$.*

Proof. We first outline the strategy of the proof. We build the function h as a set-theoretic union of functions $h_i\colon \{a_0, \dots, a_i\} \to \mathbb{Q}.\mathbb{Z}$. We force that at each step i, the function h_i is in F_i and satisfies $h_i(a_k) - h_i(a_l) < \infty$ if and only if (a_k, a_l) are (g,s)-connected for some $j \geq i$ and some $g \in F_j$. The technicality of the proof comes from the fact that although we build the functions h_i by induction, we have to look ahead before choosing whether two points have to be mapped to different copies of $\mathbb{Z}$ in $\mathbb{Q}.\mathbb{Z}$ and to which copy they can be mapped.

We define the function h by induction. We require that at each step, the function $h_i\colon \{a_0, \dots, a_i\} \to \mathbb{Q}.\mathbb{Z}$ that we define is in F_i and

- whenever $a, b \in \{a_0, \dots, a_i\}$ are not (g,s)-connected for *any* function g in some F_j, then $h_i(a) - h_i(b) = \infty$, and

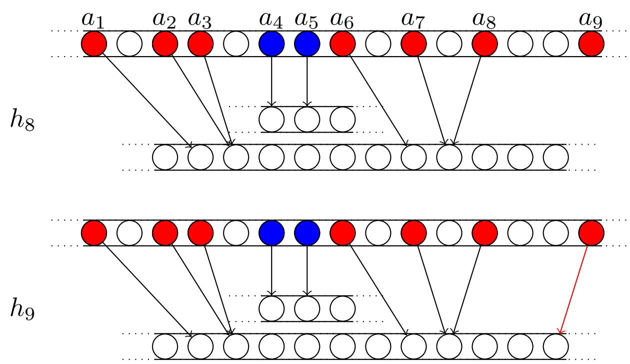

Figure 1: Illustration for item (1) of the proof of Lemma 4. We consider the case $i = 8$. The domain of h_8 is depicted above, and the copies of $\mathbb{Z}$ intersecting the image of h_8 are depicted below. Here, the image of h_8 intersects two copies of $\mathbb{Z}$. The colours represent equivalence classes of (g, s)-connectedness, where g is a function of some F_j that connects all the points of $\{a_1, \ldots, a_9\}$ that eventually become connected. Since a_9 is in the same class as some previous point, we are in case (1) of the construction. Supposing that $g(a_9) = g(a_8) + 3$, we build h_9 by setting $h_9(a_9) = h_9(a_8) + 3$.

- if $a, b \in \{a_0, \ldots, a_i\}$ are (g, s)-connected for *some* function $g \in F_j$, then $h_i(a) - h_i(b) = g(a) - g(b)$.

For $i = 0$, let h_0 be any function in F_0. Suppose now that h_i has been defined, and let $h_{i+1}(a_k) := h_i(a_k)$ for $k \in \{0, \ldots, i\}$. Let $g \in F_j$ be such that for every pair $a_k, a_l \in \{a_0, \ldots, a_{i+1}\}$, if there exist $j' \geq 0$ and $g' \in F_{j'}$ such that (a_k, a_l) are (g', s)-connected, then (a_k, a_l) are (g, s)-connected: such a function exists, by taking j sufficiently large so that $\{a_0, \ldots, a_j\}$ contains all the elements that witness that a_k, a_l are (g', s)-connected for some g'. From the induction hypothesis and the assumptions, we know that $h_i \sim_s g|_{\{a_0,\ldots,a_i\}}$. Define $h_{i+1}(a_{i+1})$ as follows:

1. If there exists $k \in \{0, \ldots, i\}$ such that a_{i+1} and a_k are (g, s)-connected. Define $h_{i+1}(a_{i+1}) := h_i(a_k) - g(a_k) + g(a_{i+1})$. This first case is depicted in Figure 1.

2. Otherwise consider the sets

$$U := \{u \in \mathbb{Q} \mid \exists k \in \{0, \ldots, i\} : g(a_k) < g(a_{i+1}) \text{ and } h_i(a_k) \in u.\mathbb{Z}\}$$

   and

$$V := \{v \in \mathbb{Q} \mid \exists k \in \{0, \ldots, i\} : g(a_{i+1}) < g(a_k) \text{ and } h_i(a_k) \in v.\mathbb{Z}\}.$$

   We have $U < V$. Indeed, let $u \in U$, $v \in V$, and let $k, l \in \{0, \ldots, i\}$ be such that $h_i(a_k) \in u.\mathbb{Z}$ with $g(a_k) < g(a_{i+1})$ and $h_i(a_l) \in v.\mathbb{Z}$ with $g(a_{i+1}) < g(a_l)$. Since a_{i+1} is not (g, s)-connected to some element of $\{a_0, \ldots, a_i\}$, we have that a_k and a_l are not (g, s)-connected. By construction, we therefore have that $h_i(a_k) - h_i(a_l) = \infty$. Since a_k and a_l are not (g, s)-connected and since $g(a_k) < g(a_l)$, we have that $h_i(a_k) < h_i(a_l)$. It follows that $u < v$. Thus, there exists $r \in \mathbb{Q}$ such that $U < r < V$. Define $h_{i+1}(a_{i+1}) := (r, 0)$. The situation is depicted in Figure 2.

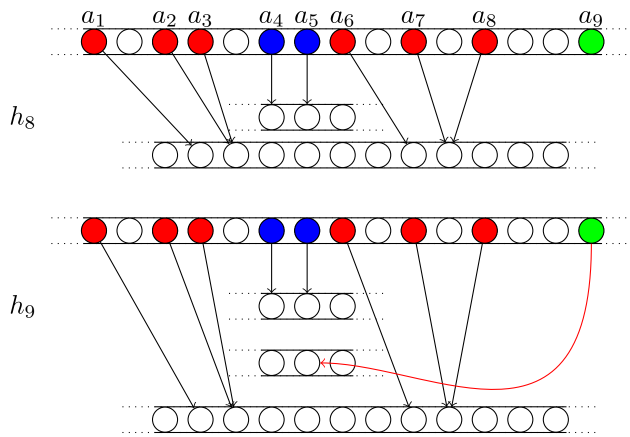

Figure 2: Illustration for item (2) of the proof of Lemma 4. Here, a_9 is not in the same equivalence class as any of the previous points. Assume that $g(a_8) < g(a_9) < g(a_4)$. We then find a copy of $\mathbb{Z}$ between the copies containing $h_8(a_8)$ and $h_8(a_4)$ and not containing any points of the image of h_8. We set $h_9(a_8)$ to be an arbitrary point in this new copy.

We now prove that the induction hypothesis remains true for h_{i+1}. We claim that $h_{i+1} \sim_s g|_{\{a_0,\ldots,a_{i+1}\}}$. Remember that we already know that $h_i \sim_s g|_{\{a_0,\ldots,a_i\}}$ since $h_i \in F_i$ by induction and $g \in F_j$ for $j > i$. Let $a_j \in \{a_0, \ldots, a_i\}$. If $h_{i+1}(a_{i+1})$ is at finite distance from $h_{i+1}(a_j)$, then by definition a_j, a_{i+1} are (g, s)-connected. Let $k \in \{0, \ldots, i\}$ be the index used in the definition of h_{i+1}. We then have

$$
\begin{aligned}
& h_{i+1}(a_{i+1}) - h_{i+1}(a_j) \\
&= h_i(a_k) - g(a_k) + g(a_{i+1}) - h_i(a_j) \\
&= g(a_k) - g(a_j) - g(a_k) + g(a_{i+1}) && \text{(since } h_i(a_k) - h_i(a_j) = g(a_k) - g(a_j)\text{)} \\
&= g(a_{i+1}) - g(a_j).
\end{aligned}
$$

It follows that a_{i+1}, a_j are (h_{i+1}, s)-connected iff they are $(g|_{\{a_0,\ldots,a_{i+1}\}}, s)$-connected. If $h_{i+1}(a_{i+1})$ and $h_{i+1}(a_j)$ are at infinite distance, then a_{i+1}, a_j are neither (h_{i+1}, s)-connected nor (g, s)-connected. Then $h_{i+1}(a_{i+1}) < h_{i+1}(a_j) \Leftrightarrow g(a_{i+1}) < g(a_j)$ from the construction of U and V. It follows that $h_{i+1} \sim_s g|_{\{a_0,\ldots,a_{i+1}\}}$. Moreover, h_{i+1} indeed separates integers that are never (g, s)-connected for any $g \in F_j$. Finally, if $g' \in F_{j'}$ is such that a, b are (g', s)-connected then a and b are also (g, s)-connected and $g'(a) - g'(b) = g(a) - g(b)$. This proves that h_{i+1} satisfies the induction hypothesis. Then $h := \bigcup_{i \geq 0} h_i$ satisfies the conclusion of the statement. □

The two previous lemmas will be applied frequently; one application is in the proof of the following proposition. Note that this makes essential use of the saturated model.

Proposition 3. *Let Γ be a finite-signature first-order reduct of $(\mathbb{Z}; <)$. Then for all $a_1, a_2 \in \mathbb{Z}$ either*

- *there is an $r \geq 0$ and a finite $S \subseteq \mathbb{Z}$ that contains $\{a_1, a_2\}$ such that for all homomorphisms f from $\Gamma[S]$ to Γ we have $|f(a_1) - f(a_2)| \leq r$, or*

- *there is a homomorphism h from Γ to $\mathbb{Q}.\Gamma$ such that $h(a_1) - h(a_2) = \infty$.*

Proof. Let $a_1, a_2 \in \mathbb{Z}$ be arbitrary. Suppose that for all $r \geq 0$ and all finite $S \subset \mathbb{Z}$ containing $\{a_1, a_2\}$ there is a homomorphism f from $\Gamma[S]$ to Γ such that $|f(a_1) - f(a_2)| > r$. We will describe how to construct the desired homomorphism h.

Let $a_1, a_2, a_3, \ldots$ be an enumeration of $\mathbb{Z}$, and let q be the qe-degree of Γ. Consider the following infinite tree $\mathcal{T}$ whose vertices lie on levels $1, 2, \ldots$ The vertices at the n-th level are the $\sim_q$-equivalence classes of homomorphisms f from $\Gamma[\{a_1, \ldots, a_{n+1}\}] \to \mathbb{Q}.\Gamma$ such that a_1, a_2 are not (f, q)-connected (note that by Lemma 3, every element in the equivalence class of such a homomorphism is also a homomorphism). We have an arc in $\mathcal{T}$ from an equivalence class F on level n to an equivalence class G on level $n+1$ if there are $f \in F$, $g \in G$ such that f is the restriction of g. By assumption, $\mathcal{T}$ has vertices on each level n: indeed, at level n it suffices to take an f such that $|f(a_1) - f(a_2)| > qn$, and such an f exists by assumption. The tree $\mathcal{T}$ has finitely many vertices on each level, since the number of $\sim_q$-equivalence classes of homomorphisms from $\Gamma[\{a_1, \ldots, a_n\}] \to \mathbb{Q}.\Gamma$ is finite.

It follows by König's lemma that there is an infinite branch $\mathcal{B}$ of $\mathcal{T}$. By Lemma 4 applied with $S := \mathbb{Z}$ and $\ell := q$ and using the elements of $\mathcal{B}$ for the sequence $(F_i)_{i \in \mathbb{N}}$, there exists a function $h \colon \mathbb{Z} \to \mathbb{Q}.\mathbb{Z}$ such that $h|_{\{a_1, \ldots, a_i\}}$ is in the branch $\mathcal{B}$ for every $i \in \mathbb{N}$, and $h(a_1) - h(a_2) = \infty$ (since a_1, a_2 are not connected by any function in the branch $\mathcal{B}$). Finally, h is a homomorphism $\Gamma \to \mathbb{Q}.\Gamma$ by Lemma 3. $\square$

Definition 6. A mapping $h \colon \kappa_1.\mathbb{Z} \to \kappa_2.\mathbb{Z}$ is called *isometric* if $|h(x) - h(y)| = |x - y|$ for all $x, y \in \kappa_1.\mathbb{Z}$.

The following proposition can be shown by straightforward modifications of the proof of Proposition 3.

Proposition 4. *Let Γ be a finite-signature first-order reduct of $(\mathbb{Z}; <)$. Then either*

- *for every $r \in \mathbb{N}$ there is a finite $S \subseteq \mathbb{Z}$ containing $\{0, r\}$ such that for all homomorphisms f from $\Gamma[S]$ to Γ we have $|f(0) - f(r)| = r$, or*

- *there is a homomorphism h from Γ to $\mathbb{Q}.\Gamma$ which is not isometric.*

5.3 Finite-range Endomorphisms

In this section we present a lemma that gives a useful sufficient condition for Γ to have endomorphisms with finite range. Note that Γ has a finite-range endomorphism if and only if there exists a finite structure Δ such that $\mathrm{CSP}(\Gamma) = \mathrm{CSP}(\Delta)$. We need the following combinatorial definitions and lemmas about the integers.

We say that $T \subseteq \mathbb{Z}$ *contains arbitrarily long intervals* when for every $m \in \mathbb{N}$ there exists $z \in \mathbb{Z}$ so that $[z, z+m] \subset T$. A sequence $u_1, \ldots, u_r$ is called a *$(\leq m)$-progression* if $1 \leq u_{i+1} - u_i \leq m$ for all $i < r$. We say that T has *arbitrarily long $(\leq m)$-progressions* if for every $r \in \mathbb{N}$ the set T contains a $(\leq m)$-progression $u_1, \ldots, u_r$. Clearly, if $\mathbb{Z} \setminus T$ does not have arbitrarily long intervals then there exists an $m \in \mathbb{N}$ so that T has arbitrarily long $(\leq m)$-progressions.

Lemma 5. *Let $T \subseteq \mathbb{Z}$ contain arbitrarily long $(\leq m)$-progressions, and let $T = T_1 \cup \cdots \cup T_k$ be a partition of T into finitely many sets. Then there exists an $i \leq k$ and an $m' \in \mathbb{N}$ such that T_i contains arbitrarily long $(\leq m')$-progressions.*

Proof. If there exists an $m' \in \mathbb{N}$ such that T_1 contains arbitrarily long $(\leq m')$-progressions, then there is nothing to show. So suppose that this is not the case.

We will show that $T' := T \setminus T_1$ contains arbitrarily long $(\leq m)$-progressions; the statement then follows by induction. Let $s \in \mathbb{N}$ be arbitrary. We want to find a $(\leq m)$-progression $u_1, \ldots, u_s$ in T'. By the above assumption, T_1 does not contain arbitrarily long $(\leq ms)$-progressions, and hence there exists an r such that T_1 does not contain a $(\leq ms)$-progression of length r.

Since T contains arbitrarily long $(\leq m)$-progressions, it contains in particular an $(\leq m)$-progression ρ of length msr. Consider the first s elements of ρ. If all those elements are in T' we have found the desired $(\leq m)$-progression of length s, and are done. So suppose otherwise; that is, at least one of those first s elements must be from T_1. We apply the same argument to the next s elements of ρ, and can again assume that at least one of those elements must be from T_1. Continuing like this, we find a subsequence of ρ of elements of T_1 which form a $(\leq ms)$-progression. The length of this subsequence is $msr/ms = r$. But this contradicts our assumption that T_1 does not contain $(\leq ms)$-progression of length r. □

Lemma 6. *Let $m \in \mathbb{N}$ and let $T \subseteq \mathbb{Z}$ be with arbitrarily long $(\leq m)$-progressions. Then for all $S \subset \mathbb{Z}$ of cardinality $m+1$ there are $x_1, x_2 \in S$ and $y_1, y_2 \in T$ such that $x_1 - x_2 = y_1 - y_2$.*

Proof. Let r be greater than $\max(S) - \min(S)$. Then there exists an $(\leq m)$-progression $w_1, \ldots, w_r$ in T. Define $T_i := \{z - w_1 + \min(S) + i \mid z \in T\}$. Then $T_0 \cup \cdots \cup T_{m-1}$ includes the entire interval $[\min(S), \max(S)]$. By the pigeon-hole principle there is an i such that $|T_i \cap S| \geq 2$, which clearly implies the statement. □

Lemma 7. *Let Γ be a finite-signature first-order reduct of $(\mathbb{Z}; <)$ and h a homomorphism from $\Gamma \to \mathbb{Q}.\Gamma$. Let $S \subseteq \mathbb{Z}$ be finite and $z_0 \in \mathbb{Z}$. If $(\mathbb{Z} \setminus h^{-1}(S)) \cap \{z \in \mathbb{Z} : z \geq z_0\}$ does not contain arbitrarily long intervals then Γ has a finite-range endomorphism.*

Proof. Since $(\mathbb{Z} \setminus h^{-1}(S)) \cap \{z \in \mathbb{Z} : z \geq z_0\}$ does not contain arbitrarily long intervals, there exists an $m' \in \mathbb{N}$ such that $T := h^{-1}(S)$ contains arbitrarily long $(\leq m')$-progressions. Suppose that $S = \{s_1, \ldots, s_k\}$, and define $T_i := h^{-1}(s_i)$ for $i \in \{1, \ldots, k\}$. Then by Lemma 5 there exists an $m \in \mathbb{N}$ and an $i \leq k$ such that T_i contains arbitrarily long $(\leq m)$-progressions.

Our argument is based on König's tree lemma, involving the finitely branching infinite tree $\mathcal{T}$ defined similarly as in the proof of Proposition 3. Let $a_0, a_1, \ldots$ be an enumeration of $\mathbb{Z}$, and let q be the qe-degree of Γ. The vertices of $\mathcal{T}$ on the n-th level are the $\sim_q$-equivalence classes of homomorphisms g from $\Gamma[\{a_0, \ldots, a_n\}]$ to Γ such that $|g(\{a_0, \ldots, a_n\})| \leq m$. Adjacency is defined by restriction, and $\mathcal{T}$ is finitely branching, as in the proof of Proposition 3.

We show that $\mathcal{T}$ has vertices on all levels n by induction on n. We prove that for any finite $X \subset \mathbb{Z}$ there exists a homomorphism $g \colon \Gamma[X] \to \Gamma$ whose range has size at most m. For $|X| \leq m$, this is witnessed by the restriction of the identity function to X. Now let $|X| = n+1$ for $n \geq m$. By Lemma 6, there are $x_1, x_2 \in X$ and $y_1, y_2 \in T_i$ such that $x_1 - x_2 = y_1 - y_2$. We therefore have that $f \colon x \mapsto h(x - x_1 + y_1)$ is a homomorphism $\Gamma[X] \to \mathbb{Q}.\Gamma$ whose range has size at most n. Indeed, we have $f(x_1) = h(y_1) = h(y_2) = h(x_2 - x_1 + y_1) = f(x_2)$. Let g be given by the induction hypothesis applied to the image of f. We then have that $g \circ f$ is a homomorphism $\Gamma[X] \to \Gamma$ whose range has size at most m, and the claim is proved.

Hence, $\mathcal{T}$ has vertices on all levels, and therefore an infinite branch $\mathcal{B}$ by König's lemma. By Proposition 4 applied to this infinite branch, $S := \mathbb{Z}$, and $\ell := q$ there exists a function $h \colon \mathbb{Z} \to \mathbb{Q}.\mathbb{Z}$ such that $h|_{\{a_0,\ldots,a_i\}} \in \mathcal{B}$ for all $i \in \mathbb{N}$. In particular, the range of h has size at most m. Up to $\sim_q$-equivalence, we can assume that the image of h lies in one copy of $\mathbb{Z}$ in $\mathbb{Q}.\mathbb{Z}$, say in $\mathbb{Z}$. Then Lemma 3 implies that h is a homomorphism $e \colon \Gamma \to \Gamma$ whose range has cardinality at most m, concluding the proof. □

The next lemma is an important consequence of Lemma 7.

Lemma 8. *Let Γ be a finite-signature first-order reduct of $(\mathbb{Z};<)$ without finite-range endomorphisms, $\ell \in \mathbb{N}$, and h a homomorphism from Γ to $\mathbb{Q}.\Gamma$. Then there exists an $e \in \mathrm{End}(\mathbb{Q}.\Gamma)$ such that for all $x, y \in \mathbb{Q}.\mathbb{Z}$ with $x - y = \infty$ we have $e(x) - e(y) = \infty$, and such that $h \sim_\ell e|_{\mathbb{Z}}$.*

Proof. We first give an idea about the proof. Since Γ does not have finite-range endomorphisms, we know from the previous lemma that the preimage of any finite subset of $\mathbb{Q}.\mathbb{Z}$ under h leaves arbitrarily large gaps in $\mathbb{Z}$. It follows that for every finite subset S of $\mathbb{Q}.\mathbb{Z}$, there exists a homomorphism $p \colon \mathbb{Q}.\Gamma[S] \to \Gamma$ such that $h \circ p$ does not connect any pair of integers that sit in different copies. Since we have such homomorphisms for arbitrarily large finite subsets $S \subset \mathbb{Q}.\mathbb{Z}$, an application of König's lemma and Lemma 4 give the desired endomorphism of $\mathbb{Q}.\Gamma$.

We now give the detailed argument. Note that if $h \sim_{\ell'} g$ and $\ell < \ell'$, then $h \sim_\ell g$. It follows that without loss of generality, we can assume that ℓ is greater than the qe-degree of Γ. As in the proof of Proposition 3, we build e through an argument involving König's lemma and an infinite tree $\mathcal{T}$. Let $a_1, a_2, \ldots$ be an enumeration of $\mathbb{Q}.\mathbb{Z}$. For the n-th level of $\mathcal{T}$ we will consider $\sim_\ell$-classes of homomorphisms f from $\mathbb{Q}.\Gamma[\{a_1, \ldots, a_n\}]$ to $\mathbb{Q}.\Gamma$ with the property that

- for all $x, y \in \{a_1, \ldots, a_n\}$ with $x - y = \infty$ the elements x, y are not (f, ℓ)-connected, and
- $f|_{\{a_1,\ldots,a_n\}} \sim_\ell h|_{\{a_1,\ldots,a_n\}}$.

Adjacency is defined by restriction as in the proof of Proposition 3.

The only difficulty of the proof is to show that $\mathcal{T}$ has vertices on all levels n. We will first construct a homomorphism p from $\mathbb{Q}.\Gamma[\{a_1, \ldots, a_n\}]$ to Γ with the property that $p(a_i) = a_i$ for a_i in the domain of h, and if $a_i - a_j = \infty$ for $i, j \leq n$, then $p(a_i)$ and $p(a_j)$ are not (h, ℓ)-connected. Let S be the set of points that are at distance at most ℓ from some $a_1, \ldots, a_n$. Let $S_1 \cup \cdots \cup S_k$ be the partition of S induced by the copies of $\mathbb{Z}$ in $\mathbb{Q}.\mathbb{Z}$, that is, $S_1, \ldots, S_k$ are pairwise disjoint and each S_i only contains points that lie in the same copy of $\mathbb{Z}$ in $\mathbb{Q}.\mathbb{Z}$. Suppose without loss of generality that $S_1 < \cdots < S_{m-1} < S_m < S_{m+1} < \cdots < S_k$ and that $S_m \subset \mathbb{Z}$, the standard copy in $\mathbb{Q}.\mathbb{Z}$. For every $i \in \{1, \ldots, k\}$, let s_i and t_i be the minimal and the maximal element of S_i, respectively. The situation is represented in Figure 3.

For the elements $x \in S_m$ we set $p(x) := x$. Let $Q_m = \{z \in \mathbb{Q}.\mathbb{Z} \mid \exists z' \in S_m : |h(z') - z| \leq \ell\}$. Write S'_m for $h^{-1}(Q_m)$. If $\mathbb{Z} \setminus S'_m \cap \{z \mid z > t_m\}$ does not contain arbitrarily long intervals, then Γ has a finite-range endomorphism by Lemma 7, contrary to our assumptions. So there exists a $z_m \in \mathbb{Z}$ greater than t_m such that $[z_m, z_m + t_{m+1} - s_{m+1} + 2\ell] \cap S'_m = \emptyset$. For $x \in S_{m+1}$, we set $p(x) := x - s_{m+1} + z_m + \ell$. The mapping is illustrated in Figure 4. As above, set Q_{m+1} to be the set of points that are at distance at most ℓ from a point in $h(p(S_m \cup S_{m+1}))$. Now, set $S'_{m+1} := h^{-1}(Q_{m+1})$. Then there exists a $z_{m+1} \in \mathbb{Z}$ such that $[z_{m+1}, z_{m+1} + t_{m+2} - s_{m+2} + 2\ell] \cap S'_{m+1} = \emptyset$. For $x \in S_{m+2}$, we set $p(x) := x - s_{m+2} + z_{m+1} + \ell$. Continuing in this way, we define p for all $x \in \{a_1, \ldots, a_n\}$ (the construction for $i < m$ is symmetric). We have that p is a homomorphism $\mathbb{Q}.\Gamma[\{a_1, \ldots, a_n\}] \to \Gamma$

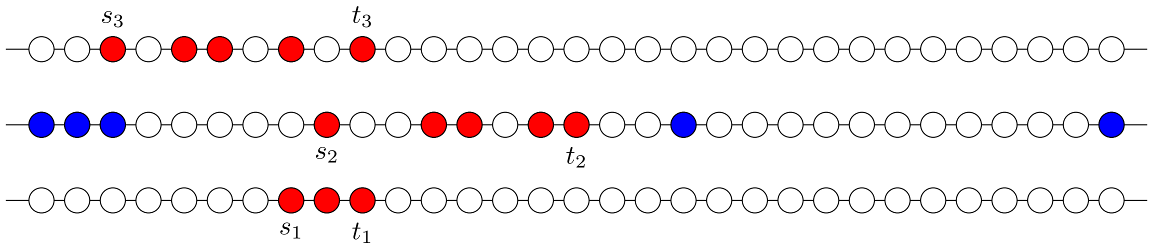

Figure 3: Illustration of the proof of Lemma 8. Here, $k = 3$, $\ell = 1$ and $m = 2$. The nodes coloured in red (light grey) are the integers in S_1, S_2, S_3. The nodes coloured in blue (dark grey) are the integers in $S'_2 \setminus S_2$, that is, the integers that are mapped under h to integers near $h(S_2)$. The assumption that Γ does not have finite-range endomorphisms guarantees that there are arbitrarily long intervals of white nodes in the middle line, both on the left of s_2 and the right of t_2.

since it is $\sim_\ell$-equivalent to the identity function on $\mathbb{Q}.\mathbb{Z}[\{a_1, \ldots, a_n\}]$. Observe that by construction of p, when $a_i - a_j = \infty$, then a_i, a_j are not $(h \circ p, \ell)$-connected. Therefore the $\sim_q$-equivalence class of $h \circ p$ is a vertex of $\mathcal{T}$ on level n.

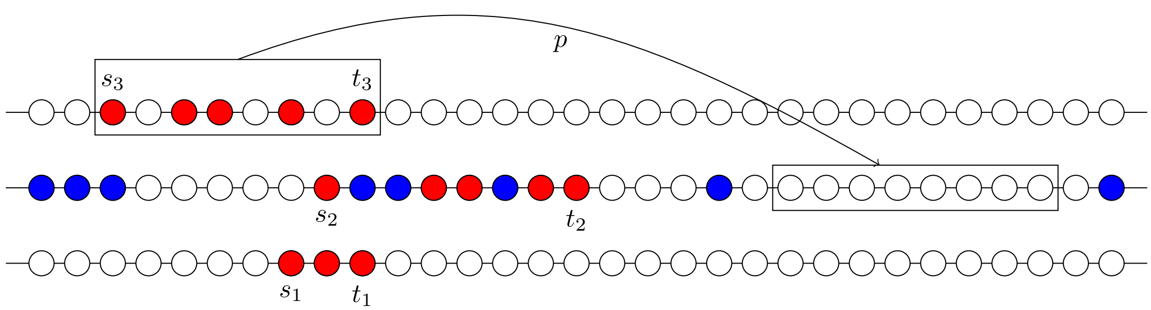

Figure 4: Illustration of the proof of Lemma 8, after the first step of the construction. The blue nodes (light grey) are now the integers in S'_3 that are not in S_2 or in $p(S_3)$, that is, the integers that are mapped by h to integers near $h(S_2 \cup p(S_3))$.

The tree $\mathcal{T}$ is finitely branching, and by König's lemma it contains an infinite branch $\mathcal{B}$. By Lemma 4 applied to this branch, $S := \mathbb{Q}.\mathbb{Z}$, and ℓ as in the statement of Lemma 8 there exists a function $e\colon \mathbb{Q}.\mathbb{Z} \to \mathbb{Q}.\mathbb{Z}$ such that $e|_{\{a_1,\ldots,a_i\}} \in \mathcal{B}$ for all $i \in \mathbb{N}$ and if $x - y = \infty$ then $e(x) - e(y) = \infty$. By Lemma 3, e is an endomorphism of $\mathbb{Q}.\Gamma$. We also have that $e|_{\mathbb{Z}} \sim_\ell h$ and hence e has the required properties. $\square$

5.4 Petrus

The following theorem is the rock upon which we build our church.

Theorem 5 (Petrus ordinis)**.** *Let Γ be a first-order reduct of $(\mathbb{Z};<)$ with finite relational signature and without an endomorphism of finite range. Then the following are equivalent:*

1. *there exists a first-order reduct Δ of $(\mathbb{Q};<)$ such that $\mathrm{CSP}(\Delta)$ equals $\mathrm{CSP}(\Gamma)$;*
2. *for all $t \geq 1$, there is an $e \in \mathrm{End}(\mathbb{Q}.\Gamma)$ and $z \in \mathbb{Q}.\mathbb{Z}$ such that $|e(z+t) - e(z)| > t$;*
3. *all binary relations with a primitive positive definition in $\mathbb{Q}.\Gamma$ are either empty, the equality relation, or have unbounded distance degree;*
4. *for all distinct $z_1, z_2 \in \mathbb{Z}$ there is a homomorphism $h\colon \Gamma \to \mathbb{Q}.\Gamma$ such that $h(z_1) - h(z_2) = \infty$;*
5. *for all distinct $z_1, z_2 \in \mathbb{Z}$ there is an $e \in \mathrm{End}(\mathbb{Q}.\Gamma)$ such that $e(z_1) - e(z_2) = \infty$; and for all $z_1', z_2' \in \mathbb{Q}.\mathbb{Z}$ with $z_1' - z_2' = \infty$ we have $e(z_1') - e(z_2') = \infty$;*
6. *there exists an $e \in \mathrm{End}(\mathbb{Q}.\Gamma)$ with infinite range such that $e(x) - e(y) = \infty$ or $e(x) = e(y)$ for any two distinct $x, y \in \mathbb{Q}.\Gamma$.*

Proof. Throughout the proof, let q be the qe-degree of Γ, which is finite since Γ has a finite signature.

(1) $\Rightarrow$ (2). Since Δ has the same CSP as Γ, and Δ is ω-categorical, Lemma 3.1.5 in [2]) states that there is a homomorphism f from the countable structure $\mathbb{Q}.\Gamma$ to Δ. Lemma 1 asserts the existence of a homomorphism g from Δ to $\mathbb{Q}.\Gamma$, because every pp-sentence that is true in Δ is also true in $\mathbb{Q}.\Gamma$, and $\mathbb{Q}.\Gamma$ is saturated.

Let $t \geq 1$. It is not possible that $f(z) = f(z+t)$ for all $z \in \mathbb{Q}.\mathbb{Z}$, for otherwise Γ would have a finite-range endomorphism. Indeed, we can restrict $g \circ f$ to a homomorphism $\Gamma \to \mathbb{Q}.\Gamma$ whose range is finite. We can then construct a function $e\colon \mathbb{Z} \to \mathbb{Q}.\mathbb{Z}$ such that $g \circ f \sim_q e$ and such that the range of e is contained in $\mathbb{Z}$. This e would then be an endomorphism of Γ by Lemma 3, a contradiction. Pick a $z \in \mathbb{Q}.\mathbb{Z}$ such that $f(z) \neq f(z+t)$. The range of g is infinite, for otherwise the range of $g \circ f$ would be finite. Thus, there are two rationals $p \neq p'$ such that $|g(p) - g(p')| > t$. Let α be an automorphism of Δ that maps $\{f(z), f(z+t)\}$ to $\{p, p'\}$. We now have $|(g \circ \alpha \circ f)(z+t) - (g \circ \alpha \circ f)(z)| = |g(p) - g(p')| > t$.

(2) $\Rightarrow$ (3). Let R be a binary relation with a primitive positive definition in $\mathbb{Q}.\Gamma$. Suppose that R is not empty and is not the equality relation. Let k be the supremum of the integers t such that there exists $(z_1, z_2) \in R$ with $|z_1 - z_2| = t$. Since R is neither empty nor the equality relation, it follows that k is positive. If k is ∞, then R has infinite distance degree. Otherwise let (z_1, z_2) be a pair in R such that $|z_1 - z_2| = k$. Let e be an endomorphism of $\mathbb{Q}.\Gamma$ and z be such that $|e(z+k) - e(z)| > k$. Let α be an automorphism of $\mathbb{Q}.\Gamma$ that maps $\{z_1, z_2\}$ to $\{z, z+k\}$. Then $(e \circ \alpha)(z_1, z_2)$ is in R since R is preserved by the endomorphisms of $\mathbb{Q}.\Gamma$ and by construction $|(e \circ \alpha)(z_1) - (e \circ \alpha)(z_2)| > k$, a contradiction to the choice of k.

(3) $\Rightarrow$ (4). Suppose that (4) does not hold, that is, there are distinct $a_1, a_2 \in \mathbb{Z}$ such that for all homomorphisms h from Γ to $\mathbb{Q}.\Gamma$ we have that $h(a_1) - h(a_2) < \infty$. Then by Proposition 3 there is an $r \geq 0$ and a finite $S \subseteq \mathbb{Z}$ containing $\{a_1, a_2\}$ such that for all homomorphisms $f\colon \Gamma[S] \to \Gamma$ we have $|f(a_1) - f(a_2)| \leq r$. Now consider the following primitive positive formula φ: the variables of φ are the elements of S, all existentially quantified except a_1 and a_2, which are free. The formula φ contains the conjunct $R(x_1, \ldots, x_n)$ for a relation R from Γ if and only if $\Gamma[S] \models R(x_1, \ldots, x_n)$. Then φ defines a binary relation, which has bounded distance degree by the previous discussion, and which is not the equality relation since it contains the pair (a_1, a_2).

(4) $\Rightarrow$ (5). Let $z_1, z_2 \in \mathbb{Z}$ be distinct, let h be given by item (4), and let e be given by Lemma 8 applied to h for $\ell := q$. Pick any function $p\colon e(\mathbb{Q}.\mathbb{Z}) \to \mathbb{Q}.\mathbb{Z}$ such that if $x, y \in \mathbb{Q}.\mathbb{Z}$ are not (e, q)-connected then $(p \circ e)(x) - (p \circ e)(y) = \infty$ and such that $p \sim_q id$. It is clear that such a function

exists because $(\mathbb{Q}; <)$ embeds all countable linear orders. Indeed, consider the equivalence relation on $e(\mathbb{Q}.\mathbb{Z})$ where $x \sim y$ if there are $x := u_1, \dots, u_k =: y \in e(\mathbb{Q}.\mathbb{Z})$ such that $|u_i - u_{i+1}| \le q$ for all $i \in \{1, \dots, k-1\}$. The equivalence classes induced by this relation are naturally ordered by setting $\rho < \pi$ if for all $x \in \rho, y \in \pi$, we have $x < y$. There are at most countably many equivalence classes, hence there exists an increasing function f from the set of equivalence classes to $\mathbb{Q}$. We let $p(a, z) := (f(\rho), z)$ where ρ is the equivalence class of (a, z). Then we have that $p \sim_q id$, and this implies that $p \circ e \sim_q e$ so that $p \circ e$ is an endomorphism of $\mathbb{Q}.\Gamma$ by the substitution lemma. Moreover, p is such that $x - y = \infty \Rightarrow (p \circ e)(x) - (p \circ e)(y) = \infty$. Finally, z_1 and z_2 are not (e, q)-connected because $e|_{\mathbb{Z}} \sim_q h$, so that $(p \circ e)(z_1) - (p \circ e)(z_2) = \infty$.

(5) $\Rightarrow$ (6). Again an argument based on König's tree lemma. Let $a_1, a_2, \dots$ be an enumeration of $\mathbb{Q}.\mathbb{Z}$. Let $\mathcal{T}$ be a tree whose vertices on the n-th level are the $\sim_q$-equivalence classes of homomorphisms g from $\mathbb{Q}.\Gamma[\{a_1, \dots, a_n\}]$ to $\mathbb{Q}.\Gamma$ such that for all $i, j \le n$ either a_i and a_j are not (g, q)-connected or $g(a_i) = g(a_j)$. Adjacency of vertices is defined by restriction between representatives. We have to show that the tree has vertices on all levels. Let $\{u_1, v_1\}, \dots, \{u_k, v_k\}$ be an enumeration of all 2-element subsets of $\{a_1, \dots, a_n\}$. We will show by induction on $i \ge 0$ that there exists an endomorphism f_i such that $f_j(u_j) - f_j(v_j) = \infty$ or $f_j(u_j) = f_j(v_j)$ for all $j \le i$. The statement is trivial for $i = 0$. So suppose we have already found f_i for some $i \ge 0$, and want to find f_{i+1}. If $f_i(u_{i+1}) - f_i(v_{i+1}) = \infty$ or $f_i(u_{i+1}) = f_i(v_{i+1})$ then there is nothing to show. Otherwise, let α be an automorphism of $\mathbb{Q}.\Gamma$ that maps $f_i(u_{i+1})$ and $f_i(v_{i+1})$ to $\mathbb{Z}$. By (5), there exists an $e \in \mathrm{End}(\mathbb{Q}.\Gamma)$ such that $e(\alpha(f_i(u_{i+1}))) - e(\alpha(f_i(v_{i+1}))) = \infty$, and such that for all $x, y \in \mathbb{Q}.\mathbb{Z}$ with $x - y = \infty$ we have that $e(x) - e(y) = \infty$. Hence, $f_{i+1} := e \circ \alpha \circ f_i$ has the desired property. The tree $\mathcal{T}$ has finitely many vertices on each level and hence must contain an infinite branch, which gives rise to an endomorphism of $\mathbb{Q}.\Gamma$ by Lemmas 4 and 3.

(6) $\Rightarrow$ (1). Let Δ be the structure induced by $\mathbb{Q}.\Gamma$ on the image of the endomorphism e whose existence has been asserted in (6). The structures Δ and Γ have the same CSP. Note that a literal $x \le y + k$ for $k \in \mathbb{Z}$ is true in Δ iff $x \le y$ is true. Therefore the relations of Δ are definable with quantifier-free formulas using only $x < y$ and $x = y$. It follows that Δ has the same CSP as a first-order reduct of $(\mathbb{Q}; <)$. □

5.5 Boundedness and Rank

Let Γ be a finite-signature first-order reduct of $(\mathbb{Z}; <)$ without a finite-range endomorphism. Theorem 5 (Petrus) characterized the "degenerate case" when CSP(Γ) is the CSP for a first-order reduct of $(\mathbb{Q}; <)$. For such Γ, as we have mentioned before, the complexity of the CSP has already been classified. In the following we will therefore assume that the equivalent items of Theorem 5, and in particular, item (2), do *not* apply. To make the best use of those findings, we introduce the following terminology.

Definition 7. Let $k \in \mathbb{N}^+, c \in \mathbb{N}$. A function $e \colon \kappa_1.\mathbb{Z} \to \kappa_2.\mathbb{Z}$ is (k, c)-*bounded* if for all $u \in \kappa_1.\mathbb{Z}$ we have $|e(u + k) - e(u)| \le c$.

We say that e is *tightly-k-bounded* if it is (k, k)-bounded, and *k-bounded* if it is (k, c)-bounded for some $c \in \mathbb{N}$. For given k, c, we say that $\kappa.\Gamma$ is (k, c)-*bounded* (resp. *k-bounded*, tightly-k-bounded) if all its endomorphisms are. We call the smallest t such that $\kappa.\Gamma$ is tightly-t-bounded the *tight rank* of $\kappa.\Gamma$. Similarly, we call the smallest r such that $\kappa.\Gamma$ is r-bounded the *rank* of $\kappa.\Gamma$.

The negation of item (2) in Theorem 5 says that there exists a $t \in \mathbb{N}$ such that $\mathbb{Q}.\Gamma$ is tightly-t-bounded. Clearly, being tightly-t-bounded implies being t-bounded. Hence, the negation of item (2) in Theorem 5 also implies that $\mathbb{Q}.\Gamma$ has finite rank $r \le t$.

Example 4. For $p > 0$, the structure $(\mathbb{Z}; succ^p)$ of Example 1 (1) has rank and tight rank equal to p. The structure $(\mathbb{Z}; succ^2, \mathrm{Diff}_{\{-2,-1,0,1,2\}})$ of Example 1 (3) is an example whose rank is 1 and whose tight rank is greater (it is equal to 2).

Sections 5.5.1 and 5.5.2 are devoted to proving that one can replace Γ by another first-order reduct Δ of $(\mathbb{Z}; <)$ which has the same CSP and such that $\mathbb{Q}.\Delta$ has both rank one and tight rank one.

Example 5. There are rank one first-order reducts of $(\mathbb{Z}; <)$ which do have non-injective endomorphisms, but no finite-range endomorphisms. Consider the third structure in Example 1:

$$\Gamma := (\mathbb{Z}; succ^2, \mathrm{Diff}_{\{-2,-1,0,1,2\}}) .$$

Note that Γ has rank one: as every endomorphism e preserves the relation $\mathrm{Diff}_{\{-2,-1,0,1,2\}}$ we have $|e(x+1) - e(x)| \leq 2$. Also note that Γ has the non-injective endomorphism e defined by $e(x) = x$ for even x, and $e(x) = x + 1$ for odd x.

Corollary 1. *Let Γ be a finite-signature reduct of $(\mathbb{Z}; <)$ without finite-range endomorphisms. Then $\mathbb{Q}.\Gamma$ has finite rank if and only if $\mathbb{Q}.\Gamma$ has finite tight rank.*

Proof. We have just seen that having finite tight rank implies having finite rank. Conversely, when $\mathbb{Q}.\Gamma$ has finite rank, then item (5) in Theorem 5 is false. Then Theorem 5 implies that item (2) is false, too, which is to say that $\mathbb{Q}.\Gamma$ has finite tight rank. □

We also make the following important observation.

Lemma 9. *Let Γ be a finite-signature reduct of $(\mathbb{Z}; <)$ without finite-range endomorphisms and such that $\mathbb{Q}.\Gamma$ has finite rank r. Then there exists a $c \geq 0$ such that every $e \in \mathrm{End}(\Gamma)$ is (r, c)-bounded.*

Proof. Let $a_1 < a_2$ be two integers at distance r. We know from the negation of item (4) in Theorem 5 that every homomorphism $h \colon \Gamma \to \mathbb{Q}.\Gamma$ satisfies $h(a_1) - h(a_2) < \infty$. Proposition 3 gives a $c \geq 0$ and a finite $S \subset \mathbb{Z}$ containing a_1, a_2 such that every homomorphism $f \colon \Gamma[S] \to \Gamma$ satisfies $|f(a_1) - f(a_2)| \leq c$. In particular, every endomorphism f of Γ also satisfies this.

To prove that every endomorphism of Γ is (r, c)-bounded, let now $f \in \mathrm{End}(\Gamma)$ and $a \in \mathbb{Z}$. Let α be the automorphism of $(\mathbb{Z}; <)$ that maps a_1 to a. By the paragraph above applied to the endomorphism $f \circ \alpha$, we have $|(f \circ \alpha)(a_1) - (f \circ \alpha)(a_2)| \leq c$, i.e., $|f(a) - f(a + r)| \leq c$. This proves that f is (r, c)-bounded. □

The next lemma connects the rank and the tight rank of a structure and its countable saturated extension.

Lemma 10. *Let Γ be a first-order reduct of $(\mathbb{Z}; <)$ with finite relational signature such that $\mathbb{Q}.\Gamma$ has rank r and tight rank t. Then Γ has rank $r' \leq r$ and tight rank $t' \leq t$.*

Proof. Let f be an endomorphism of Γ, and let $a \in \mathbb{Z}$. Let $\ell = \max\{|f(a + r) - f(a)|, q\}$. We view f as a homomorphism $\Gamma \to \mathbb{Q}.\Gamma$ and find an endomorphism e of $\mathbb{Q}.\Gamma$ such that $e|_{\mathbb{Z}} \sim_\ell f$ by Lemma 8. There exists a $c > 0$ such that the endomorphism e is (r, c)-bounded, by assumption on $\mathbb{Q}.\Gamma$. This gives $|f(a + r) - f(a)| = |e(a + r) - e(a)| \leq c$, i.e., f is (r, c)-bounded. Therefore, every endomorphism of Γ is r-bounded and Γ has finite rank $r' \leq r$. We prove similarly that every endomorphism of Γ is tightly-t-bounded, which implies that Γ has finite tight rank $t' \leq t$. □

5.5.1 The Rank One Case

The main result of this section, Theorem 7, implies that for each rank one first-order reduct Γ of $(\mathbb{Z};<)$ without finite range endomorphisms there exists a first-order reduct Δ of $(\mathbb{Z};<)$ which has the same CSP as Γ and where *succ* is pp-definable, or for all $k \geq 1$ the relation Dist_k is pp-definable. By Theorem 2, it suffices to show that the endomorphisms of $\mathbb{Q}.\Delta$ preserve *succ*, or that the endomorphisms of $\mathbb{Q}.\Delta$ preserve Dist_k and Dist_k is 1-generated under $\mathrm{End}(\mathbb{Q}.\Delta)$. The endomorphisms of Γ are better behaved than the endomorphisms of $\mathbb{Q}.\Gamma$, as the latter endomorphisms can exhibit different behaviours in each copy of $\mathbb{Z}$, and can collapse copies, whereas the former endomorphisms are more uniform, as we will show below. Theorem 6 is the first milestone in our strategy, as it allows us to replace Γ with a first-order reduct Δ of $(\mathbb{Z};<)$ such that $\mathbb{Q}.\Delta$ has tight rank one.

Lemma 11. *Let $e\colon \mathbb{Z} \to \mathbb{Z}$ be tightly-t-bounded and $(1,c)$-bounded for some $c,t \in \mathbb{N}$. Then for all $n \in \mathbb{N}$, and $z \in \mathbb{Z}$, $|e(z+n) - e(z)| \leq n + ct$.*

Proof. Let $n = pt + k$ for $0 \leq k < t$. We have $|e(z+pt+k) - e(z+pt)| \leq kc$ by k applications of $(1,c)$-boundedness, and $|e(z+pt) - e(z)| \leq pt$ by p applications of tight rank t. We obtain

$$\begin{aligned}|e(z+n) - e(z)| &\leq |e(z+pt+k) - e(z+pt)| + |e(z+pt) - e(z)| \\ &\leq kc + pt = n + c(k-1) \leq n + ct\end{aligned}$$

by the triangle inequality. □

The following can be shown by the same proof as the proof of Lemma 6 in [3]; since our statement is more general, and since we use rank and tight rank instead of bounded distance degree, we still give the proof here for the convenience of the reader.

Lemma 12. *Let $e\colon \mathbb{Z} \to \mathbb{Z}$ be tightly-t-bounded and $(1,c)$-bounded. Then either $\{e\} \cup \mathrm{Aut}(\mathbb{Z};<)$ locally generates a function with finite range, or there exists $k > ct+1$ such that for all $x,y \in \mathbb{Z}$ with $|x-y| = k$ we have $|e(x) - e(y)| \geq k$.*

Proof. Assume for all $k > ct+1$ there are $x,y \in \mathbb{Z}$ with $|x-y| = k$ and $|e(x)-e(y)| < k$. We will prove that e locally generates a function with range of size at most $2ct+1$. We again use an argument based on König's tree lemma, albeit with a different flavour than in the previous proofs. Enumerate $\mathbb{Z}$ as $a_1, a_2, \ldots$. The vertices of the tree on level n are the functions $h\colon \{a_1,\ldots,a_n\} \to \mathbb{Z}$ generated by $\{e\} \cup \mathrm{Aut}(\mathbb{Z};<)$ such that the diameter of the image of h is bounded above by $2ct+1$ and such that $h(a_1) = 0$. The edges of the tree between the levels n and $n+1$ are defined by function restriction. The condition on the diameter of the image of h implies that the tree is finitely branching, and we now prove that the tree is infinite.

Let $A \subseteq \mathbb{Z}$ be a finite set. Enumerate the pairs $(x,y) \in A^2$ with $x < y$ by $(x_1,y_1),\ldots,(x_r,y_r)$. Let m be the smallest number with the property that $\mathcal{F} := \{e\} \cup \mathrm{Aut}(\mathbb{Z};<)$ generates a function h_1 such that $|h_1(x_1) - h_1(y_1)| = m$. We claim that $m \leq ct+1$. Otherwise, by assumption there are $x,y \in \mathbb{Z}$ with $|x-y| = m$ and $|e(x)-e(y)| < m$. Let a be the automorphism of $(\mathbb{Z};<)$ such that $a(\{h_1(x),h_1(y)\}) = \{x_1,y_1\}$. Then $\mathcal{F}$ also generates $h_1' := e \circ a \circ h_1$, but $|h_1'(x_1) - h_1'(y_1)| < m$ in contradiction to the choice of m. We conclude that $\mathcal{F}$ generates a function h_1 such that $|h_1(x_1) - h_1(y_1)| \leq ct+1$.

Similarly, there exists h_2 generated by $\mathcal{F}$ such that $|h_2(h_1(x_2)) - h_2(h_1(y_2))| \leq ct+1$. Continuing like this we arrive at a function h_r generated by $\mathcal{F}$ such that

$$|h_r h_{r-1} \cdots h_1(x_r) - h_r h_{r-1} \cdots h_1(y_r)| \leq ct+1.$$

Now consider $h := h_r \circ \cdots \circ h_1$. Set $f_j := h_r \circ \cdots \circ h_{j+1}$ and $g_j := h_j \circ \cdots \circ h_1$, for all $1 \leq j \leq r$; so $h = f_j \circ g_j$. Then, since by construction $|g_j(x_j) - g_j(y_j)| \leq ct + 1$, we have that for all $j \in \mathbb{Z}$ with $1 \leq j \leq r$,

$$\begin{aligned} |h(x_j) - h(y_j)| &= |f_j(g_j(x_j)) - f_j(g_j(y_j))| \\ &\leq |g_j(x_j) - g_j(y_j)| + ct && \text{(by Lemma 11)} \\ &\leq 2ct + 1\,, \end{aligned}$$

and our claim follows. ☐

Definition 8. For $e\colon \kappa_1.\mathbb{Z} \to \kappa_2.\mathbb{Z}$, we call $s \in \mathbb{N}^+$ *stable for* e if for every $p \in \kappa_1$, one of the following applies:

- $e(z + s) = e(z) + s$ for all $z \in p.\mathbb{Z}$,
- $e(z + s) = e(z) - s$ for all $z \in p.\mathbb{Z}$.

Note that if a function e has a stable number, it does not generate a function with finite range. Indeed, it follows from the definition that for all $k \in \mathbb{Z}$ we have $|e(z + kt) - e(z)| = kt$.

Lemma 13. *Let* $e\colon \mathbb{Z} \to \mathbb{Z}$ *be tightly-*t*-bounded and* 1*-bounded. Then* t *is stable for* e*, or* $\{e\} \cup \mathrm{Aut}(\mathbb{Z}; <)$ *locally generates a function with finite range.*

Proof. Let $c \in \mathbb{N}$ be such that e is $(1, c)$-bounded, and assume that e does not locally generate a function with finite range. By Lemma 12, there exists $k > ct + 1$ such that for all z we have $|e(z + k) - e(z)| \geq k$, and hence either $e(z + k) \geq e(z) + k$ or $e(z + k) \leq e(z) - k$ for each $z \in \mathbb{Z}$. We will first show that either $e(z + k) \geq e(z) + k$ for all $z \in \mathbb{Z}$, or $e(z + k) \leq e(k) - k$ for all $z \in \mathbb{Z}$. Suppose otherwise that there are $z_1, z_2 \in \mathbb{Z}$ such that $e(z_1 + k) \geq e(z_1) + k$ and $e(z_2 + k) \leq e(z_2) - k$. Clearly, we can choose z_1, z_2 such that $|z_1 - z_2| = 1$. We only treat the case that $z_2 = z_1 + 1$, since the other case is symmetric. Then

$$\begin{aligned} e(z_2) - e(z_2 + k) &\geq k && \text{by assumption,} \\ -e(z_2) + e(z_1) &\geq -c && \text{by 1-boundedness,} \\ e(z_2 + k) - e(z_1 + k) &\geq -c && \text{by 1-boundedness,} \\ e(z_1 + k) - e(z_1) &\geq k && \text{by assumption.} \end{aligned}$$

Summing over those inequalities yields $0 \geq 2k - 2c$, a contradiction since $k > c$.

In the following we assume without loss of generality that $e(z + k) \geq e(z) + k$ for all $z \in \mathbb{Z}$. Recall that $|e(z + t) - e(z)| \leq t$ for all $z \in \mathbb{Z}$ because e is tightly-t-bounded. We next claim that $e(z + kt) = e(z) + kt$ for all $z \in \mathbb{Z}$. Since points at distance t cannot be mapped to points at larger distance, we get that $e(z + kt) - e(z) \leq kt$. On the other hand, since $e(z + k) \geq e(z) + k$ for all $z \in \mathbb{Z}$, we obtain that $e(z + kt) \geq e(z) + kt$, proving the claim.

We now show that $e(z + t) \geq e(z) + t$ for all $z \in \mathbb{Z}$. Note that

$$\begin{aligned} e(z) + kt &= e(z + kt) \\ &= e(z + t + (k - 1)t) \\ &\leq e(z + t) + (k - 1)t \end{aligned}$$

the latter inequality holding since $e(z+mt)-e(z) \leq mt$ for each $m \in \mathbb{N}$. Subtracting $(k-1)t+e(z)$ on both sides, our claim follows. Since $|e(z+t)-e(z)| \leq t$ for all $z \in \mathbb{Z}$, we obtain that $e(z+t) = e(z)+t$ and have proved the lemma. ☐

Corollary 2. *Let Γ be a finite-signature reduct of $(\mathbb{Z};<)$ without finite range endomorphism such that $\mathbb{Q}.\Gamma$ has rank one. Then Γ has finite tight rank t and t is stable for every $e \in \mathrm{End}(\mathbb{Q}.\Gamma)$.*

Proof. By Corollary 1, $\mathbb{Q}.\Gamma$ has finite tight rank t', and by Lemma 10, Γ has tight rank $t \leq t'$ and rank one. Let $e \in \mathrm{End}(\mathbb{Q}.\Gamma)$. Since $\mathbb{Q}.\Gamma$ has rank one, we have $e(z+k) - e(z) < \infty$ for all $z \in \mathbb{Q}.\mathbb{Z}$ and $k \in \mathbb{Z}$. As a consequence, for any $p \in \mathbb{Q}$, the function e induces an endomorphism $e' \colon \Gamma \to \Gamma$ by restricting e to $p.\mathbb{Z}$. By Lemma 13, t is stable for e', and we conclude that t is stable for e. $\square$

Lemma 14. *Every stable number of a function $e \colon \mathbb{Z} \to \mathbb{Z}$ is divisible by the smallest stable number of e.*

Proof. Suppose that p is stable but not divisible by s. Write $p = ms + r$ where m, r are positive integers and $0 < r < s$. Since r is not stable there exists $z \in \mathbb{Z}$ such that $e(z+r) - e(z) \neq r$. But this is impossible since

$$\begin{aligned} e(z+r) - e(z) &= e(z+p-ms) - e(z) \\ &= e(z-ms) + p - e(z) \\ &= e(z) - ms + p - e(z) = r\ . \end{aligned}$$

$\square$

Lemma 15. *Let Γ be a finite-signature reduct of $(\mathbb{Z};<)$ without finite-range endomorphisms and such that $\mathbb{Q}.\Gamma$ has rank one. Let e be an endomorphism of Γ, and let s be the smallest stable number for e. Then $\{e\} \cup \mathrm{Aut}(\mathbb{Z};<)$ generates a function f such that $f(\mathbb{Z}) = \{s \cdot z : z \in \mathbb{Z}\}$.*

Proof. We prove by induction on $i \in \{0, \ldots, s-1\}$ that there exists a function f_i, generated by $\{e\} \cup \mathrm{Aut}(\mathbb{Z};<)$, such that $f_i(j) \in \{s \cdot z : z \in \mathbb{Z}\}$ for all $j \in \{0, \ldots, i\}$, and $f_i(0) = 0$. Without loss of generality, assume that $e(0) = 0$. The base case $i = 0$ is trivial: the identity function on $\mathbb{Z}$ satisfies the requirements. Let f_{i-1} be given. If $f_{i-1}(i)$ is a multiple of s there is nothing to do. Otherwise, $f_{i-1}(i)$ is not stable for e by Lemma 14. Since e has a stable number, it does not generate a function with finite range, so by Lemma 13 it is not tightly-$f_{i-1}(i)$-bounded. It follows that there exist $x_0, y_0 \in \mathbb{Z}$ with $x_0 - y_0 = f_{i-1}(i)$ and $|e(x_0) - e(y_0)| > |f_{i-1}(i)|$. Write $r_1 := |e(x_0) - e(y_0)|$. If r_1 is a multiple of s, then we are done: let α_0 be the automorphism of $(\mathbb{Z};<)$ that maps $\{0, f_{i-1}(i)\}$ to $\{x_0, y_0\}$, let β be the automorphism of $(\mathbb{Z};<)$ that maps $(e \circ \alpha_0 \circ f_{i-1})(0)$ to 0, and let $f_i = \beta \circ e \circ \alpha_0 \circ f_{i-1}$. Since s is stable for e and α_0, we have that $f_i(j) \in \{s \cdot z : z \in \mathbb{Z}\}$ for $j \in \{0, \ldots, i-1\}$ and $|f_i(i)| = |e(y_0) - e(x_0)|$ is a multiple of s by hypothesis. Otherwise, using again Lemma 13 and Lemma 14, we know that e is not tightly-r_1-bounded. Therefore there exist $x_1, y_1 \in \mathbb{Z}$ with $|x_1 - y_1| = r_1$ and $|e(x_1) - e(y_1)| =: r_2 > r_1$. Continuing this way, we obtain a sequence of pairs $(x_0, y_0), (x_1, y_1), \ldots$ such that $r_j = |x_j - y_j|$, and $r_{j+1} > r_j$. Up to exchanging x_j and y_j, we can assume that $e(x_j) < e(y_j)$ iff $x_{j+1} < y_{j+1}$. Since $\mathbb{Q}.\Gamma$ has rank one, Lemma 9 gives a $c \geq 0$ such that every endomorphism of Γ is $(1, c)$-bounded. This implies that the sequence built above must stop in at most c steps. By construction, this can only happen when r_k is a multiple of s. For $j \in \{1, \ldots, k-1\}$, set α_j an automorphism of $(\mathbb{Z};<)$ such that $\alpha_{j+1}(e(x_j)) = x_{j+1}$ and $\alpha_{j+1}(e(y_j)) = y_{j+1}$. Let β be the translation that maps x_k to 0. Finally, set $f_i := \beta \circ \alpha_k \circ e \circ \alpha_{k-1} \circ e \circ \cdots \circ \alpha_1 \circ e \circ \alpha_0 \circ f_{i-1}$. Since s is stable for e and automorphisms of $(\mathbb{Z};<)$, we have that $f_i(j)$ is a multiple of s for every $j \in \{0, \ldots, i-1\}$. Finally we have $f_i(i) = y_k - x_k$ which is a multiple of s by construction. This finishes the inductive proof.

The function f whose existence is claimed in the statement is then f_{s-1}. Indeed, s is stable for f as f is obtained as the composition of e and automorphisms of $(\mathbb{Z};<)$. Therefore $f(\mathbb{Z})$ contains the set $\{s \cdot z : z \in \mathbb{Z}\}$. For the other inclusion, let $v \in \mathbb{Z}$ be arbitrary, and write $v = s \cdot z + r$, where $z \in \mathbb{Z}$ and $0 \leq r < s$. Then $f(s \cdot z + r) - f(r)$ is a multiple of s since s is stable for f. By construction, $f(r)$ is a multiple of s as well, so that $f(v) \in \{s \cdot z \mid z \in \mathbb{Z}\}$. ☐

The following definition arises naturally from the statement of Lemma 15.

Definition 9. Let Γ be a structure over $\mathbb{Z}$ and let $k \in \mathbb{N}^+$. Then we write Γ/k for the substructure of Γ induced by the set $\{z \in \mathbb{Z} : z = 0 \bmod k\}$.

Lemma 16. *For all first-order reducts Γ of $(\mathbb{Z};<)$ and $k \in \mathbb{N}^+$, the structure Γ/k is isomorphic to a first-order reduct of $(\mathbb{Z};<)$, the isomorphism being the function $x \mapsto x/k$.*

Proof. Let R be an n-ary relation of Γ, and let φ be a quantifier-free formula defining R. Construct a formula φ' as follows: For all $i \in \mathbb{Z}$, replace every atomic formula of the form $x \leq y + i$ by $x \leq y + \lfloor i/k \rfloor$. We prove by structural induction on φ that for all $z_1, \ldots, z_n \in \Gamma/k$ we have $(\mathbb{Z},<) \models \varphi(z_1, \ldots, z_n) \Leftrightarrow (\mathbb{Z};<) \models \varphi'(z_1/k, \ldots, z_n/k)$. If φ is $x \leq y + i$ for some $i \in \mathbb{Z}$, then $\Gamma/k \models \varphi(x,y)$ iff $x \leq y + i$ iff $x/k \leq y/k + \lfloor i/k \rfloor$. The cases of conjunction, disjunction, and negation follow immediately from the induction hypothesis. ☐

For instance, in Example 5 the structure $\Gamma/2$ is isomorphic to $(\mathbb{Z}; succ, \{(x,y) : |x-y| \leq 1\})$ which has tight rank one.

Theorem 6. *Let Γ be a finite-signature first-order reduct of $(\mathbb{Z};<)$ without finite range endomorphisms and such that $\mathbb{Q}.\Gamma$ has rank one. Then Γ has an endomorphism that maps Γ to Γ/k for some $k \in \mathbb{N}^+$, which is isomorphic to a reduct Δ of $(\mathbb{Z};<)$ such that $\mathbb{Q}.\Delta$ has tight rank one.*

Proof. Let t be the tight rank of $\mathbb{Q}.\Gamma$, and let c be such that $\mathbb{Q}.\Gamma$ is $(1,c)$-bounded (which exists by Corollary 1). By Lemma 10, Γ has tight rank t', with $t' \leq t$. By Corollary 2, every endomorphism of $\mathbb{Q}.\Gamma$ has a stable number, and in particular each endomorphism has a minimal one. If the minimal stable number of every endomorphism is 1, then $\mathbb{Q}.\Gamma$ has tight rank one and we are done, choosing $k = 1$. Otherwise there exists an $e \in \mathrm{End}(\mathbb{Q}.\Gamma)$ such that 1 is not stable for e. So there exists a copy of $\mathbb{Z}$ and some integer $s > 1$ such that s is stable for the restriction of e to that copy, which we call $\hat{e}$, and so that no s' with $s' < s$ is stable for $\hat{e}$. Since $\mathbb{Q}.\Gamma$ has rank one, e sends copies of $\mathbb{Z}$ to copies of $\mathbb{Z}$. By composing $\hat{e}$ with an automorphism of $(\mathbb{Q}.\mathbb{Z};<)$ we can assume that $\hat{e} \in \mathrm{End}(\Gamma)$. By Lemma 15, there exists a function f generated by $\{\hat{e}\} \cup \mathrm{Aut}(\mathbb{Z};<)$ such that $f(\mathbb{Z}) = \{s \cdot z \mid z \in \mathbb{Z}\}$. By Lemma 13, t' is stable for f, and t' is divisible by s since $|f(z+t') - f(z)| = t'$ and $f(z+t'), f(z) \in \{s \cdot z \mid z \in \mathbb{Z}\}$. Also note that s is stable for f since f is generated by $\hat{e}$.

Observe that Γ/s cannot have a finite range endomorphism: if g were such an endomorphism, then $g \circ f$ would be a finite range endomorphism for Γ, contrary to our assumption. By Lemma 16, Γ/s is isomorphic to a first-order reduct Δ of $(\mathbb{Z};<)$ via the function $x \mapsto x/s$. It is also clear that the function $(a,z) \mapsto (a, sz)$ from $\mathbb{Q}.\mathbb{Z}$ to $\mathbb{Q}.\mathbb{Z}$ is a homomorphism from $\mathbb{Q}.\Delta$ to $\mathbb{Q}.\Gamma$. We claim that $\mathbb{Q}.\Delta$ has rank one and tight rank at most t'/s.

Let $e \in \mathrm{End}(\mathbb{Q}.\Delta)$. Let $x \in \mathbb{Z} \subset \mathbb{Q}.\mathbb{Z}$. Define $e'(z) = s \cdot e(f(z)/s)$ which is a homomorphism $\Gamma \to \mathbb{Q}.\Gamma$. Note that every homomorphism from Γ to $\mathbb{Q}.\Gamma$ is $(1,c)$-bounded, since otherwise by Lemma 8 we can find an endomorphism of $\mathbb{Q}.\Gamma$ which is not $(1,c)$-bounded. Since f is surjective as a function $\mathbb{Z} \to \{s \cdot z \mid z \in \mathbb{Z}\}$, there exists $y \in \mathbb{Z}$ such that $f(y)/s = x$. Since s is stable for f, we have either $f(y+s) = f(y) + s$ or $f(y-s) = f(y) + s$. If $f(y+s) = s + f(y)$, then $e(\frac{f(y)+s}{s}) = \frac{1}{s} \cdot e'(y+s)$. In the other case, $e(\frac{f(y)+s}{s}) = \frac{1}{s} \cdot e'(y-s)$. In any case, by applying s times the $(1,c)$-boundedness of e', we obtain that

$$|e(x+1) - e(x)| = \left| e\left(\frac{f(y)}{s} + 1\right) - e\left(\frac{f(y)}{s}\right) \right| \leq c.$$

The same argument works for all $x \in \mathbb{Q}.\mathbb{Z}$, so all the endomorphisms of $\mathbb{Q}.\Delta$ are $(1,c)$-bounded and $\mathbb{Q}.\Delta$ has rank one. Similarly, we have

$$\left| e\left(x + \frac{t'}{s}\right) - e(x) \right| = \left| e\left(\frac{f(y)}{s} + \frac{t'}{s}\right) - e\left(\frac{f(y)}{s}\right) \right| \leq \frac{t'}{s}$$

i.e., e is tightly-t'/s-bounded and $\mathbb{Q}.\Delta$ has tight rank at most t'/s.

Since Δ satisfies all the assumptions that we had on Γ, we may repeat the argument. If $\mathbb{Q}.\Delta$ has tight rank 1, then we are done. This process terminates, since the tight rank of $\mathbb{Q}.\Delta$ is bounded above by t'/s, which is strictly smaller than the tight rank of $\mathbb{Q}.\Gamma$. Observe furthermore that if Δ' is the first-order reduct of $(\mathbb{Z};<)$ that is isomorphic to Δ/s', then Δ' is isomorphic to Γ/ss' by the obvious composition of isomorphisms, so that the resulting structure at termination is indeed of the form Γ/k for some $k \in \mathbb{N}$. $\square$

Theorem 7. *Let Γ be a finite-signature first-order reduct of $(\mathbb{Z};<)$ such that $\mathbb{Q}.\Gamma$ has rank one. Then* $\mathrm{CSP}(\Gamma)$ *equals* $\mathrm{CSP}(\Delta)$ *where Δ is one of the following:*

1. *a finite structure;*
2. *a first-order reduct of $(\mathbb{Z};<)$ where* Dist_k *is pp-definable for all $k \geq 1$;*
3. *a first-order reduct of $(\mathbb{Z};<)$ where succ is pp-definable.*

Proof. If Γ has a finite-range endomorphism f, then the image of the endomorphism induces a finite structure with the same CSP as Γ, thus we are in case (1) of the statement and done. So assume that this is not the case. Then by Theorem 6, Γ has an endomorphism g that maps Γ to Γ/k which is isomorphic to a reduct Δ of $(\mathbb{Z};<)$ such that $\mathbb{Q}.\Delta$ has tight rank one. Lemma 10 implies that Δ has tight rank one, too. The structure Γ/k cannot have finite-range endomorphisms f since otherwise $f \circ g$ would be a finite-range endomorphism for Γ. Hence, Δ does not have finite-range endomorphisms. Since $\mathbb{Q}.\Delta$ has rank one, Corollary 2 is applicable, and implies that 1 is stable for every endomorphism of $\mathbb{Q}.\Delta$. Hence all endomorphisms of $\mathbb{Q}.\Delta$ are isometries and the relation Dist_k is preserved by the endomorphisms of $\mathbb{Q}.\Delta$.

If *succ* is preserved by all the endomorphisms of $\mathbb{Q}.\Delta$, then Theorem 2 implies that *succ* is pp-definable in $\mathbb{Q}.\Delta$ since *succ* is 1-generated under $\mathrm{End}(\mathbb{Q}.\Delta)$. In this case, *succ* is pp-definable in Δ, too, and we are in case (3) of the statement.

Otherwise, there exists an endomorphism e of $\mathbb{Q}.\Delta$ that does not preserve *succ*. Therefore, there exists an $x \in \mathbb{Q}.\mathbb{Z}$ such that $e(x+k) = e(x) - k$ for all $k \geq 1$. For each $k \geq 1$, the relation Dist_k is then 1-generated under $\mathrm{End}(\mathbb{Q}.\Delta)$, the pair $(x, x+k)$ being a generator. Since Dist_k is preserved by all endomorphisms of $\mathbb{Q}.\Delta$, it follows from Theorem 2 that Dist_k is pp-definable in $\mathbb{Q}.\Delta$ for all $k \geq 1$. Finally, this implies that Dist_k is pp-definable in Δ for all $k \geq 1$ and we are in case (2) of the statement. $\square$

5.5.2 Arbitrary Rank

In this section we study first-order reducts of $(\mathbb{Z};<)$ of arbitrary finite rank. The goal is to reduce this to the rank one situation (in Proposition 6). For this, we need the following proposition, which is quite similar, but formally unrelated, to the implication from item (2) to item (4) in Theorem 5.

Lemma 17. *Let Γ be a finite-signature first-order reduct of $(\mathbb{Z};<)$ and $k \in \mathbb{N}$ such that $\mathbb{Q}.\Gamma$ is not k-bounded. Then for all $x, y \in \mathbb{Z}$ such that $x - y = k$ there exists an endomorphism h of $\mathbb{Q}.\Gamma$ such that $h(x) - h(y) = \infty$ and for all z, z' with $z - z' = \infty$ we have $h(z) - h(z') = \infty$.*

Proof. Since $\mathbb{Q}.\Gamma$ is not k-bounded, for any $r \geq 0$ there exist $x_0, y_0 \in \mathbb{Q}.\mathbb{Z}$ with $|x_0 - y_0| = k$ and an endomorphism $e\colon \mathbb{Q}.\Gamma \to \mathbb{Q}.\Gamma$ such that $e(x_0) - e(y_0) > r$. Composing e with an automorphism we can take $\{x_0, y_0\} = \{x, y\}$. For every finite set $S \subset \mathbb{Z}$, we then have a homomorphism $e\colon \mathbb{Q}.\Gamma[S] \to \mathbb{Q}.\Gamma$ such that $e(x) - e(y) > r$. It follows from an analog of Proposition 3 that there exists a homomorphism $h\colon \mathbb{Q}.\Gamma \to \mathbb{Q}.\Gamma$ such that $h(x) - h(y) = \infty$ and for all z, z' with $z - z' = \infty$ we have $h(z) - h(z') = \infty$. $\square$

Proposition 5. *Let Γ be a finite-signature first-order reduct of $(\mathbb{Z};<)$ such that $\mathbb{Q}.\Gamma$ has rank r, and let e be an endomorphism of $\mathbb{Q}.\Gamma$. Then $e(z_1) = e(z_2) \mod r$ for all $z_1, z_2 \in \mathbb{Q}.\mathbb{Z}$ such that $z_1 = z_2 \mod r$.*

Proof. Suppose that $e \in \mathrm{End}(\mathbb{Q}.\Gamma)$ is (r, c)-bounded and $z_1, z_2 \in \mathbb{Q}.\mathbb{Z}$ contradict the statement of the proposition. Choose z_1, z_2 such that $z_1 > z_2$ and $z_1 - z_2$ is minimal.

Claim 1. $z_1 - z_2 = r$.

Suppose otherwise; then there are $p_1, \dots, p_k$ for $k > 2$ such that $p_1 = z_1$, $p_k = z_2$, and $p_i - p_{i+1} = r$ for all $i \in \{1, \dots, k-1\}$ because r divides $z_1 - z_2$. By the choice of z_1, z_2 we have $e(p_i) = e(p_j) \bmod r$. But then $e(p_1) = e(p_k) \bmod r$, a contradiction to the assumption that $e(z_1) \neq e(z_2) \bmod r$.

Let $w, v \in \mathbb{N}$ be such that $|e(z_1) - e(z_2)| = wr + v$ and $v < r$. Note that $v > 0$ because $e(z_1) \neq e(z_2) \bmod r$. Assume that $e(z_1) > e(z_2)$; the proof when $e(z_2) > e(z_1)$ is analogous. Let $e' \in \mathrm{End}(\mathbb{Q}.\Gamma)$ be arbitrary, and $u_1, u_2 \in \mathbb{Z}$ be arbitrary such that $u_1 - u_2 = v$.

Claim 2. $|e'(u_1) - e'(u_2)| \leq (w+1)c + 1$.

To prove the claim, suppose the contrary. Let $\alpha \in \mathrm{Aut}(\mathbb{Z};<)$ be such that $\alpha(e(z_1)) = u_1$. Note that $\alpha(e(z_2) + wr) = u_2$. Set $e'' := e' \circ \alpha \circ e$. Then

$$\begin{aligned}
|e''(z_1) - e''(z_2))| &\geq |e''(z_1) - e'(u_2)| - |e'(u_2) - e''(z_2))| \\
&= |e'(u_1) - e'(u_2)| - |e'(\alpha(e(z_2) + wr)) - e'(\alpha(e(z_2)))| \\
&\geq (w+1)c + 1 - wc = c + 1
\end{aligned}$$

where the first inequality is the triangle inequality, and the second inequality is by assumption and (r, c)-boundedness. But $|e''(z_1)) - e''(z_2))| > c$ contradicts the assumption that $\mathbb{Q}.\Gamma$ is (r, c)-bounded, and this finishes the proof of Claim 2.

Since e' was chosen arbitrarily, we obtain that $\mathbb{Q}.\Gamma$ is $(v, w(c+1)+1)$-bounded, and hence has rank $v < r$, a contradiction. $\square$

Lemma 18. *Let Γ be a finite-signature first-order reduct of $(\mathbb{Z};<)$ such that $\mathbb{Q}.\Gamma$ has rank $r \in \mathbb{N}$. Then there exists an endomorphism e of $\mathbb{Q}.\Gamma$ with the property that for all $x, y \in \mathbb{Q}.\mathbb{Z}$,*

$$\begin{aligned}
&\textit{either } e(y) - e(x) = \infty \\
&\textit{or } e(y) - e(x) = 0 \mod r.
\end{aligned}$$

Proof. We construct e by an application of König's tree lemma as follows. Let $a_1, a_2, \ldots$ be an enumeration of the elements of $\mathbb{Q}.\mathbb{Z}$. Given a partial function $f\colon \{a_1,\ldots,a_n\} \to \mathbb{Q}.\mathbb{Z}$, we say that f has property (†) if for all $x, y \in \{a_1,\ldots,a_n\}$, either $f(y)-f(x)=\infty$ or $f(x)=f(y) \mod r$. The vertices on level n of the tree are $\sim_q$-equivalence classes of homomorphisms h from $\mathbb{Q}.\Gamma[\{a_1,\ldots,a_n\}]$ to $\mathbb{Q}.\Gamma$ that satisfy property (†). Adjacency between vertices is defined by restriction of representatives.

The interesting part of the proof is to show that the tree has vertices on all levels. Let g be a homomorphism from $\mathbb{Q}.\Gamma[\{a_1,\ldots,a_n\}]$ to $\mathbb{Q}.\Gamma$ such that the number m of pairs $i, j \in \{1,\ldots,n\}$ with $g(a_i)-g(a_j)=\infty$ or $g(a_i)=g(a_j) \mod r$ is maximal. If $m=\binom{n}{2}$ then we are done; so suppose that there are $p, q \in \{1,\ldots,n\}$ such that $g(a_p)-g(a_q) \in \mathbb{Z}$ is not divisible by r. Let $k \in \{1,\ldots,r-1\}$ and $l \in \mathbb{Z}$ be such that $g(a_p)-g(a_q)=lr+k$, $0<k<r$. Since $\mathbb{Q}.\Gamma$ is not k-bounded, by Lemma 17 there exists an endomorphism f of $\mathbb{Q}.\Gamma$ such that $f(g(a_p)+lr)-f(g(a_q))=\infty$. By Proposition 5 we have $f(g(a_p))=f(g(a_p)+lr) \mod r$, and hence $f(g(a_p))-f(g(a_q))=\infty$. We claim that the number m' of pairs $i, j \in \{1,\ldots,n\}$ such that $f(g(a_i))-f(g(a_j))=\infty$ or $f(g(a_i))=f(g(a_j)) \mod r$ is larger than m. If $g(a_i)-g(a_j)=\infty$ then $f(g(a_i))-f(g(a_j))=\infty$; if $g(a_i)=g(a_j) \mod r$ then $f(g(a_i))=f(g(a_j)) \mod r$. Therefore, $m' \geq m$. Moreover, we have $f(g(a_p))-f(g(a_q))=\infty$, and hence $m'>m$. Then $f \circ g$ is a homomorphism from $\mathbb{Q}.\Gamma[a_1,\ldots,a_n] \to \mathbb{Q}.\Gamma$, contradicting the maximality of m.

By Lemma 4, we obtain an endomorphism $e\colon \mathbb{Q}.\mathbb{Z} \to \mathbb{Q}.\mathbb{Z}$ such that for every n, $e|_{\{a_1,\ldots,a_n\}}$ is $\sim_q$-equivalent to some function g_n satisfying (†). Let $x, y \in \mathbb{Q}.\mathbb{Z}$. If x, y are (e,q)-connected, then they are $(e|_{\{a_1,\ldots,a_n\}}, q)$-connected for some n, so that they are (g_n,q)-connected. It follows that $e(x)-e(y)=g_n(x)-g_n(y)=0 \mod r$. If x, y are not (e,q)-connected, they are not (g_n,q)-connected for any function g_n in the tree, and we have $e(x)-e(y)=\infty$ by Lemma 4. Therefore, e satisfies (†). $\square$

Proposition 6. *Let Γ be a finite-signature reduct of $(\mathbb{Z};<)$ without finite-range endomorphism and such that $\mathbb{Q}.\Gamma$ has rank $r \in \mathbb{N}$ and tight rank $t \in \mathbb{N}$. Then Γ/r has the same CSP as Γ, and is isomorphic to a first-order reduct Δ of $(\mathbb{Z};<)$ such that $\mathbb{Q}.\Delta$ has tight rank at most t/r.*

Proof. By Lemma 16, there is a first-order reduct Δ of $(\mathbb{Z};<)$ such that $x \mapsto r\cdot x$ is an isomorphism between Δ and Γ/r. Let e be the endomorphism of $\mathbb{Q}.\Gamma$ constructed in Lemma 18. Replacing e by $\alpha \circ e$ for an appropriate automorphism α of $(\mathbb{Q}.\mathbb{Z};<)$, we can assume that the range of e lies within $S := \{r\cdot z : z \in \mathbb{Q}.\mathbb{Z}\}$. Since $x \mapsto r\cdot x$ is an isomorphism between $\mathbb{Q}.\Delta$ and the structure induced by S in $\mathbb{Q}.\Gamma$, we obtain that Γ, $\mathbb{Q}.\Gamma$, $\mathbb{Q}.\Delta$, and Δ all have the same CSP.

It remains to be shown that $\mathbb{Q}.\Delta$ has rank at most t/r. For an arbitrary $e \in \mathrm{End}(\mathbb{Q}.\Gamma)$, the quantity $\delta(e) := \max_{z\in\mathbb{Q}.\mathbb{Z}} |e(z+t)-e(z)|$ is well-defined and finite, since $\mathbb{Q}.\Gamma$ has tight-rank t. Let e be an endomorphism of $\mathbb{Q}.\Gamma$ as in Lemma 18 such that $\delta(e)$ is maximal among all endomorphisms satisfying the conclusion of Lemma 18. Let $z_0 \in \mathbb{Q}.\Gamma$ be a witness for the maximum taken in $\delta(e)$. If $e(z_0+t)=e(z_0)$, then for all $z \in \mathbb{Q}.\mathbb{Z}$ we have $e(z+t)=e(z)$. As in the proof of (1) $\Rightarrow$ (2) in Theorem 5, this would imply that Γ has a finite-range endomorphism, a contradiction to the assumption. So we have $e(z_0+t) \neq e(z_0)$. Suppose that $e(z_0+t) > e(z_0)$, the other case being treated similarly. Since e satisfies the property of Lemma 18, the distance $e(z_0+t)-e(z_0)$ is equal to kr for some $k \leq t/r$. We prove that $\mathbb{Q}.\Delta$ has tight rank k. Let f be an endomorphism of $\mathbb{Q}.\Delta$, and suppose that there exists a $y \in \mathbb{Q}.\Delta$ such that $|f(y+k)-f(y)|>k$. Up to composition of f with an automorphism of $\mathbb{Q}.\Delta$, we can assume that $y=\frac{e(z_0)}{r}$. Note that $y+k=\frac{e(z_0)}{r}+k=\frac{e(z_0)+kr}{r}=\frac{e(z_0+t)}{r}$. Let $e'\colon \mathbb{Q}.\Gamma \to \mathbb{Q}.\Gamma$ be defined by $e'(x)=r\cdot f(\frac{e(x)}{r})$. Note that e' satisfies the property of Lemma 18. We have furthermore $|e'(z_0+t) - e'(z_0)| = r \cdot |f(y+k) - f(y)| > kr$. This contradicts the fact that e was chosen to maximise the distance $|e(z_0+t) - e(z_0)|$. $\square$

Iterating the previous proposition, we finally obtain a reduction to the rank one case.

Corollary 3. *Let Γ be a finite-signature reduct of $(\mathbb{Z};<)$ such that $\mathbb{Q}.\Gamma$ has rank $r \in \mathbb{N}$. Then there exists a $k \in \mathbb{N}$ such that Γ/k has the same CSP as Γ, and is isomorphic to a reduct Δ of $(\mathbb{Z};<)$ such that $\mathbb{Q}.\Delta$ has rank one.*

Proof. If $\mathbb{Q}.\Gamma$ has rank one there is nothing to prove, so assume that $r > 1$. By Proposition 6, Γ/r has the same CSP as Γ, is isomorphic to a reduct Δ_1 of $(\mathbb{Z};<)$, and the tight rank t_1 of $\mathbb{Q}.\Delta_1$ is strictly smaller than that of $\mathbb{Q}.\Gamma$. Write $\Delta_0 := \Gamma$. We iterate this construction, obtaining reducts $\Delta_0, \Delta_1, \ldots, \Delta_{n+1}$ of $(\mathbb{Z};<)$ with ranks $r_0, r_1, \ldots, r_{n+1}$ and tight ranks $t_0 > t_1 > \cdots > t_n = t_{n+1}$ until the sequence of tight ranks stabilises, which can only happen if the rank of $\mathbb{Q}.\Delta_n$ is one. The structure Δ_n is isomorphic to $\Gamma/(r_0 \ldots r_{n-1})$, which proves the corollary. $\square$

5.6 Defining *succ* and $<$

In the remainder of this section, we prove the following dichotomy: a first-order reduct of $(\mathbb{Z};<)$ that pp-defines *succ* either pp-defines $<$, or is a first-order reduct of $(\mathbb{Z}; succ)$. Call a binary relation R with a first-order definition over $(\mathbb{Z};<)$ *one-sided infinite* if there exist $c, d \in \mathbb{Z}$ with $c \leq d$ so that

- $R(x, x+z)$ holds for no $z < c$,
- $R(x, x+z)$ holds for all $z \geq d$.

Note that this definition does not depend on $x \in \mathbb{Z}$, since R is first-order definable over $(\mathbb{Z};<)$.

Lemma 19. *Let Γ be a first-order reduct of $(\mathbb{Q}.\mathbb{Z};<)$ such that succ is pp-definable in Γ. Then $<$ is pp-definable in Γ if and only if some one-sided infinite binary relation is pp-definable in Γ.*

Proof. Since $<$ is one-sided infinite we only have to show the reverse implication. Choose a binary one-sided infinite relation R with a pp-definition in Γ such $d - c$ is minimal, where c and d are as in the definition of one-sided infinity of R. If $c = d$ then R is a relation of the form $x < y + k$ for $k \in \mathbb{Z}$, and using *succ* we can pp-define $<$ in Γ. We now show that $c \neq d$ is impossible. Replace R by the relation T defined by the formula $R(x, y) \wedge R(x, y + d - c - 1)$, which is equivalent to a pp-formula over Γ. Then $(0, z)$ is in T for all $z \geq d$. On the other hand, for $z < c + 1$, we have $(0, z) \notin T$. Indeed, if $z < c$ then $(0, z)$ is not in R, so not in T. If $z = c$, then $(0, d-1)$ is not in R by the minimality of d, so that $(0, c)$ is not in T. Therefore, the integers c', d' as defined for T in place of R have a smaller difference than $d - c$, contradicting the choice of R such that $d - c$ is minimal. $\square$

If R is a relation of arity n, and $i_1, \ldots, i_k \in \{1, \ldots, n\}$ are distinct indices, the projection of R onto $\{i_1, \ldots, i_k\}$, denoted by $\pi_{i_1,\ldots,i_k}(R)$, is the relation defined by $\exists x_{j_1}, \ldots, x_{j_{n-k}}.R(x_1, \ldots, x_n)$ over $(\mathbb{Z}; R)$ where $\{j_1, \ldots, j_{n-k}\} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$. A *binary projection* of R is a projection of R onto a set of size 2.

Lemma 20. *Let Γ be a first-order reduct of $(\mathbb{Z};<)$ in which succ is pp-definable. Then, either Γ pp-defines $<$ or Γ is a first-order reduct of $(\mathbb{Z}; succ)$.*

Proof. Let R be a relation of Γ of arity k. If $E = \{a - b \mid (a,b) \in \pi_{i,j}(R)\}$, for distinct $i, j \in \{1, \ldots, k\}$, is a finite or cofinite set, there is a definition of $R(x_1, \ldots, x_k)$ over $(\mathbb{Z}; <)$ without literals of the form $x_i < x_j + k$. Indeed, such a literal can be replaced by a disjunction of literals $succ^p(x_i, x_j)$ for suitable integers p if E is finite, or a by a conjunction of literals $\neg succ^p(x_i, x_j)$ if E is cofinite. Therefore, if Γ is not a first-order reduct of $(\mathbb{Z}; succ)$ there exists a relation R of Γ and integers i, j such that the set $\{a - b \mid (a,b) \in R'\}$ for $R' := \pi_{i,j}(R)$ is neither finite nor cofinite. Let q be the quantifier-elimination degree of R'. It is clear that if $(a,b) \in R'$ with $a - b > q$, then $(a', b') \in R'$ whenever $a' - b' > q$. It follows that R' or $\{(b,a) \in \mathbb{Z}^2 \mid (a,b) \in R'\}$ is one-sided infinite. From Lemma 19 and the fact that R' is pp-definable in Γ follows that $<$ is pp-definable in Γ. $\square$

Combining the results of the preceding subsections, we can finally prove Theorem 4, which we restate here for the convenience of the reader.

Theorem 8. *Let Γ be a first-order reduct of $(\mathbb{Z}; <)$ with finite signature. Then* $\mathrm{CSP}(\Gamma)$ *equals* $\mathrm{CSP}(\Delta)$ *where Δ is one of the following:*

1. *a finite structure;*
2. *a first-order reduct of $(\mathbb{Q}; <)$;*
3. *a first-order reduct of $(\mathbb{Z}; <)$ where Dist_k is pp-definable for all $k \geq 1$;*
4. *a first-order reduct of $(\mathbb{Z}; <)$ where succ and $<$ are pp-definable;*
5. *a first-order reduct of $(\mathbb{Z}; succ)$ where succ is pp-definable.*

Proof. Let Γ be a first-order reduct of $(\mathbb{Z}; <)$ with finite signature. If Γ has an endomorphism with finite range, then Γ is homomorphically equivalent to a finite structure; hence item (1) of Theorem 4 holds and we are done. So suppose that this is not the case. If there exists a first-order reduct of $(\mathbb{Q}; <)$ with the same CSP, then item (2) of Theorem 4 holds and we are done. Otherwise, the equivalence of (2) and (1) in Theorem 5 implies that $\mathbb{Q}.\Gamma$ has bounded tight rank t and bounded rank r. If $r > 1$, then by Proposition 6 we have that Γ has the same CSP as a first-order reduct Δ of $(\mathbb{Z}; <)$ such that $\mathbb{Q}.\Delta$ has rank 1. It follows from Theorem 7 that there exists a first-order reduct Δ' of $(\mathbb{Z}; <)$ that has the same CSP as Γ and such that Dist_k is pp-definable in Δ' for all $k \geq 1$ or *succ* is pp-definable in Δ'. In the former case, item (3) of Theorem 4 holds. In the latter case, we finally have by Lemma 20 that $<$ has a pp-definition in Δ', in which case item (4) holds, or that Δ' is a first-order reduct of $(\mathbb{Z}; succ)$, in which case item (5) holds. $\square$

6 Tractable Classes

We treat the algorithmic part of our main result, that is, we prove that if Γ is a first-order reduct of $(\mathbb{Z}; <)$ that is preserved by $\max_d$ or $\min_d$, or if Γ is a first-order reduct of $(\mathbb{Z}; succ)$ such that $\mathbb{Q}.\Gamma$ is preserved by a binary injective operation preserving *succ*, then $\mathrm{CSP}(\Gamma)$ is in P (items (3) and (4) in Theorem 3).

6.1 The Horn Case

The two structures $(\mathbb{Q}.\mathbb{Z}, succ)^2$ and $(\mathbb{Q}.\mathbb{Z}, succ)$ are isomorphic. Let si be an isomorphism from $(\mathbb{Q}.\mathbb{Z}, succ)^2$ to $(\mathbb{Q}.\mathbb{Z}, succ)$. In the following we will also consider si as a binary operation on $\mathbb{Q}.\mathbb{Z}$ that preserves $succ$. Remember that relations that are first-order definable over $(\mathbb{Z}; succ)$ are also definable by quantifier-free formulas with (positive or negative) literals of the form $succ^p(x, y)$ for $p \in \mathbb{Z}$ (see items (1)-(5) in Example 1). A quantifier-free formula in conjunctive normal form (CNF) is called *Horn* if each clause of the formula contains at most one positive literal. A relation is said to be *Horn-definable* if there exists a Horn formula that defines the relation.

We use the following characterisation of Horn definability, which is Proposition 5.9 in [4]: if Δ is a structure with an embedding e of Δ^2 into Δ (such as for instance $\Delta = (\mathbb{Q}.\mathbb{Z}; succ)$) then a relation R with a quantifier-free definition in Δ is Horn-definable over Δ if and only if R is preserved by e. Applied to our situation, we obtain the following.

Proposition 7. *Let Γ be a first-order reduct of $(\mathbb{Z}; succ)$. Then the following are equivalent.*

- *every relation of Γ is Horn-definable over $(\mathbb{Z}; succ)$;*
- *$\mathbb{Q}.\Gamma$ is preserved by si;*
- *$\mathbb{Q}.\Gamma$ has a binary injective polymorphism that preserves succ.*

Proposition 8. *Let Γ be a finite-signature first-order reduct of $(\mathbb{Z}; succ)$ such that si is a polymorphism of $\mathbb{Q}.\Gamma$. Then $\mathrm{CSP}(\Gamma)$ is in P.*

Proof. From Proposition 7 we know that the relations of Γ are definable with quantifier-free Horn formulas over $(\mathbb{Z}; succ)$. It is easy to see that there is a polynomial-time algorithm that decides whether a set of constraints of the form $succ^{p_i}(x_i, y_i)$ is satisfiable. Moreover, we can also efficiently decide whether it implies another constraint of this form. Indeed, to see if the set of constraints is satisfiable, consider the graph whose vertices are the variables, and whose arcs consists of those pairs (x_i, y_i), labelled by p_i, such that there is a constraint $succ^{p_i}(x_i, y_i)$ in the input. For each variable x, using a graph traversal we can check whether all the directed paths going from x to some other variable y have the same weight (which is given by the sum of the labels over the arcs); If this is not the case, the constraints are unsatisfiable. Otherwise, to decide whether the constraints imply $succ^p(x, y)$, check whether there is a directed path from x to y where the sum of the labels equals p.

We view the instance of $\mathrm{CSP}(\Gamma)$ as a set of Horn-clauses over $(\mathbb{Z}; succ)$. We iterate the following algorithm: form the set C of clauses that consist of only one positive literal (these clauses are called *positive unit clauses*). For each negative literal $\neg\ell$ appearing in the instance, we can use the algorithm above to test whether C is consistent and whether it implies ℓ. If C is inconsistent, we reject the instance. If ℓ is implied by C, we remove every occurrence of $\neg\ell$ in the input. If we derive the empty clause, we reject the input. Otherwise, the resolution stabilises in a polynomial number of steps with a set of Horn clauses; in this case, accept the input. Since the resulting clauses are Horn, they are preserved by si. We apply si to show that in this case indeed there exists a solution. By assumption, for each Horn clause $\bigwedge_i succ^{p_i}(x_i, y_i) \Rightarrow succ^p(x, y)$ there exists an assignment that falsifies some literal $succ^{p_i}(x_i, y_i)$ and additionally satisfies all the positive unit clauses: otherwise the literal would have been removed by the resolution procedure. Let $s_1, \ldots, s_r$ be those assignments for the r clauses. Since si is an isomorphism, the assignment $s := si(s_1, \ldots, si(s_{r-1}, s_r) \ldots)$ simultaneously breaks all the equalities in the premises of all the clauses. Moreover, since si preserves *succ*, the resulting assignment s preserves the positive unit clauses, and hence is a valid assignment for the input. $\square$

6.2 Modular Minimum and Modular Maximum

Theorem 9. *Let* Γ *be a finite-signature first-order reduct of* $(\mathbb{Z}; <)$ *that admits a modular max or modular min polymorphism. Then* $\mathrm{CSP}(\Gamma)$ *is in P.*

Proof. Suppose that Γ is preserved by max, the regular maximum operation. Then $\mathrm{CSP}(\Gamma)$ is solvable in polynomial time as follows. Let q be the qe-degree of Γ. Let φ be an instance of $\mathrm{CSP}(\Gamma)$ with n variables. We already noted in the proof of Proposition 1 that φ is satisfiable in Γ iff it is satisfiable in $\Gamma[\{0, \dots, (q+1)n\}]$, and the latter structure can be constructed in polynomial time, and is preserved by the maximum function on $\{0, \dots, (q+1)n\}$. We can then decide whether $\Gamma[\{0, \dots, (q+1)n\}] \models \varphi$ using the arc-consistency algorithm, noting that the arc-consistency procedure can be implemented in such a way that the running time is polynomial in both the size of the formula and of the structure [20].

Suppose now that Γ is preserved by $\max_d$ for $d \geq 2$. It follows that $<$ is not pp-definable in Γ, as $\max_d$ does not preserve $<$. We can suppose that Γ pp-defines *succ*, because this only increases the complexity of $\mathrm{CSP}(\Gamma)$ and *succ* is preserved by $\max_d$. By Lemma 20, Γ is a first-order reduct of $(\mathbb{Z}; \mathit{succ})$. In [3], the authors prove that the CSP of a first-order reduct of $(\mathbb{Z}; \mathit{succ})$ with finite distance degree and which is preserved by a modular maximum or minimum is decidable in polynomial time. An inspection of the proof shows that the finite distance degree hypothesis is not necessary. Indeed, the critical idea of the algorithm is that if Γ is preserved by the d-modular maximum, then $\mathrm{CSP}(\Gamma)$ reduces in polynomial time to $\mathrm{CSP}(\Delta)$, where Δ is a reduct of $(\mathbb{Z}; \mathit{succ})$ which is preserved by the usual maximum or minimum. Then, arc-consistency can be used to solve $\mathrm{CSP}(\Delta)$ in polynomial time (for the details, we have to refer to [3]). The reduction and the algorithm for $\mathrm{CSP}(\Delta)$ do not rely on the distance degree of Γ being finite to work. $\square$

7 The Classification

In this section we prove our complexity classification result, Theorem 3. By Theorem 4 and the comments before and after Proposition 2, we are left with the task to classify the CSP for finite-signature reducts Γ of $(\mathbb{Z}; <)$ where the binary relation *succ* is among the relations of Γ (that is, when we are in case (4) or (5) of Theorem 4).

An important case distinction in this section is whether the order relation $<$ is primitive positive definable in Γ. The situation when this is the case is treated in Section 7.1. Otherwise, if *succ* is pp-definable in Γ, but $<$ is not, then Γ is a first-order reduct of $(\mathbb{Z}; \mathit{succ})$ by Lemma 20. In this case, we further distinguish whether Γ is *positive* in the sense that each of its relations can be defined over $(\mathbb{Z}; \mathit{succ})$ with a *positive* quantifier-free formula, that is, a first-order formula without negation symbols. Positivity of reducts of $(\mathbb{Z}; \mathit{succ})$ has several natural different characterisations, which is the topic of Section 7.2. We first treat the case of non-positive reducts of $(\mathbb{Z}; \mathit{succ})$ in Section 7.3, and then the case of positive reducts of $(\mathbb{Z}; \mathit{succ})$ in Section 7.4. All the formulas considered here are quantifier-free unless stated otherwise.

7.1 First-order expansions of $(\mathbb{Z}; succ, <)$

We have already seen that the CSP for first-order reducts of $(\mathbb{Z}; <)$ preserved by max or by min is in P. The following lemma provides the matching hardness result for first-order expansions of $(\mathbb{Z}; <, succ)$.

Definition 10. A *d-progression* is a set of the form $[a, b \mid d] := \{a, a + d, a + 2d, \dots, b\}$, for $a \leq b$ with $b - a$ divisible by d. A d-progression is *trivial* if it has cardinality one.

We need the following, which is Proposition 47 from [3]. Remember that a structure definable over $(\mathbb{Z}; succ)$ is *locally finite* if every relation has finite distance degree.

Proposition 9. *Let Γ be a locally finite first-order expansion of $(\mathbb{Z}; succ)$ such that Diff_S is pp-definable in Γ for a non-trivial 1-progression S. If Γ is neither preserved by max nor min then $\mathrm{CSP}(\Gamma)$ is NP-hard.*

Lemma 21. *Let Γ be a first-order expansion of $(\mathbb{Z}; <, succ)$. If Γ is preserved by neither* max *nor* min*, then $\mathrm{CSP}(\Gamma)$ is NP-hard.*

Proof. Let R be a relation of Γ which is not preserved by max, and let T be a relation of Γ which is not preserved by min. Then there are tuples $\overline{a}, \overline{b}$ in R such that $\max(\overline{a}, \overline{b}) \notin R$. Let m be $\max_{i,j}(|a_i - a_j|, |b_i - b_j|)$. Since the binary relation defined by $x \leq y + m$ has a pp-definition in Γ, the relation R^* defined by

$$R(x_1, \dots, x_n) \wedge \bigwedge_{i,j} x_i \leq x_j + m$$

is pp-definable in Γ, too. Note that $\overline{a}$ and $\overline{b}$ are in R^*, and that $\max(\overline{a}, \overline{b}) \notin R^*$. Also note that R^* is first-order definable over *succ* and has finite distance degree. Dually, we find a pp-definition over Γ of a relation T^* which is not preserved by min, first-order definable over *succ* and with finite distance degree. The primitive positive formula $\exists u(u = succ^3(x) \wedge x < y \wedge y < u)$ defines $\mathrm{Diff}_{\{1,2\}}$. It then follows from Proposition 9 that $\mathrm{CSP}(\mathbb{Z}; succ, R^*, T^*)$ is NP-hard, and hence $\mathrm{CSP}(\Gamma)$ is NP-hard, too. $\square$

7.2 Endomorphisms of and Definability in Positive Reducts

Positivity of reducts Γ of $(\mathbb{Z}; succ)$ can be characterised via the endomorphisms of $\mathbb{Q}.\Gamma$, but also via the non-definability of certain binary relations with primitive positive formulas (Lemma 22). These binary relations then play an important role in the complexity classification of the non-positive reducts of $(\mathbb{Z}; succ)$.

Binary relations R with a first-order definition in $(\mathbb{Z}; succ)$ come in two flavours. Indeed, the set $\{x - y \mid (x, y) \in R\}$ is either finite or cofinite. This easily follows from the quantifier elimination in $(\mathbb{Z}; succ)$. Remember that a binary relation R that is first-order definable over $(\mathbb{Z}; succ)$ (or over $(\mathbb{Q}.\mathbb{Z}; succ)$) is called *basic* if it is empty, $\mathbb{Z}^2$, or defined by the formula $y = x + c$ for some $c \in \mathbb{Z}$, and *non-basic* otherwise.

In the following, we use expressions of the form $succ^p(x, y)$ (see Example 1) as if they were atomic symbols of the language. Since they are pp-definable in a first-order expansion of $(\mathbb{Z}; succ)$, this is without loss of generality. Recall that a formula over *succ* is *positive* if it only includes literals of the form $succ^p(x, y)$. A formula over the signature of $(\mathbb{Z}; succ)$ in disjunctive normal form (DNF) is called *reduced* when every formula obtained by removing literals or conjunctive clauses is not logically equivalent over $(\mathbb{Z}; succ)$. It is clear that every first-order formula on $(\mathbb{Z}; succ)$ is equivalent to a reduced formula in DNF.

Lemma 22. *For a first-order expansion Γ of $(\mathbb{Z}; succ)$, the following are equivalent:*

1. *Every formula in reduced DNF that defines a relation of Γ is positive;*
2. *$\mathbb{Q}.\Gamma$ has an endomorphism that violates the binary relation given by $x - y = \infty$;*
3. *Γ does not pp-define a non-basic binary relation with infinite distance degree.*

Proof. (2) implies (1). Let e be an endomorphism of $\mathbb{Q}.\Gamma$ that violates $x - y = \infty$, and let a, b be such that $a - b = \infty$ and $e(a) - e(b) < \infty$. Using automorphisms of $(\mathbb{Q}.\mathbb{Z}; succ)$, we may assume that $e(a) = e(b) = b$ without loss of generality. For contradiction, suppose that Γ has a relation with a reduced DNF definition $\varphi(x_1, \ldots, x_n)$ which is not positive.

We now show that we can choose $s\colon \{x_1, \ldots, x_n\} \to \mathbb{Z}$ such that s is a satisfying assignment for φ but $e \circ s$ is not. For this, let us write one of the non-positive disjuncts ψ of φ as $\neg succ^p(z_2, z_1) \wedge \varphi'$ where φ' is a conjunction of literals, $z_1, z_2 \in \{x_1, \ldots, x_n\}$, and $p \in \mathbb{Z}$. Moreover, let $\psi_2, \ldots, \psi_m$ be the other disjuncts of φ. Suppose that all assignments that satisfy $\varphi' \wedge succ^p(z_2, z_1)$ also satisfy $\bigvee_{2 \leq i \leq m} \psi_i$. Then we could rewrite φ as $\varphi' \vee \bigvee_{i \geq 2} \psi_i$, which is impossible since φ is reduced. Hence, there exists $t\colon \{x_1, \ldots, x_n\} \to \mathbb{Z}$ such that t is a satisfying assignment for $\varphi' \wedge succ^p(z_2, z_1)$ but not for ψ_i for every $i \geq 2$; in particular, t does not satisfy φ. Using an automorphism of $(\mathbb{Q}.\mathbb{Z}; succ)$, we can assume that $t(z_1) = b$. Moreover, we can assume that the image S of t lies in only one copy of $\mathbb{Z}$. To see this, let $g\colon S \to \mathbb{Q}.\mathbb{Z}$ be any function that maps S to the first copy of $\mathbb{Z}$ in such a way that if $t(x_i)$ and $t(x_j)$ are in different copies, then $g(t(x_i))$ and $g(t(x_j))$ are at distance at least $q + 1$, where q is the qe-degree of φ. We have that g is $\sim_q$-equivalent to an embedding of S into the first copy of $\mathbb{Z}$ in $\mathbb{Q}.\mathbb{Z}$. Therefore, by the substitution lemma (Lemma 3), the function $g \circ t$ is a satisfying assignment to the variables of φ that only occupies one copy of $\mathbb{Z}$.

We now derive from t an assignment s that satisfies $\neg succ^p(z_2, z_1)$, that gives the same truth value as t to all the other literals of ψ, and such that $e \circ s = t$. If we consider φ' as a graph on $\{z_1, \ldots, z_k\}$ where edges represent positive literals, then z_1 and z_2 are in different connected components. Indeed, if there were a path from z_1 to z_2 in this graph we would have that φ' implies a statement of the form $succ^q(z_2, z_1)$. But then the conjunction $\neg succ^p(z_2, z_1) \wedge succ^q(z_2, z_1)$ is either contradictory or is equivalent to $succ^q(z_2, z_1)$, which is in contradiction to φ being reduced. Let V be the variables in the connected component of z_1. Define s on V by $s(v) := a - t(z_1) + t(v)$ (in particular $s(z_1) = a$) and define $s(v) := t(v)$ on the variables v that are not in V. We have that s satisfies $\neg succ^p(z_2, z_1)$ and that the other literals in φ' are satisfied by s, too:

- The truth of positive literals is preserved since we performed a translation on variables that are connected by positive literals.
- Negative literals between the variables in V and the other variables are also true, since for $v \in V$ and $v' \notin V$ we have that $s(v) - s(v') = \infty$ ($s(v)$ lies in the same component as a and $s(v')$ lies in the same component as b.)
- Finally, negative literals between variables not in V are preserved: they are satisfied by t, and for $v \notin V$ we have $s(v) = t(v)$ by definition.

Hence, s is a satisfying assignment of φ. We have $e \circ s = t$: If v is a variable in V, then $e(s(v)) = e(a - t(z_1) + t(v)) = e(a) - t(z_1) + t(v) = t(v)$ since e preserves $succ$ and $e(a) = b = t(z_1)$, and if

$v \notin V$ we defined $s(v)$ to be $t(v)$, so that $e(s(v)) = e(t(v)) = t(v)$. Since t does not satisfy φ, this contradicts the assumption that e is an endomorphism of $\mathbb{Q}.\Gamma$.

(1) implies (3). Let R be a binary relation with a pp-definition $\varphi(x, y)$ in Γ of the form $\exists \overline{z} \bigwedge_i \varphi_i$ where φ_i is for each i an atomic formula over Γ. Let us replace φ_i by its definition ψ_i over $(\mathbb{Z}; \mathit{succ})$ in quantifier-free reduced DNF. By assumption, all the literals in ψ_i are positive. The formula $\varphi(x, y)$ is therefore equivalent to a formula $\psi(x, y) := \bigvee_j \exists \overline{z}.\psi_j$ where ψ_j is a conjunction of positive literals of the form $\mathit{succ}^k(u, v)$. If one of the disjuncts of ψ is vacuously true, then ψ defines a basic binary relation. So let us assume that this is not the case. Since all the literals in ψ_j are positive, the relations defined by the disjuncts have finite distance degree. Their disjunction therefore also defines a binary relation of finite distance degree. In either case, ψ does not define a non-basic binary relation of infinite distance degree.

(3) implies (2). Suppose that all the endomorphisms of $\mathbb{Q}.\Gamma$ preserve the binary relation defined by $x - y = \infty$. Then all the endomorphisms preserve the relation defined by $x \neq y$. Indeed, if $x - y < \infty$ then $e(x) - e(y) = x - y$ since e preserves succ, and hence $x \neq y$ implies $e(x) \neq e(y)$. On the other hand, if $x - y = \infty$, then $e(x) - e(y) = \infty$ by assumption. It follows from Theorem 2 that $x \neq y$ has an existential positive definition in $\mathbb{Q}.\Gamma$ and in Γ. Let $\bigvee \varphi_i(x, y)$ be such a definition, where each φ_i is a primitive positive formula over Γ. Since $\neq$ has infinite distance degree, one of the φ_i must define a binary relation with infinite distance degree. This relation is also distinct from $(\mathbb{Q}.\mathbb{Z})^2$ because it is contained in the relation defined by $x \neq y$, so the infinite distance degree implies that it is non-basic. Thus, item (3) does not hold. $\square$

7.3 The Non-positive Case

Let Γ be a non-positive reduct of $(\mathbb{Z}; \mathit{succ})$ such that $\mathbb{Q}.\Gamma$ is not preserved by si. Our aim in this section is to show that Γ has an NP-hard CSP. Together with Proposition 8, this completes the complexity classification for the CSP of non-positive reducts of $(\mathbb{Z}; \mathit{succ})$. Note that si is an arbitrary isomorphism $(\mathbb{Q}.\mathbb{Z}; \mathit{succ})^2 \to (\mathbb{Q}.\mathbb{Z}; \mathit{succ})$, but the discussion below does not depend on which function we take for si. Indeed, given two isomorphisms $\mathit{si}, \mathit{si}'$ as above, there exists an automorphism α of $(\mathbb{Q}.\mathbb{Z}; \mathit{succ})$ such that $\mathit{si} = \alpha \circ \mathit{si}'$.

In order to show that CSP(Γ) is NP-hard, we show in Proposition 10 that when $\mathbb{Q}.\Gamma$ is not preserved by si then there is a non-basic binary relation with finite distance degree that is pp-definable in Γ. This binary relation will serve to define the set of vertices of a certain finite undirected graph. The edge relation of that graph comes from the binary relation of Lemma 22 which provided an alternative characterisation of non-positivity of Γ. We finally use the classification of the CSPs for finite undirected graphs [15] to conclude that CSP(Γ) is NP-hard.

A formula φ in CNF is called *reduced* when removing any literal in a clause yields a formula that is not equivalent to φ. This is equivalent to saying that for any literal ℓ in a clause ψ of φ, there exists an assignment that satisfies φ and that satisfies only ℓ in ψ. This assignment witnesses the fact that the given literal cannot be removed from the formula without changing the set of satisfying assignments.

Lemma 23. *Let φ be a quantifier-free formula over $(\mathbb{Z}; \mathit{succ})$, and suppose that φ is equivalent to a Horn formula over $(\mathbb{Z}; \mathit{succ})$. If φ is reduced, then it is Horn.*

Proof. Note that φ is equivalent to a Horn formula over $(\mathbb{Z}; \mathit{succ})$ if and only if it is equivalent to a Horn formula over $(\mathbb{Q}.\mathbb{Z}; \mathit{succ})$, since both structures have the same first-order theory. By Proposition 7 the formula φ is preserved by si.

Suppose for contradiction that φ is not Horn, that is, it contains a clause ψ of the form $(succ^p(y,x) \vee succ^q(v,u) \vee \dots)$. Since this formula is reduced, there exist satisfying assignments s,t of φ such that s satisfies only $succ^p(y,x)$ in ψ, and t satisfies only $succ^q(v,u)$ in ψ. The assignment (s,t) that maps a variable x_i of φ to the pair $(s(x_i),t(x_i))$ in $(\mathbb{Q}.\mathbb{Z})^2$ is not a satisfying assignment for φ. Since si is an isomorphism between $(\mathbb{Q}.\mathbb{Z};succ)^2$ and $(\mathbb{Q}.\mathbb{Z};succ)$, we have that the assignment $si(s,t)$ does not satisfy ψ, which contradicts the fact that φ is preserved by si. $\square$

Clearly, every formula φ in CNF is equivalent to a reduced one, since we can repeatedly remove logically redundant literals until we obtain a reduced formula φ': in this case we say that we *obtain φ' from reducing φ.*

Lemma 24. *A binary relation $R \subseteq \mathbb{Z}^2$ is Horn definable over $(\mathbb{Z};succ)$ if and only if it is basic or has infinite distance-degree.*

Proof. The backward implication is clear, since a binary relation with infinite distance-degree and different from $\mathbb{Z}^2$ can be defined by a conjunction of literals of the form $\neg succ^p(x,y)$. Basic relations can be defined by a formula of the form $succ(x,x)$, $x=x$, or $succ^c(x,y)$, which are all Horn formulas.

Let us prove the forward implication. Let $\varphi(x,y)$ be a reduced Horn quantifier-free formula. In every clause of φ, there is at most one positive literal. Note that two negative literals cannot appear in the same clause of φ, for the disjunction $\neg succ^c(x,y) \vee \neg succ^d(x,y)$ is either trivial if $c \neq d$ or equivalent to a single literal if $c=d$, and φ is assumed to be reduced. Similarly, a positive literal and a negative literal cannot appear in the same clause, because $succ^c(x,y) \vee \neg succ^d(x,y)$ is equivalent to $\neg succ^d(x,y)$ if $c \neq d$, and is vacuously true if $c=d$. Therefore every clause of φ contains exactly one literal, so that φ is a conjunction of literals. If one of those literals is positive, φ is equivalent to $succ^c(x,y)$ for some c or defines the empty relation, so that the relation that φ defines is basic. Otherwise all the literals in φ are negative, and φ has infinite distance-degree. $\square$

Proposition 10. *Let Γ be a first-order expansion of $(\mathbb{Z};succ)$, and suppose that Γ pp-defines a relation that is not Horn-definable over $(\mathbb{Z};succ)$. Then Γ also pp-defines a binary relation that is not Horn-definable over $(\mathbb{Z};succ)$.*

Proof. Let R be a relation with a pp-definition in Γ that is not Horn-definable over $(\mathbb{Z};succ)$, and whose arity n is minimal among all the relations with the same property. We claim that for all $i,j \leq n$ and $p \in \mathbb{Z}$ the relation defined by the formula $R(x_1,\dots,x_n) \wedge succ^p(x_j,x_i)$ is Horn-definable over $(\mathbb{Z};succ)$. Otherwise, any reduced definition φ of this relation over $(\mathbb{Z};succ)$ has a clause ψ with at least two positive literals ℓ_1 and ℓ_2. Hence, there are satisfying assignment s_1 and s_2 for φ such that s_1 only satisfies ℓ_1 in ψ and s_2 only satisfies ℓ_2 in ψ. Let φ' be the formula obtained from φ by replacing literals of the form $succ^{p'}(x_j,x_k)$ or $succ^{-p'}(x_k,x_j)$, for $p' \in \mathbb{Z}$, by $succ^{p'-p}(x_i,x_k)$. Then the variable x_j no longer occurs in φ', and φ' is equivalent to $\exists x_j(R(x_1,\dots,x_n) \wedge succ^p(x_j,x_i))$. In particular, the restrictions of s_1 and s_2 to $\{x_1,\dots,x_n\} \setminus \{x_j\}$ are satisfying assignments for φ', and they witness that the literals ℓ_1 and ℓ_2 of φ (or the literals that correspond to those literals in φ') cannot be removed from φ'. Lemma 23 implies that φ' is not equivalent to a Horn formula. Note that φ' defines a relation of arity $n-1$ that is not Horn and that is pp-definable in Γ, a contradiction to the choice of R.

If a binary projection of R is non-basic and has finite distance-degree, then it is not Horn by Lemma 24 and we are done. If a binary projection of R is basic, then we have a contradiction to the minimality of n as we have seen above. So we can assume that the binary projections of R have infinite distance degree.

Suppose for contradiction that $n > 2$. Let $\varphi(x_1, \dots, x_n)$ be a reduced quantifier-free formula that defines R in $(\mathbb{Z}; succ)$ whose number of non-Horn clauses is minimal. We first prove that every non-Horn clause of φ is positive, i.e., consists of positive literals only. Pick a non-Horn clause ψ of φ with two positive literals ℓ_1, ℓ_2, and suppose ψ also contains the negative literal $\neg succ^p(x_j, x_i)$ for some $i, j \in \{1, \dots, n\}$ and $p \in \mathbb{Z}$. Since φ is reduced, there are satisfying assignment s_1 and s_2 for φ such that s_1 only satisfies ℓ_1 in ψ and s_2 only satisfies ℓ_2 in ψ; in particular, both s_1 and s_2 satisfy $succ^p(x_j, x_i)$. Then these two assignments show that both ℓ_1 and ℓ_2 cannot be removed when reducing $\varphi \wedge succ^p(x_j, x_i)$; by Lemma 23, this contradicts the fact that $\varphi \wedge succ^p(x_j, x_i)$ is equivalent to a Horn formula, which was established in the first paragraph of the proof.

Therefore, there exists a positive non-Horn clause ψ in φ. Let φ' denote the rest of φ, and define

$$E_{i,j} := \{s(x_j) - s(x_i) \mid s\colon \{x_1, \dots, x_n\} \to \mathbb{Z} \text{ satisfies } \varphi' \wedge \neg\psi\}.$$

If $E_{i,j}$ is empty for some $i, j \in \{1, \dots, n\}$, then the formulas φ and φ' are equivalent. But φ' contains fewer non-Horn clauses than φ, contradicting the choice of φ. By the first paragraph, for all distinct i, j and $p \in E_{i,j}$, the formula $\varphi \wedge succ^p(x_i, x_j)$ is equivalent to a Horn formula, and by Lemma 23, it even *reduces* to a Horn formula. Note that since ψ is a positive clause, the only way to reduce $\varphi \wedge succ^p(x_i, x_j)$ to a Horn formula is to remove all literals in ψ but one. Also note that at least one literal of ψ must remain when reducing $\varphi \wedge succ^p(x_i, x_j)$ because we chose p from $E_{i,j}$. This means that there exists a literal $\ell_p^{i,j}$ of ψ such that

$$\varphi \wedge succ^p(x_i, x_j) \models \ell_p^{i,j}\,.$$

Let q be the qe-degree of φ. If $p \in E_{i,j}$ is greater than nq, then we may take $\ell_p^{i,j}$ to be $\ell_{nq+1}^{i,j}$, by the substitution lemma.

First consider the case that there are distinct i, j such that $E_{i,j}$ is finite. Then φ is equivalent over $(\mathbb{Z}; succ)$ to the formula

$$\chi := \varphi' \wedge \bigwedge_{p \in E_{i,j}} (succ^p(x_i, x_j) \Rightarrow \ell_p^{i,j}).$$

Indeed, φ implies χ directly from the hypotheses we have. Conversely, if s satisfies χ one of two cases occur. Either some $\ell_p^{i,j}$, for $p \in E_{i,j}$, is satisfied by s, and then s satisfies ψ and φ. Or we must have $s(x_j) \neq s(x_i) + p$ for every $p \in E_{i,j}$, i.e., $s(x_j) - s(x_i) \notin E_{i,j}$. Since s is known to satisfy φ', by definition of $E_{i,j}$ it must also satisfy ψ, whence we get that s satisfies φ. Note that χ contains fewer non-Horn clauses than φ, which contradicts the choice of φ.

Therefore, $E_{i,j}$ is not finite, and thus cofinite for all distinct $i, j \leq n$. We claim that φ has a satisfying assignment s such that $|s(x_i) - s(x_j)| > 2(n+1)q$ for all distinct $i, j \in \{1, \dots, n-1\}$. The binary projections of R all have infinite distance degree, so by the substitution lemma we find for each pair (i, j) such that $1 \leq i < j \leq n$ a satisfying assignment $s_{i,j}\colon \{x_1, \dots, x_n\} \to \mathbb{Q}.\mathbb{Z}$ for φ such that $s_{i,j}(x_i) - s_{i,j}(x_j) = \infty$. Also note that the $(n-1)$-projection R' of R onto $\{1, \dots, n-1\}$ is Horn, and hence preserved by si. Then for $s'\colon \{x_1, \dots, x_{n-1}\} \to \mathbb{Q}.\mathbb{Z}$ defined by $s'(x) := si(s_{1,2}(x), \dots si(s_{n-3,n-1}(x), s_{n-2,n-1}(x)) \dots)$ we have that $s'(x_i) - s'(x_j) = \infty$ for all distinct i, j, and that $(s'(x_1), \dots, s'(x_{n-1})) \in R'$. Since R' is a projection of R, we can extend s' to a satisfying assignment s'' for φ. Again using the substitution lemma, we obtain from s'' a satisfying assignment $s\colon \{x_1, \dots, x_n\} \to \mathbb{Z}$ for φ such that $|s(x_i) - s(x_j)| > 2(n+1)q$ for distinct $i, j \in \{1, \dots, n-1\}$, and this concludes the proof of the claim.

For ψ to be satisfied by s, there must exist an $i \in \{1, \ldots, n-1\}$ such that $|s(x_i) - s(x_n)| \leq q$, since ψ only contains positive literals of degree at most q. Let $k \in \{1, \ldots, n-1\}$ be different from i. Note that $|s(x_k) - s(x_i)| > nq$ and $|s(x_k) - s(x_n)| > nq$. Also note that s satisfies the literals $\ell^{k,i}_{nq+1}$ and $\ell^{k,n}_{nq+1}$ by the definition of $\ell^{k,i}_{nq+1}$ and $\ell^{k,n}_{nq+1}$. Then the literal $\ell^{k,i}_{nq+1}$ relates x_i and x_n, and so does the literal $\ell^{k,n}_{nq+1}$, because x_i and x_n are with respect to s the only variables that are able to satisfy a positive literal. Since all binary projections of R have infinite distance degree, there is a satisfying assignment t of φ such that $|t(x_i) - t(x_n)| > 2(n+1)q$. Either $|t(x_k) - t(x_i)| > nq$ or $|t(x_k) - t(x_n)| > nq$. In the first case, $\ell^{k,i}_{nq+1}$ must be satisfied by t. But $\ell^{k,i}_{nq+1}$ is a literal of the form $succ^{p_1}(x_n, x_i)$ for $|p_1| \leq q$, and $|t(x_i) - t(x_n)| > nq$, so t cannot satisfy $\ell^{k,i}_{nq+1}$. Similarly, in the second case, t must satisfy $\ell^{k,n}_{nq+1}$, which is impossible since this literal is of the form $succ^{p_2}(x_n, x_i)$ for $|p_2| \leq q$. We have reached a contradiction. Therefore, we must have $n = 2$, and R is the desired binary non-Horn relation with a pp-definition over Γ. $\square$

We can finally conclude the complexity classification for non-positive first-order expansions of $(\mathbb{Z}; succ)$.

Proposition 11. *Let Γ be a non-positive first-order expansion of $(\mathbb{Z}; succ)$. Then $\mathbb{Q}.\Gamma$ is preserved by si and $\mathrm{CSP}(\Gamma)$ is in P, or $\mathrm{CSP}(\Gamma)$ is NP-hard.*

Proof. If $\mathbb{Q}.\Gamma$ is preserved by si then $\mathrm{CSP}(\Gamma)$ is in P by Proposition 8. Otherwise, Proposition 7 implies that Γ has a non-Horn relation. By Proposition 10, a binary non-Horn relation is pp-definable in Γ. A binary relation which is definable over $(\mathbb{Z}; succ)$ but not Horn is non-basic and has finite distance degree, by Lemma 24. Hence, a non-basic binary relation T of finite distance degree is pp-definable in Γ.

By Lemma 22, there exists a non-basic binary relation N pp-definable in Γ and which has infinite distance degree. The relation defined by $N(x,y) \wedge N(y,x)$ in Γ is symmetric and has infinite distance degree, and is again pp-definable in Γ, so we will assume that N is already symmetric. Let a be the smallest positive integer such that $(0, b)$ is in N for all $b \geq a$. With $succ$ and pp-definition, we may assume that T contains $(0,0)$ and that N does not contain $(0,0)$. Then by repeatedly replacing T by the pp-definable relation $\{(x,y) \in \mathbb{Z}^2 \mid \exists z \in \mathbb{Z} : (x,z) \in T \wedge (z,y) \in T\}$ we may assume that $(0,b), (0,2b) \in T$ with $b \geq a$. Let G be the undirected graph whose vertices are the integers v such that $(0,v) \in T$, and where v and w are adjacent if $(v,w) \in N$. This graph has no loop and contains the triangle $\{0,b\}, \{b,2b\}, \{0,2b\}$, so that G is not bipartite and $\mathrm{CSP}(G)$ is NP-hard [15]. Furthermore, $\mathrm{CSP}(G)$ is polynomial-time reducible to $\mathrm{CSP}(\Gamma)$: if $\exists x_1, \ldots, x_n.\varphi$ is an instance of $\mathrm{CSP}(G)$, create an instance of $\mathrm{CSP}(\Gamma)$ by adding an existentially quantified variable z, and by adding the constraints $T(z, x_i)$ for all i. This instance is satisfiable if and only if the original instance is satisfiable in G, using the fact that the automorphism group of Γ is transitive. This proves that $\mathrm{CSP}(\Gamma)$ is NP-hard. $\square$

7.4 The Positive Case

We prove in this section that a positive first-order expansion Γ of $(\mathbb{Z}; succ)$ which is not preserved by any d-modular maximum or minimum has an NP-hard CSP. As in the non-positive case, an important step of the classification is to show that there exists a non-basic binary relation with a pp-definition in Γ.

Let R be a relation of arity n with a first-order definition φ over Γ. We say that R is *r-decomposable* if $\varphi(x_1, \ldots, x_n)$ is equivalent over Γ to

$$\bigwedge_{1 \leq i_1 < \cdots < i_{n-r} \leq n} \exists x_{i_1}, \ldots, x_{i_{n-r}}.\varphi(x_1, \ldots, x_n)\,.$$

The following lemma states that a positive first-order expansion of $(\mathbb{Z}; succ)$ which is not preserved by a modular maximum or minimum pp-defines a non-basic binary relation. It is a positive variant of Lemma 38 in [3], and its proof is essentially the same. Intuitively this is because in both cases the binary relations that are pp-definable in Γ have either finite distance degree or are $\mathbb{Z}^2$ (if Γ has finite distance degree this is immediate, and when Γ is positively definable in $(\mathbb{Z}; succ)$ this is the content of Lemma 22). For the sake of completeness, we reproduce the proof with the necessary adjustments.

Lemma 25. *Let Γ be a positive first-order expansion of $(\mathbb{Z}; succ)$ without a modular max or a modular min as polymorphism. Then there is a non-basic binary relation pp-definable in Γ which has a finite distance degree.*

Proof. The binary relations pp-definable in Γ are either basic, or non-basic and of finite distance degree, by the fact that Γ is positive and Lemma 22. Suppose for contradiction that all the binary relations with a pp-definition in Γ are basic.

If every relation S of Γ were 2-decomposable then Γ would be invariant under max: indeed, we assumed that the binary relations pp-definable in Γ are already pp-definable in $(\mathbb{Z}; succ)$, so that a 2-decomposable relation S already has a pp-definition in $(\mathbb{Z}; succ)$, and is thus preserved by max. Therefore, Γ contains a relation that is not 2-decomposable. This implies that, by projecting out coordinates from S, we can obtain a relation R of arity $r \geq 3$ which is not $(r-1)$-decomposable.

This implies, in particular, that there exists a tuple $(a_1, \ldots, a_r) \notin R$ such that for all $i \in \{1, \ldots, r\}$, there exists some integer p_i such that $(a_1, \ldots, p_i, \ldots, a_r) \in R$. By replacing R by the relation with the pp-definition

$$\exists y_1, \ldots, y_r \big(\bigwedge_{i \in [r]} (y_i = x_i + a_i) \wedge R(y_1, \ldots, y_r) \big)$$

we can further assume that $a_i = 0$ for all $i \in [r]$. We can also assume, w.l.o.g., that $p_1 \neq -p_2$ because $r \geq 3$.

Suppose that the arity of R is greater than 3, and consider the ternary relation $T(x_1, x_2, x_3)$ defined by $R(x_1, x_2, x_3, \ldots, x_3)$. If there is a $z \in \mathbb{Z}$ so that $R(0, 0, z, \ldots, z)$, then T would not be 2-decomposable, since $(0, 0, 0) \notin T$, although $(p_1, 0, 0), (0, p_2, 0)$, and $(0, 0, z)$ are all in T. This contradicts the minimality of the arity of R. If there is no such z then $\exists x_3.R(x_1, x_2, x_3, \ldots, x_3)$ defines a binary relation omitting $(0, 0)$ and containing $(0, -p_1)$ and $(0, p_2)$. This relation is binary, pp-definable in Γ, and non-basic, contradiction.

Suppose now that $r = 3$. We claim that every binary projection of R is $\mathbb{Z}^2$. Suppose otherwise that one such binary projection, say the one defined by $\exists x_1.R(x_1, x_2, x_3)$, is not of this form. By assumption, all binary relations with a pp-definition in Γ are basic, so this formula is equivalent to $x_3 = x_2 + p$ for some $p \in \mathbb{Z}$. Let $(a, b, c) \in \mathbb{Z}^3$ be such that

- (a, b) is in the projection of R onto $\{1, 2\}$,
- (a, c) is in the projection of R onto $\{1, 3\}$, and

- (b,c) is in the projection of R onto $\{2,3\}$ (i.e., $c = b + p$).

Since (a,b) is in the projection of R onto $\{1,2\}$, there exists $d \in \mathbb{Z}$ such that $(a,b,d) \in R$. Since the projection of R onto $\{2,3\}$ is basic we have $d = b + p = c$, so that (a,b,c) is in R. Hence, R is 2-decomposable, contradicting our assumptions. This shows the claim.

Let $\varphi(x_1,x_2,x_3)$ be a positive formula in reduced DNF defining R over $(\mathbb{Z}; \mathit{succ})$. This formula has at least two disjuncts, otherwise R would be pp-definable over $(\mathbb{Z}; \mathit{succ})$. Each disjunct contains at most two literals, because it suffices to describe only two distances between three variables to determine the type of a triple of integers. We claim that there is a disjunct in φ that consists of only one literal. If that was not the case, every disjunct would have two literals and would be equivalent to $\mathit{succ}^{p_i}(x_2,x_1) \wedge \mathit{succ}^{q_i}(x_3,x_1)$ for some $p_i, q_i \in \mathbb{Z}$. In this case, the formula $\exists x_2.\varphi(x_1,x_2,x_3)$ defines a binary relation with finite distance degree, contradicting the claim established in the previous paragraph. Furthermore, there are at least two such disjuncts: if there were only one, say $\mathit{succ}^{p}(x_2,x_1)$, the relation defined by $\exists x_3.\varphi(x_1,x_2,x_3)$ is binary and has a finite distance degree, a contradiction. Hence, there are at least two disjuncts in φ that contain only one literal. One of x_1,x_2,x_3 must appear twice in those literals, and we may assume by permuting the variables that it is x_1. Let us write these literals as $\mathit{succ}^{p}(x_2,x_1)$ and $\mathit{succ}^{q}(x_3,x_1)$, for $p,q \in \mathbb{Z}$. Then the formula $\exists x_3 \left(\varphi(x_1,x_2,x_3) \wedge \mathit{succ}^{p-q+1}(x_2,x_3)\right)$ is equivalent to a binary DNF which is reduced and contains the two disjuncts $\mathit{succ}^{p}(x_2,x_1)$ and $\mathit{succ}^{p+1}(x_2,x_1)$. The relation defined by this formula has finite distance degree, again contradicting our assumptions.

It follows that there exists a non-basic binary relation pp-definable in Γ, and this relation has finite distance degree by positivity of Γ. ☐

The following is Lemma 43 in [3].

Lemma 26. *Let $S \subset \mathbb{Z}$ be finite with $|S| > 1$, and let d be the greatest common divisor of all $a - a'$ for $a, a' \in S$. Then for any d-progression T, the relation Diff_T is pp-definable in $(\mathbb{Z}; \mathit{succ}, \mathrm{Diff}_S)$.*

Proposition 12. *Let Γ be a first-order expansion of $(\mathbb{Z}; \mathit{succ})$, and $S \subset \mathbb{Z}$ a 1-progression with $|S| > 1$, such that Diff_S is pp-definable in Γ. Then Γ is preserved by max or min; or $\mathrm{CSP}(\Gamma)$ is NP-hard.*

Proof. Suppose that Γ is not preserved by max nor min. Therefore, there exist in Γ a relation $R \subseteq \mathbb{Z}^n$ that is not preserved by max and a relation $T \subseteq \mathbb{Z}^m$ which is not preserved by min. This means that there are tuples $\overline{a}, \overline{b}$ in R such that $\max(\overline{a}, \overline{b})$ is not in R and similarly for T. By hypothesis and Lemma 26, all the 1-progressions are definable in Γ. Let M be $\max_{i,j}\{|a_i - a_j|, |b_i - b_j|\}$, and let φ be the pp-definition of $\mathrm{Dist}_{[0,M|1]}$ in Γ. Define the relation R^* by $R(x_1,\ldots,x_n) \wedge \bigwedge_{i,j \leq n} \varphi(x_i,x_j)$ and analogously define T^* from T. We have that $\overline{a}, \overline{b} \in R^*$ by construction, and still $\max(\overline{a}, \overline{b}) \notin R^*$ since $R^* \subseteq R$. Also note that R^* has finite distance degree. Analogously, S^* is not preserved by min and has finite distance degree. It follows from Proposition 9 that $\mathrm{CSP}(\mathbb{Z}; \mathit{succ}, \mathrm{Diff}_S, R^*, T^*)$ is NP-hard. Therefore, $\mathrm{CSP}(\Gamma)$ is also NP-hard. ☐

We can now prove the complexity classification for positive first-order expansions of $(\mathbb{Z}; \mathit{succ})$.

Proposition 13. *Let Γ be a positive first-order expansion of $(\mathbb{Z}; \mathit{succ})$. Then Γ is preserved by a modular max or a modular min, and $\mathrm{CSP}(\Gamma)$ is in P, or $\mathrm{CSP}(\Gamma)$ is NP-hard.*

Proof. If Γ is preserved by a modular max or a modular min, then $\mathrm{CSP}(\Gamma)$ is in P by Theorem 9, so assume that this is not the case. Lemma 25 implies there exists a non-basic binary relation R

with finite distance degree and a pp-definition in Γ. Lemma 44 in [3] states that if $S \subset \mathbb{Z}$ is finite, but not a d-progression, for any $d > 0$, then $\mathrm{CSP}(\mathbb{Z}; succ, \mathrm{Diff}_S)$ is NP-hard. Hence, if R is not a d-progression for any $d \geq 1$, then $\mathrm{CSP}(\Gamma)$ is NP-hard. So let us assume that R is a d-progression for some $d \geq 1$.

Since Γ is not preserved by $\max_d$, it contains a relation T_1 containing tuples $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ such that $(\max_d(a_1, b_1), \ldots, \max_d(a_n, b_n)) \notin T_1$. Since Γ is positive, the binary projections of T_1 have finite distance degree by Lemma 22. By the same argument as above, we can suppose that these binary projections are arithmetic progressions unless $\mathrm{CSP}(\Gamma)$ is NP-hard. Lemma 45 in [3] establishes that $\mathrm{CSP}(\Gamma)$ is NP-hard unless all these arithmetic progressions are d-progressions. Using $succ$, we easily see that T_1 pp-defines a relation T_1' that is not preserved by $\max_d$ and such that all the differences $a_i - a_j$, for $(a_1, \ldots, a_n) \in T_1'$, are divisible by d. Therefore we can pick two tuples $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ in T_1' whose entries are divisible by d, and such that $(\max_d(a_1, b_1), \ldots, \max_d(a_n, b_n)) \notin T_1'$. Similarly, we obtain a relation T_2' which has a pp-definition in Γ and which contains tuples $(c_1, \ldots, c_m), (d_1, \ldots, d_m)$ whose entries are all divisible by d and such that $(\min_d(c_1, d_1), \ldots, \min_d(c_m, d_m))$ is not in T_2'. Let Γ' be $(\mathbb{Z}; succ^d, R, T_1', T_2')$. It follows from our construction that Γ'/d is not preserved by max nor by min. Moreover Γ' contains the non-basic d-progression R, so that Γ'/d contains a non-basic 1-progression. By Proposition 12, we obtain that $\mathrm{CSP}(\Gamma'/d)$ is NP-hard.

Now we reduce $\mathrm{CSP}(\Gamma'/d)$ to $\mathrm{CSP}(\Gamma)$ to prove that the latter is also NP-hard. By Lemma 43 in [3], the relation $\mathrm{Dist}_{[0,d|d]}$ has a pp-definition in Γ. Let q be the qe-degree of Γ and note that an instance of Γ on n variables has a solution iff it has a solution on the interval $[0, qn]$. From an instance Φ of $\mathrm{CSP}(\Gamma'/d)$ we build an instance Ψ of $\mathrm{CSP}(\Gamma)$. The variables of Ψ consist of the variables of Φ and additionally $qn - 1$ new variables $x_1, \ldots, x_{qn-1}$ for each extant variable x of Φ, and finally an additional new variable z. The constraints of Ψ are as follows:

- for each constraint of Φ using the relations $succ^d, R, T_1'$, and T_2', we add a constraint to Ψ using the pp-definitions of these relations in Γ,
- for each variable x of Φ, Ψ contains the constraint $\mathrm{Dist}_{[0,qdn|d]}(x, z)$, that we define by the conjunction $\mathrm{Dist}_{[0,d|d]}(x, x_1) \wedge \mathrm{Dist}_{[0,d|d]}(x_1, x_2) \wedge \ldots \wedge \mathrm{Dist}_{[0,d|d]}(x_{qn-1}, z)$.

It is straightforward to see that $\Gamma/d \models \Phi$ iff $\Gamma \models \Psi$ and the result follows. □

7.5 Concluding the Classification

We can finally combine the structural classification of first-order reducts of $(\mathbb{Z}; <)$ (Theorem 4) with the complexity classification of the previous sections.

Theorem 10. *Let Γ be a reduct of $(\mathbb{Z}; <)$ with finite signature. Then there exists a structure Δ such that $\mathrm{CSP}(\Delta)$ equals $\mathrm{CSP}(\Gamma)$ and one of the following cases applies.*

1. *Δ has a finite domain, and the CSP for Γ is conjectured to be in P or NP-complete [13].*
2. *Δ is a reduct of $(\mathbb{Q}; <)$ and $\mathrm{CSP}(\Delta)$ is either in P or NP-complete [5].*
3. *Δ is a reduct of $(\mathbb{Z}; <)$ and preserved by max or by min. In this case, $\mathrm{CSP}(\Delta)$ is in P.*
4. *Δ is a reduct of $(\mathbb{Z}; succ)$ such that Δ is preserved by a modular max or min, or $\mathbb{Q}.\Delta$ is preserved by a binary injective function preserving succ. In this case, $\mathrm{CSP}(\Delta)$ is in P.*

5. CSP(Γ) *is NP-complete.*

Proof. Let Γ be a finite signature reduct of $(\mathbb{Z}; <)$. By Proposition 1, CSP(Γ) is in NP. If Γ is homomorphically equivalent to a finite structure, we are in case (1) of the statement and there is nothing to be shown. Otherwise, Theorem 4 implies that there exists a reduct Δ of $(\mathbb{Z}; <)$ such that CSP(Γ) equals CSP(Δ), and one of the following cases applies.

(a) Δ is a reduct of $(\mathbb{Q}; <)$. We are in case (2) of the statement; the complexity of CSP(Δ) has been classified in Theorem 50 in [5].

(b) For all $k \geq 1$, the relation $\mathrm{Dist}_{\{k\}}$ is pp-definable; in this case, CSP(Γ) and CSP(Δ) are NP-hard by Proposition 2. Hence, we are in case (5) of the statement.

(c) The relation *succ* is pp-definable in Δ. If $<$ is pp-definable in Δ, then Lemma 21 and Theorem 9 imply that we are in case (3) or (5) of the statement. Otherwise Δ is a reduct of $(\mathbb{Z}; succ)$, by Lemma 20. If Δ is non-positive then the statement follows from Proposition 11, if it is positive then the statement follows from Proposition 13.

□

8 Discussion and Future Work

In this article, we proved that for finitely many relations $R_1, \dots, R_k$ that are first-order definable over $(\mathbb{Z}; <)$, the constraint satisfaction problem for $\Gamma = (\mathbb{Z}; R_1, \dots, R_k)$ satisfies a trichotomy: it is equivalent to the CSP of a finite structure, in P, or NP-complete. In the case that Γ contains the successor relation (we showed that the complexity classification can be reduced to this situation), the trichotomy has an elegant algebraic formulation:

- either $\mathbb{Q}.\Gamma$ has an endomorphism with finite range, or
- $\mathbb{Q}.\Gamma$ has a modular maximum/minimum polymorphism or a binary injective polymorphism and the CSP of Γ is in P, or
- $\mathbb{Q}.\Gamma$ omits these polymorphisms and the CSP of Γ is NP-complete.

These results are important foundations for the future investigation of the complexity of CSPs for constraint languages that are definable in Presburger arithmetic, i.e., definable over $(\mathbb{Z}; +, <)$. We mention here two possible classification projects that can improve our understanding of the complexity of problems expressible in Presburger arithmetic, as well as an interesting connection to open problems in other areas.

We believe that the techniques we employed can be developed further in order to apply them for first-order reducts of $(\mathbb{Z}; succ, 0)$. The principal difference is that the automorphism group of $(\mathbb{Z}; <)$ is transitive, while $(\mathbb{Z}; succ, 0)$ is *rigid*, i.e., it has no automorphism besides the identity function. Its countable saturated extension, however, has non-trivial automorphisms. Note that every integer is first-order definable in $(\mathbb{Z}; succ, 0)$. The complexity of the CSPs of first-order reducts of $(\mathbb{Z}; 0, 1, -1, \dots)$ have been classified recently [6]. An interesting question is whether this classification can be employed together with result from the present article to obtain a complexity classification for the reducts of $(\mathbb{Z}; succ, 0)$. We believe that the techniques of the present article (rather than the techniques from [6] which are tied to the realm of ω-categoricity) are the basis for solving this question. More generally, we believe that the way in which we use the universal-algebraic approach for sufficiently saturated models of the template can be applied for even larger classes of constraints over the integers.

Secondly, a result for Presburger arithmetic would in particular give a complexity classification for the CSPs of reducts of $(\mathbb{Z}; +, 1)$. This class of CSPs captures many well-known problems, such as the feasibility of linear diophantine equations. Note that both classes above contain all finite-domain CSPs.

Finally, we note that the important Max-Atom problem [1] can be formulated as the CSP for a first-order reduct of $(\mathbb{Z}; <)$ whose signature is *infinite*. In order to define the computational problem $\mathrm{CSP}(\Gamma)$ for a structure with an infinite relational signature we have to discuss how the relation symbols of Γ are represented in the input. In this article, we have only studied $\mathrm{CSP}(\Gamma)$ for finite-signature structures Γ, and there the choice of the representation of the relation symbols in Γ does not affect the computational complexity of the problem. This changes for infinite signatures: indeed, if we represent a relation symbol R in a first-order reduct Γ of $(\mathbb{Z}; <)$ by the first-order formula that defines R^Γ, we can no longer expect polynomial-time algorithms for $\mathrm{CSP}(\Gamma)$ since already the problem to decide whether a single constraint in the input is satisfiable becomes PSPACE-complete. However, for first-order reducts of $(\mathbb{Z}; <)$ with infinite signature there is a natural candidate for an input encoding of the relation symbols of Γ that still allows for polynomial-time algorithms for $\mathrm{CSP}(\Gamma)$, which we describe in the following. Each constraint $R(x_1, \ldots, x_k)$ is represented by a quantifier-free definition of R over $(\mathbb{Z}; <)$ in reduced disjunctive normal form, where a literal $x_i \leq x_j + k$ is encoded by giving k in binary representation. When the input is represented in this way, $\mathrm{CSP}(\Gamma)$ is still in NP, by the same argument as in Proposition 1.

The Max-Atom problem is the CSP for the first-order reduct of $(\mathbb{Z}; <)$ that contains all the relations of the form

$$\{(x, y, z) \in \mathbb{Z}^3 \mid x \leq y + p \vee x \leq z + p\}.$$

Many decision problems reduce in polynomial time to Max-Atom: this is for example the case of mean-payoff games [12] (which are in fact polynomial-time equivalent to Max-Atom [21]), parity games, and the model-checking problem for the modal μ-calculus [19]. The precise complexity of the Max-Atom problem is still unknown: it is known to be in NP $\cap$ coNP, but not known to be in P. Note that if the constants p in the Max-Atom constraints are encoded in unary, then there is a simple reduction of the Max-Atom problem to a discrete temporal CSP (which is max-closed and with finite signature; also see Example 1, (6)): the max-atom constraint $x \leq \max(y, z) + p$ is equivalent to

$$\exists x_1, \ldots, x_p \big(x = succ(x_1) \wedge \cdots \wedge x_{p-1} = succ(x_p) \wedge x_p \leq \max(y, z)\big).$$

The hardness proofs in this article can be used even for infinite-signature reducts Γ of $(\mathbb{Z}; <)$: for any structure Γ' obtained by keeping finitely many relations from Γ, there is a trivial polynomial-time reduction from $\mathrm{CSP}(\Gamma')$ to $\mathrm{CSP}(\Gamma)$. Hence, Theorem 3 implies that if Γ contains the successor relation, and if $\max_d$, $\min_d$, and si are not polymorphisms of Γ or $\mathbb{Q}.\Gamma$, then $\mathrm{CSP}(\Gamma)$ is NP-hard. In particular, if Γ contains additionally the relation $<$, then $\mathrm{CSP}(\Gamma)$ is NP-hard unless Γ is preserved by max or min.

Acknowledgements

The authors express their gratitude to the referees for their thorough work.

References

[1] Marc Bezem, Robert Nieuwenhuis, and Enric Rodríguez-Carbonell. The max-atom problem and its relevance. In *Proceedings of the International Conference on Logic for Programming, Artificial Intelligence, and Reasoning (LPAR)*, pages 47–61, 2008.

[2] Manuel Bodirsky. Complexity classification in infinite-domain constraint satisfaction. Mémoire d'habilitation à diriger des recherches, Université Diderot – Paris 7. Available at arXiv:1201.0856, 2012.

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

[4] Manuel Bodirsky, Martin Hils, and Barnaby Martin. On the scope of the universal-algebraic approach to constraint satisfaction. *Logical Methods in Computer Science (LMCS)*, 8(3:13), 2012. An extended abstract that announced some of the results appeared in the proceedings of Logic in Computer Science (LICS'10).

[5] Manuel Bodirsky and Jan Kára. The complexity of temporal constraint satisfaction problems. *Journal of the ACM*, 57(2):1–41, 2009. An extended abstract appeared in the Proceedings of the Symposium on Theory of Computing (STOC).

[6] Manuel Bodirsky and Antoine Mottet. Reducts of finitely bounded homogeneous structures, and lifting tractability from finite-domain constraint satisfaction. In *Proceedings of the 31th Annual IEEE Symposium on Logic in Computer Science – LICS'16*, pages 623–632, 2016. Preprint available under ArXiv:1601.04520.

[7] Manuel Bodirsky and Jaroslav Nešetřil. Constraint satisfaction with countable homogeneous templates. *Journal of Logic and Computation*, 16(3):359–373, 2006.

[8] Manuel Bodirsky and Michael Pinsker. Schaefer's theorem for graphs. *Journal of the ACM*, 62(3):52 pages (article number 19), 2015. A conference version appeared in the Proceedings of STOC 2011, pages 655–664.

[9] V. G. Bodnarčuk, L. A. Kalužnin, V. N. Kotov, and B. A. Romov. Galois theory for Post algebras, part I and II. *Cybernetics*, 5:243–539, 1969.

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

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

[12] Andrzej Ehrenfeucht and Jan Mycielski. Positional strategies for mean payoff games. *International Journal of Game Theory*, 8(2):109–113, 1979.

[13] Tomás Feder and Moshe Y. 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.

[14] David Geiger. Closed systems of functions and predicates. *Pacific Journal of Mathematics*, 27:95–100, 1968.

[15] Pavol Hell and Jaroslav Nešetřil. On the complexity of H-coloring. *Journal of Combinatorial Theory, Series B*, 48:92–110, 1990.

[16] Wilfrid Hodges. *A shorter model theory*. Cambridge University Press, Cambridge, 1997.

[17] Peter Jeavons, David Cohen, and Marc Gyssens. Closure properties of constraints. *Journal of the ACM*, 44(4):527–548, 1997.

[18] Peter Jonsson and Tomas Lööw. Computation complexity of linear constraints over the integers. *Artificial Intelligence*, 195:44–62, 2013.

[19] Dexter Kozen. Results on the propositional μ-calculus. *Theoretical Computer Science*, 27:333–354, 1983.

[20] A. K. Mackworth. Consistency in networks of relations. *Artificial Intelligence*, 8:99–118, 1977.

[21] Rolf H. Möhring, Martin Skutella, and Frederik Stork. Scheduling with and/or precedence constraints. *SIAM Journal on Computing*, 33(2):393–415, 2004.

[22] Arash Rafiey, Jeff Kinne, and Tomás Feder. Dichotomy for Digraph Homomorphism Problems. arXiv:1701.02409, 2017.

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