

# As Close as It Gets

Mike Behrisch<sup>1</sup>, Miki Hermann<sup>2</sup>, Stefan Mengel<sup>2</sup>, and Gernot Salzer<sup>1</sup>(✉)

<sup>1</sup> Technische Universität Wien, Vienna, Austria

{behrisch,salzer}@logic.at

<sup>2</sup> LIX (UMR CNRS 7161), École Polytechnique, Palaiseau, France

hermann@lix.polytechnique.fr, mengel@cril.fr

**Abstract.** We study the minimum Hamming distance between distinct satisfying assignments of a conjunctive input formula over a given set of Boolean relations (`MinSolutionDistance`, `MSD`). We present a complete classification of the complexity of this optimization problem with respect to the relations admitted in the formula. We give polynomial time algorithms for several classes of constraint languages. For all other cases we prove hardness or completeness with respect to poly-APX, or NPO, or equivalence to a well-known hard optimization problem.

## 1 Introduction

We study the following optimization problem related to Boolean constraint satisfaction problems (CSPs): Given a formula built from atomic constraint relations by means of conjunction and variable identification, the task is to produce two satisfying assignments having minimal Hamming distance among all distinct pairs in the solution space of the CSP instance represented by the formula (`MinSolutionDistance`, `MSD`). Note that the dual problem `MaxHammingDistance` has been studied in [9].

As usual our problem is parametrized by the set of atomic constraints allowed to occur in the conjunctive formulas. With respect to this parametrization we completely classify the complexity of the minimization problem `MSD`: It turns out that it is either polynomial-time solvable, or that it is complete for a well-known optimization class, or else equivalent to some classical hard optimization problem.

Restricting the allowed relations to affine Boolean relations, our problem `MSD` becomes the well-known problem `MinDistance` of computing the minimum distance of a linear code. As this quantity determines the number of errors such a code can detect and correct, it is of central importance in coding theory. Our work can thus be seen as a generalization of these questions from affine to arbitrary relations.

---

M. Behrisch and G. Salzer— Supported by Austrian Science Fund (FWF) grant I836-N23.

M. Hermann— Supported by ANR-11-ISO2-003-01 Blanc International grant ALCOCLAN.

S. Mengel— Supported by QUALCOMM grant. Now at CRIL (UMR CNRS 8188), Lens, France.

In the course of investigations it appears that MSD lacks compatibility with existential quantification, preventing classical clone theory from being applicable. Consequently, we are lead to weak co-clones that need only be closed under conjunction and equality. To deal with such structures we make use of the theory established in [15], as well as the minimal weak bases of Boolean co-clones described in [13].

This paper is part of a more general program to understand the Hamming distance between solutions of constraint satisfaction problems. The results of this program up to now, including those from this paper and some on other problems, can be found in [4, 5].

## 2 Preliminaries

An  $n$ -ary *Boolean relation*  $R$  is a subset of  $\{0, 1\}^n$ ; its elements  $(b_1, \dots, b_n)$  are also written as  $b_1 \cdots b_n$ . Let  $\mathcal{V}$  be a set of variables. An *atomic constraint*, or an *atom*, is an expression  $R(\mathbf{x})$ , where  $R$  is an  $n$ -ary relation and  $\mathbf{x}$  is an  $n$ -tuple of variables from  $\mathcal{V}$ . Let  $\mathcal{L}$  be the collection of all non-empty finite sets of Boolean relations, also called *constraint languages*. A (conjunctive)  $\Gamma$ -*formula* is a finite conjunction of atoms  $R_1(\mathbf{x}_1) \wedge \cdots \wedge R_k(\mathbf{x}_k)$ , where the  $R_i$  are relations from  $\Gamma \in \mathcal{L}$  and the  $\mathbf{x}_i$  are variable tuples of suitable arity.

An *assignment* is a mapping  $m: \mathcal{V} \rightarrow \{0, 1\}$  assigning a Boolean value  $m(x)$  to each variable  $x \in \mathcal{V}$ . If we arrange the variables in some arbitrary but fixed order, say as a vector  $(x_1, \dots, x_n)$ , then the assignments can be identified with vectors from  $\{0, 1\}^n$ . The  $i$ -th component of a vector  $m$  is denoted by  $m[i]$  and corresponds to the value of the  $i$ -th variable, i.e.,  $m[i] = m(x_i)$ . The *Hamming weight*  $\text{hw}(m) = |\{i \mid m[i] = 1\}|$  of  $m$  is the number of 1s in the vector  $m$ . The *Hamming distance*  $\text{hd}(m, m') = |\{i \mid m[i] \neq m'[i]\}|$  of  $m$  and  $m'$  is the number of coordinates on which the vectors disagree. The complement  $\bar{m}$  of a vector  $m$  is its pointwise complement,  $\bar{m}[i] = 1 - m[i]$ .

**Table 1.** Some relevant Boolean co-clones with bases

iD <sub>2</sub>	$\{x \oplus y, x \rightarrow y\}$	iN	$\{\text{dup}^3\}$
iL	$\{\text{even}^4\}$	iN <sub>2</sub>	$\{\text{nae}^3\}$
iL <sub>2</sub>	$\{\text{even}^4, \neg x, x\}$	iI	$\{\text{even}^4, x \rightarrow y\}$
iV <sub>2</sub>	$\{x \vee y \vee \neg z, \neg x, x\}$	iI <sub>0</sub>	$\{\text{even}^4, x \rightarrow y, \neg x\}$
iE <sub>2</sub>	$\{\neg x \vee \neg y \vee z, \neg x, x\}$	iI <sub>1</sub>	$\{\text{even}^4, x \rightarrow y, x\}$

An assignment  $m$  satisfies the constraint  $R(x_1, \dots, x_n)$  if  $(m(x_1), \dots, m(x_n)) \in R$  holds. It satisfies the formula  $\varphi$  if it satisfies all of its atoms;  $m$  is said to be a *model* or *solution* of  $\varphi$  in this case. We use  $[\varphi]$  to denote the Boolean relation containing all models of  $\varphi$ . In sets of relations represented this

way we usually omit the brackets. A *literal* is a variable  $v$ , or its negation  $\neg v$ . Assignments  $m$  are extended to literals by defining  $m(\neg v) = 1 - m(v)$ .

The following Boolean functions and relations are of particular relevance to us: we write  $x \oplus y$  for addition modulo 2 and  $x \equiv y$  for  $x \oplus y \oplus 1$ . Further, we define the relations  $\text{nae}^3 := \{0, 1\}^3 \setminus \{000, 111\}$ ,  $\text{dup}^3 := \{0, 1\}^3 \setminus \{010, 101\}$  and  $\text{even}^4 := \{(a_1, a_2, a_3, a_4) \in \{0, 1\}^4 \mid \bigoplus_{i=1}^4 a_i = 0\}$ .

Throughout the text we refer to different types of Boolean constraint relations following Schaefer’s terminology [14] (see also [6, 8]). A Boolean relation  $R$  is (1) *1-valid* if  $1 \cdots 1 \in R$  and it is *0-valid* if  $0 \cdots 0 \in R$ , (2) *Horn* (*dual Horn*) if  $R$  can be represented by a formula in conjunctive normal form (CNF) having at most one unnegated (negated) variable in each clause, (3) *monotone* if it is both Horn and dual Horn, (4) *bijunctive* if it can be represented by a CNF having at most two variables in each clause, (5) *affine* if it can be represented by an affine system of equations  $Ax = b$  over  $\mathbb{Z}_2$ , (6) *complementive* if for each  $m \in R$  also  $\bar{m} \in R$ . A set  $\Gamma$  of Boolean relations is called 0-valid (1-valid, Horn, dual Horn, monotone, affine, bijunctive, complementive) if *every* relation in  $\Gamma$  has the respective property.

We denote by  $\langle \Gamma \rangle$  the set of all relations that can be expressed using relations from  $\Gamma \cup \{=\}$ , conjunction, variable identification (and permutation), cylindrification and existential quantification. The set  $\langle \Gamma \rangle$  is called the *co-clone* generated by  $\Gamma$ . A *base* of a co-clone  $\mathcal{B}$  is a set of relations  $\Gamma$  such that  $\langle \Gamma \rangle = \mathcal{B}$ . The set of all co-clones constitutes a lattice with regard to set inclusion. Their bases were studied in [7]; those relevant in this paper are listed in Table 1. In particular the sets of relations being 0-valid, 1-valid, complementive, Horn, dual Horn, affine, and bijunctive each form a co-clone denoted by  $\text{iI}_0$ ,  $\text{iI}_1$ ,  $\text{iN}_2$ ,  $\text{iE}_2$ ,  $\text{iV}_2$ ,  $\text{iL}_2$ , and  $\text{iD}_2$ , respectively.

We will also use a weaker closure than  $\langle \Gamma \rangle$ , called *conjunctive closure* and denoted by  $\langle \Gamma \rangle_\wedge$ , where the constraint language  $\Gamma$  is closed under conjunctive definitions, but not under existential quantification or addition of explicit equality constraints.

*Minimal weak bases* of co-clones are bases with certain additional properties. Since we rely on only some of them, we shall not define this notion but refer the reader to [13, 15].

**Theorem 1 (Schnoor & Schnoor [15]).** *If  $\Gamma$  is a minimal weak base of a co-clone, then  $\Gamma \subseteq \langle \Gamma' \rangle_\wedge$  for any base  $\Gamma'$ .*

Lagerkvist computed minimal weak bases for all Boolean co-clones in [13]. From there we infer that  $\{\text{even}_4(x_1, x_2, x_3, x_4) \wedge (x_1 \wedge x_4 \equiv x_2 \wedge x_3)\}$  constitutes a minimal weak base of the co-clone  $\text{iN}$ ; likewise  $\{[(x_1 \equiv x_2 \wedge x_3) \wedge (\neg x_4 \equiv \neg x_2 \wedge \neg x_3)]\}$  is one of  $\text{iI}$ .

We assume that the reader has a basic knowledge of approximation algorithms and complexity theory, see e.g. [2, 8]. For reductions among decision problems we use polynomial-time many-one reduction denoted by  $\leq_m$ . Many-one equivalence between decision problems is written as  $\equiv_m$ . For reductions among optimization problems we employ approximation preserving reductions (AP-reductions), represented by  $\leq_{\text{AP}}$ , while AP-equivalence of optimization problems

is stated as  $\equiv_{AP}$ . Besides, the following approximation complexity classes in the hierarchy  $PO \subseteq APX \subseteq \text{poly-APX} \subseteq NPO$  occur.

An optimization problem  $\mathcal{P}_1$  AP-reduces to another optimization problem  $\mathcal{P}_2$  if there are two polynomial-time computable functions  $f, g$ , and a constant  $\alpha \geq 1$  such that for all  $r > 1$  on any input  $x$  for  $\mathcal{P}_1$  the following holds:

- $f(x)$  is an instance of  $\mathcal{P}_2$ ;
- for any solution  $y$  of  $f(x)$ , the result  $g(x, y)$  is a solution of  $x$ ;
- if  $y$  is an  $r$ -approximate solution for the instance  $f(x)$ , then the solution  $g(x, y)$  is  $(1 + (r - 1)\alpha + o(1))$ -approximate for  $x$ .

If  $\mathcal{P}_1$  AP-reduces to  $\mathcal{P}_2$  with constant  $\alpha \geq 1$  and  $\mathcal{P}_2$  has an  $f(n)$ -approximation algorithm, then there is an  $\alpha f(n)$ -approximation algorithm for  $\mathcal{P}_1$ .

To relate our problem to well-known optimization problems we make the following convention: For optimization problems  $\mathcal{P}$  and  $\mathcal{Q}$  we say that  $\mathcal{P}$  is  $\mathcal{Q}$ -complete if  $\mathcal{P} \equiv_{AP} \mathcal{Q}$ . We use this notion in particular with respect to the following well-studied problem.

**Problem MinDistance.** Given a matrix  $A \in \mathbb{Z}_2^{k \times l}$  any non-zero vector  $x \in \mathbb{Z}_2^l$  with  $Ax = 0$  is considered a solution. The objective is to minimize the Hamming weight  $\text{hw}(x)$ .

MinDistance is known to be NP-hard to approximate within a factor  $2^{\Omega(\log^{1-\varepsilon}(n))}$  for every  $\varepsilon > 0$ , see [10]. Thus if a problem  $\mathcal{P}$  is equivalent to it, then  $\mathcal{P} \notin APX$  unless  $P = NP$ .

We also use the classic satisfiability problem  $\text{SAT}(\Gamma)$ , asking for a conjunctive formula  $\varphi$  over a constraint language  $\Gamma$ , if  $\varphi$  is satisfiable. Schaefer presented in [14] a complete classification of complexity for SAT. His dichotomy theorem proves that  $\text{SAT}(\Gamma)$  is polynomial-time decidable if  $\Gamma$  is 0-valid ( $\Gamma \subseteq iI_0$ ), 1-valid ( $\Gamma \subseteq iI_1$ ), Horn ( $\Gamma \subseteq iE_2$ ), dual Horn ( $\Gamma \subseteq iV_2$ ), bijunctive ( $\Gamma \subseteq iD_2$ ), or affine ( $\Gamma \subseteq iL_2$ ); otherwise it is NP-complete. Moreover, we need the decision problem  $\text{AnotherSAT}(\Gamma)$ , whose complexity was completely classified in [12]. Given a conjunctive formula  $\varphi$  and a satisfying assignment  $m$ , it asks if there exists another satisfying assignment  $m' \neq m$  for  $\varphi$ .

### 3 Results

The input to our problem is a conjunctive formula over a constraint language. The satisfying assignments of the formula, being its solutions, constitute the codewords of the associated code. The minimization target is the distance of any two distinct solutions.

**Problem MinSolutionDistance( $\Gamma$ ),  $\text{MSD}(\Gamma)$**

*Input:* A conjunctive formula  $\varphi$  over relations from  $\Gamma$ .

*Solution:* Two satisfying truth assignments  $m \neq m'$  to the variables occurring in  $\varphi$ .

*Objective:* Minimum Hamming distance  $\text{hd}(m, m')$ .

**Theorem 2.** *For any constraint language  $\Gamma$  the optimization problem  $\text{MSD}(\Gamma)$  is*

- (i) *in PO if  $\Gamma$  is*
  - (a) *bijunctive ( $\Gamma \subseteq \text{iD}_2$ ) or*
  - (b) *Horn ( $\Gamma \subseteq \text{iE}_2$ ) or*
  - (c) *dual Horn ( $\Gamma \subseteq \text{iV}_2$ );*
- (ii) *MinDistance-complete if  $\Gamma$  is exactly affine ( $\text{iL} \subseteq \langle \Gamma \rangle \subseteq \text{iL}_2$ );*
- (iii) *in poly-APX if  $\Gamma$  is both 0-valid and 1-valid, but does not contain an affine relation ( $\text{iN} \subseteq \langle \Gamma \rangle \subseteq \text{iI}$ ), where  $\text{MSD}(\Gamma)$  is  $n$ -approximable but not  $(n^{1-\varepsilon})$ -approximable unless  $\text{P} = \text{NP}$ ; and*
- (iv) *NPO-complete otherwise ( $\text{iN}_2 \subseteq \langle \Gamma \rangle$  or  $\text{iI}_0 \subseteq \langle \Gamma \rangle$  or  $\text{iI}_1 \subseteq \langle \Gamma \rangle$ ).*

*Proof.* The proof is split into several propositions presented in the remainder of the paper.

- (i) See Propositions 7 and 8.
- (ii) See Proposition 14.
- (iii) For  $\Gamma \subseteq \text{iI}$ , every formula  $\varphi$  over  $\Gamma$  has at least two solutions since it is both 0-valid and 1-valid. Thus  $2\text{SolutionSAT}(\Gamma)$  is in P, and Proposition 13 yields that  $\text{MSD}(\Gamma)$  is  $n$ -approximable. By Proposition 18 this approximation is indeed tight.
- (iv) According to [12],  $\text{AnotherSAT}(\Gamma)$  is NP-hard for  $\text{iI}_0 \subseteq \langle \Gamma \rangle$ , or  $\text{iI}_1 \subseteq \langle \Gamma \rangle$ . By Lemma 10 it follows that  $2\text{SolutionSAT}(\Gamma)$  is NP-hard, too. For  $\text{iN}_2 \subseteq \langle \Gamma \rangle$  we can reduce the NP-hard problem  $\text{SAT}(\Gamma)$  to  $2\text{SolutionSAT}(\Gamma)$ . Hence  $\text{MSD}(\Gamma)$  is NPO-complete in all three cases.  $\square$

The optimization problem can be transformed into a decision problem  $\text{MSD}^d$  by adding a bound  $k \in \mathbb{N}$  to the input and asking if  $\text{hd}(m, m') \leq k$ . We obtain the following dichotomy:

**Corollary 3.**  *$\text{MSD}^d(\Gamma)$  is in P if  $\Gamma \in \mathcal{L}$  is bijunctive, Horn, or dual-Horn, and it is NP-complete otherwise.*

*Proof.* This follows immediately from Theorem 2: All cases in PO become polynomial-time decidable, whereas the other cases, which are APX-hard, become NP-complete. According to Post’s lattice this classification covers all finite sets  $\Gamma$  of relations.  $\square$

## 4 Duality and Inapplicability of Clone Closure

As the optimization problem  $\text{MSD}$  is not compatible with existential quantification, we cannot prove an AP-equivalence result between any two  $\text{MSD}$  parameterized by constraint languages generating the same co-clone. Yet, similar results hold for weak co-clones.

**Proposition 4.** *We have  $\text{MSD}^d(\Gamma') \leq_m \text{MSD}^d(\Gamma)$  and  $\text{MSD}(\Gamma') \leq_{\text{AP}} \text{MSD}(\Gamma)$  for  $\Gamma, \Gamma' \in \mathcal{L}$  satisfying  $\Gamma' \subseteq \langle \Gamma \rangle_\wedge$ .*

*Proof.* For similarity it suffices to prove that  $\Gamma' \subseteq \langle \Gamma \rangle_\wedge$  implies  $\text{MSD}(\Gamma') \leq_{\text{AP}} \text{MSD}(\Gamma)$ .

Let a  $\Gamma'$ -formula  $\varphi$  be an instance of  $\text{MSD}(\Gamma')$ . Since  $\Gamma' \subseteq \langle \Gamma \rangle_\wedge$ , every constraint  $R(x_1, \dots, x_k)$  of  $\varphi$  can be written as a conjunction of constraints upon relations from  $\Gamma$ . Substitute the latter into  $\varphi$ , obtaining  $\varphi'$ . Now  $\varphi'$  is an instance of  $\text{MSD}(\Gamma)$ , where  $\varphi'$  is only polynomially larger than  $\varphi$ . As  $\varphi$  and  $\varphi'$  have the same variables and hence the same models, also the closest distinct models of  $\varphi$  and  $\varphi'$  are the same.  $\square$

For a relation  $R \subseteq \{0, 1\}^n$ , its *dual* relation is  $\text{dual}(R) = \{\bar{m} \mid m \in R\}$ , i.e., the relation containing the complements of tuples from  $R$ . We naturally extend this to sets of relations  $\Gamma$  by putting  $\text{dual}(\Gamma) = \{\text{dual}(R) \mid R \in \Gamma\}$ . Since taking complements is involutive, duality is a symmetric relation. By inspecting the bases of co-clones in Table 1, we deduce that many co-clones are duals of each other, e.g.  $\text{iE}_2$  and  $\text{iV}_2$ .

We now show that it suffices to consider only one half of Post’s lattice of co-clones.

**Lemma 5.** *For any set  $\Gamma$  of Boolean relations we have  $\text{MSD}^d(\Gamma) \equiv_m \text{MSD}^d(\text{dual}(\Gamma))$  and  $\text{MSD}(\Gamma) \equiv_{\text{AP}} \text{MSD}(\text{dual}(\Gamma))$ .*

*Proof.* For a  $\Gamma$ -formula  $\varphi$  and an assignment  $m$  to  $\varphi$  we construct a  $\text{dual}(\Gamma)$ -formula  $\varphi'$  by substitution of every atom  $R(\mathbf{x})$  by  $\text{dual}(R)(\mathbf{x})$ . Then  $m$  satisfies  $\varphi$  if and only if  $\bar{m}$  satisfies  $\varphi'$ ,  $\bar{m}$  being the complement of  $m$ . Moreover,  $\text{hd}(m, m') = \text{hd}(\bar{m}, \bar{m}')$ .  $\square$

## 5 Finding the Minimal Distance Between Solutions

### 5.1 Polynomial-Time Cases

We use the following result based on a previous theorem of Baker and Pixley [3], showing that it suffices to consider binary relations when studying bijunctive constraint languages.

**Proposition 6 (Jeavons et al. [11]).** *Any bijunctive constraint  $R(x_1, \dots, x_n)$  is equivalent to  $\bigwedge_{1 \leq i < j \leq n} R_{ij}(x_i, x_j)$ , where  $R_{ij}$  is the projection of  $R$  to the coordinates  $i$  and  $j$ .*

**Proposition 7.** *If  $\Gamma$  is bijunctive ( $\Gamma \subseteq \text{iD}_2$ ) then  $\text{MSD}(\Gamma)$  is polynomial-time solvable.*

By Proposition 6, an algorithm for bijunctive  $\Gamma$  can be restricted to at most binary clauses. We extend the algorithm of Aspvall, Plass, and Tarjan [1].

**Algorithm**

*Input:* An  $\text{iD}_2$ -formula  $\varphi$  viewed as a collection of one- or two-element sets of literals.

*Output:* “ $\leq 1$  model” or the minimal Hamming distance of any two distinct models of  $\varphi$ .

*Method:* Let  $\mathcal{V}$  be the set of variables occurring in  $\varphi$ , let  $\mathcal{L} = \{v, \neg v \mid v \in \mathcal{V}\}$  be the set of corresponding literals, and let  $\bar{u}$  denote the complementary literal to  $u \in \mathcal{L}$ .

- Construct the relation  $R := \{(\bar{u}, v), (u, \bar{v}) \mid \{u, v\} \in \varphi\} \cup \{(\bar{u}, u) \mid \{u\} \in \varphi\}$ .  
Let  $\leq$  be the reflexive and transitive closure of  $R$ , i.e. the least preorder on  $\mathcal{L}$  extending  $R$ .  
Let  $\sim := \{(u, v) \in \mathcal{L}^2 \mid u \leq v \wedge v \leq u\}$  be the associated equivalence relation.  
If  $v \sim \neg v$  holds for some variable  $v$ , then return “ $\leq 1$  model” ( $\varphi$  is unsatisfiable).
- Otherwise, let  $\mathcal{V}_0 := \{v \in \mathcal{V} \mid v \leq \neg v\}$  and  $\mathcal{V}_1 := \{v \in \mathcal{V} \mid \neg v \leq v\}$  be sets of variables being false and true, respectively, in every model of  $\varphi$ .  
If  $\mathcal{V}_0 \cup \mathcal{V}_1 = \mathcal{V}$  holds, then return “ $\leq 1$  model” ( $\varphi$  has only one model).
- Otherwise, construct the sets  
 $F_0 := \{L \in \mathcal{L}/\sim \mid \exists v \in \mathcal{V}_0 : L \leq [v]_{\sim}\} \cup \{L \in \mathcal{L}/\sim \mid \exists v \in \mathcal{V}_1 : L \leq [\neg v]_{\sim}\}$  and  
 $F_1 := \{L \in \mathcal{L}/\sim \mid \exists v \in \mathcal{V}_0 : [\neg v]_{\sim} \leq L\} \cup \{L \in \mathcal{L}/\sim \mid \exists v \in \mathcal{V}_1 : [v]_{\sim} \leq L\}$ .  
 $F_0$  ( $F_1$ ) is the set of equivalence classes of literals whose value is forced to false (to true) by backward and forward propagation from variables in  $\mathcal{V}_0$  and  $\mathcal{V}_1$ .  
Let  $P := (\mathcal{L}/\sim) \setminus (F_0 \cup F_1)$  be the set of remaining equivalence classes.  
Return  $\min\{|L| \mid L \in P\}$  as minimal Hamming distance.

**Complexity.** The size of  $\mathcal{L}$  is linear in the number of variables, the reflexive closure can be computed in linear time in  $|\mathcal{L}|$ , the transitive closure in cubic time in  $|\mathcal{L}|$ , see [16]. The equivalence relation  $\sim$  is the intersection of  $\leq$  and its inverse (quadratic in  $|\mathcal{L}|$ ); from it we can obtain the partition  $\mathcal{L}/\sim$  in linear time in  $|\mathcal{L}|$ , and combining this with the preorder  $\leq$  we can compute the order on  $\mathcal{L}/\sim$  in polynomial time, as well. Similarly, the remaining sets from the proof can be computed with polynomial time complexity.

**Correctness.** The pairs in  $R$  arise from understanding the atomic constraints in  $\varphi$  as implications. Therefore, by transitivity of implication, in every model of  $\varphi$ , literals  $u, v \in \mathcal{L}$  satisfying  $u \leq v$  have to be evaluated so that  $(m(u), m(v))$  does not violate the Boolean order relation, i.e.  $[x \rightarrow y]$ . Hence, literals  $u \sim v$  from one equivalence class have to share the same value in any model  $m$ . Moreover, since literals and their negations have to take on opposite values, we must have  $m(v) = 0$  for all  $v \in \mathcal{V}_0$  and  $m(v) = 1$  for all  $v \in \mathcal{V}_1$ . This proves that the algorithm gives a correct answer in case there do not exist feasible solutions. Furthermore, by transitivity, we see that literals in equivalence classes in  $F_i$  must be evaluated to  $i \in \{0, 1\}$ . So any two models can only differ on the literals belonging to members of  $P$ . Therefore, clearly, the return value of the algorithm is a lower bound for the minimal solution distance. To prove the converse, we shall

exhibit two models  $m_0 \neq m_1$  of  $\varphi$  having the least cardinality of equivalence classes in  $P$  as their Hamming distance.

Let  $L \in P$  be a class of minimum cardinality. Define  $m_0(u) := 0$  for all literals  $u \in L$  and likewise,  $m_1(u) := 1$ . We extend this by  $m_1(w) := m_0(w) := 0$  for all  $w \in L$  such that  $w \leq u$  for some  $u \in L$ , and by  $m_0(w) := m_1(w) := 1$  for all  $w \in L$  such that  $u \leq w$  for some  $u \in L$ . For variables  $v \in \mathcal{V}$  satisfying  $v \leq \neg v$  or  $\neg v \leq v$  we have  $[v]_{\sim} \notin P$ ; in other words, for  $[v]_{\sim} \in P$  the classes  $[v]_{\sim}$  and  $[\neg v]_{\sim}$  are incomparable. Thus, so far, we have not defined  $m_0$  and  $m_1$  on a variable  $v \in \mathcal{V}$  and on its negation  $\neg v$  at the same time. Of course, fixing a value for a negative literal  $\neg v$  to some value implicitly means that we bind the assignment for  $v \in \mathcal{V}$  to the opposite value. We complete the definition of  $m_0$  and  $m_1$  by setting them to 0 on every  $v \in \mathcal{V}$  where  $[v]_{\sim} \in P$  and they are not yet defined. Moreover, for  $v \in \mathcal{V}$  where  $[v]_{\sim} \in F_i$  we put  $m_j(v) := i$  for  $i, j \in \{0, 1\}$ . Obviously,  $m_0$  differs from  $m_1$  only in the variables corresponding to the literals in  $L$ , so their Hamming distance is  $|L|$  as desired. Besides, both assignments respect the order constraints in  $(\mathcal{L}, \leq)$ . As these faithfully reflect all original atomic constraints,  $m_0$  and  $m_1$  are indeed models of  $\varphi$ .

**Proposition 8.** *MSD( $\Gamma$ ) is in PO for  $\Gamma \in \mathcal{L}$  satisfying  $\Gamma \subseteq \text{iE}_2$  or  $\Gamma \subseteq \text{iV}_2$ .*

We only discuss the Horn case ( $\Gamma \subseteq \text{iE}_2$ ), dual-Horn ( $\Gamma \subseteq \text{iV}_2$ ) being symmetric.

**Algorithm**

*Input:* A Horn formula  $\varphi$  viewed as a set of Horn clauses.

*Output:* “ $\leq 1$  model” or the minimal Hamming distance of any two distinct models of  $\varphi$ .

*Method:*

Step 1: For each variable  $x$  in  $\varphi$ , add the clause  $(\neg x \vee x)$ . Apply the following rules to  $\varphi$  until no more clauses and literals can be removed and no new clauses can be added.

- Unit resolution and unit subsumption: Let  $\bar{u}$  denote the complement of a literal  $u$ . If the clause set contains a unit clause  $u$ , remove all clauses containing the literal  $u$  and remove all literals  $\bar{u}$  from the remaining clauses.
- Hyper-resolution with binary implications: Resolve all negative literals of a clause simultaneously with binary implications possessing identical premises.

$$\frac{(\neg x \vee y_1) \cdots (\neg x \vee y_k) \quad (\neg y_1 \vee \cdots \vee \neg y_k \vee z)}{(\neg x \vee z)} \quad \frac{(\neg x \vee y_1) \cdots (\neg x \vee y_k) \quad (\neg y_1 \vee \cdots \vee \neg y_k)}{(\neg x)}$$

Let  $\mathcal{D}$  be the final result of this step. If  $\mathcal{D}$  is empty or contains the empty clause, return “ $\leq 1$  model”.

Step 2: Let  $\mathcal{V}$  be the set of variables occurring in  $\mathcal{D}$ , and let  $\sim \subseteq \mathcal{V}^2$  be the relation defined by  $x \sim y$  if  $\{\neg x \vee y, \neg y \vee x\} \subseteq \mathcal{D}$ . Note that  $\sim$  is an equivalence, since the tautological clauses ensure reflexivity and resolution of implications computes their transitive closure. We say that a variable  $z$  depends on variables



$y_1, \dots, y_k$ , if  $\mathcal{D}$  contains the clauses  $\neg y_1 \vee \dots \vee \neg y_k \vee z$ ,  $\neg z \vee y_1, \dots, \neg z \vee y_k$  and  $z \not\sim y_i$  holds for all  $i = 1, \dots, k$ .

Return  $\min\{|X| \mid X \in \mathcal{V}/\sim, X \text{ does not contain dependent variables}\}$  as minimal Hamming distance.

**Complexity.** The run-time of the algorithm is polynomial in the number of clauses and the number of variables in  $\varphi$ : Unit resolution/subsumption can be applied at most once for each variable, and hyper-resolution has to be applied at most once for each variable  $x$  and each clause  $\neg y_1 \vee \dots \vee \neg y_k \vee z$  and  $\neg y_1 \vee \dots \vee \neg y_k$ .

**Correctness.** Let  $\mathcal{U}$  be the set of unit clauses removed by subsumption. Adding resolvents and removing subsumed clauses maintains logical equivalence, therefore  $\mathcal{D} \cup \mathcal{U}$  is logically equivalent to  $\varphi$ , i.e., both clause sets have the same models. If  $\mathcal{D}$  is empty, the unit clauses in  $\mathcal{U}$  define a unique model of  $\varphi$ . If  $\mathcal{D}$  contains the empty clause, the sets  $\mathcal{D}$  and  $\varphi$  are unsatisfiable. Otherwise  $\mathcal{D}$  has at least two models, as we will show below. As each model  $m$  of  $\mathcal{D}$  uniquely extends to a model of  $\varphi$  by defining  $m(x) = 1$  for  $(x) \in \mathcal{U}$  and  $m(x) = 0$  for  $(\neg x) \in \mathcal{U}$ , the minimal Hamming distances of  $\varphi$  and  $\mathcal{D}$  are the same.

We are thus looking for models  $m_1, m_2$  of  $\mathcal{D}$  such that the size of the difference set  $\Delta(m_1, m_2) = \{x \mid m_1(x) \neq m_2(x)\}$  is minimal. In fact, since the models of Horn formulas are closed under minimum, we may assume  $m_1 < m_2$ , i.e., we have  $m_1(x) = 0$  and  $m_2(x) = 1$  for all variables  $x \in \Delta(m_1, m_2)$ . Indeed, given two models  $m_2$  and  $m'_2$  of  $\mathcal{D}$ ,  $m_1 = m_2 \wedge m'_2$  is also a model. Since  $\text{hd}(m_1, m_2) \leq \text{hd}(m_2, m'_2)$  holds, the minimal Hamming distance will occur between models  $m_1$  and  $m_2$  satisfying  $m_1 < m_2$ .

Note the following facts regarding the equivalence relation  $\sim$  and dependent variables.

- If  $x \sim y$  then the two variables must have the same value in every model of  $\mathcal{D}$  in order to satisfy the implications  $\neg x \vee y$  and  $\neg y \vee x$ . This means that for all models  $m$  of  $\mathcal{D}$  and all  $X \in \mathcal{V}/\sim$ , we have either  $m(x) = 0$  for all  $x \in X$  or  $m(x) = 1$  for all  $x \in X$ .

- The dependence of variables is acyclic: If  $z_i$  depends on  $z_{i+1}$  for  $i = 1, \dots, l$  and  $z_l$  depends on  $z_1$ , then we have a cycle of binary implications between the variables and thus  $z_i \sim z_j$  for all  $i, j$ , contradicting the definition of dependence.

- If a variable  $z$  depending on  $y_1, \dots, y_k$  belongs to a difference set  $\Delta(m_1, m_2)$ , then at least one of the  $y_i$ s also has to belong to  $\Delta(m_1, m_2)$ :  $m_2(z) = 1$  implies  $m_2(y_j) = 1$  for all  $j = 1, \dots, k$  (because of the clauses  $\neg z \vee y_i$ ), and  $m_1(z) = 0$  implies  $m_1(y_i) = 0$  for at least one  $i$  (because of the clause  $\neg y_1 \vee \dots \vee \neg y_k \vee z$ ). Therefore  $\Delta(m_1, m_2)$  is the union of at least two sets in  $\mathcal{V}/\sim$ , namely the equivalence class of  $z$  and the one of  $y_i$ .

Hence the difference between any two models cannot be smaller than the cardinality of the smallest set in  $\mathcal{V}/\sim$  without dependent variables. It remains to show that we can indeed find two such models.

Let  $X$  be a set in  $\mathcal{V}/\sim$  which has minimal cardinality among the sets without dependent variables, and let  $m_1, m_2$  be interpretations defined as follows: (1)  $m_1(y) = 0$  and  $m_2(y) = 1$  if  $y \in X$ ; (2)  $m_1(y) = 1$  and  $m_2(y) = 1$  if  $y \notin X$  and

$(\neg x \vee y) \in \mathcal{D}$  for some  $x \in X$ ; (3)  $m_1(y) = 0$  and  $m_2(y) = 0$  otherwise. We have to show that  $m_1$  and  $m_2$  satisfy all clauses in  $\mathcal{D}$ . Let  $m$  be any of these models.  $\mathcal{D}$  contains two types of clauses.

Type 1: Horn clauses with a positive literal  $\neg y_1 \vee \dots \vee \neg y_k \vee z$ . If  $m(y_i) = 0$  for any  $i$ , we are done. So suppose  $m(y_i) = 1$  for all  $i = 1, \dots, k$ ; we have to show  $m(z) = 1$ . The condition  $m(y_i) = 1$  means that either  $y_i \in X$  (for  $m = m_2$ ) or that there is a clause  $(\neg x_i \vee y_i) \in \mathcal{D}$  for some  $x_i \in X$ . We distinguish the two cases  $z \in X$  and  $z \notin X$ .

Let  $z \in X$ . If  $z \sim y_i$  for any  $i$ , we are done for we have  $m(z) = m(y_i) = 1$ . So suppose  $z \not\sim y_i$  for all  $i$ . As the elements in  $X$ , in particular  $z$  and the  $x_i$ s, are equivalent and the binary clauses are closed under resolution,  $\mathcal{D}$  contains the clause  $\neg z \vee y_i$  for all  $i$ . But this would mean that  $z$  is a variable depending on the  $y_i$ s, contradicting the assumption  $z \in X$ . Let  $z \notin X$ , and let  $x \in X$ . As the elements in  $X$  are equivalent and the binary clauses are closed under resolution,  $\mathcal{D}$  contains  $\neg x \vee y_i$  for all  $i$ . Closure under hyper-resolution with the clause  $\neg y_1 \vee \dots \vee \neg y_k \vee z$  means that  $\mathcal{D}$  also contains  $\neg x \vee z$ , whence  $m(z) = 1$ .

Type 2: Horn clauses with only negative literals  $\neg y_1 \vee \dots \vee \neg y_k$ . If  $m(y_i) = 0$  for any  $i$ , we are done. It remains to show that the assumption  $m(y_i) = 1$  for all  $i = 1, \dots, k$  leads to a contradiction. The condition  $m(y_i) = 1$  means that either  $y_i \in X$  (for  $m = m_2$ ) or that there is a clause  $(\neg x_i \vee y_i) \in \mathcal{D}$  for some  $x_i \in X$ . Let  $x$  be some particular element of  $X$ . Since the elements in  $X$  are equivalent and the binary clauses are closed under resolution,  $\mathcal{D}$  contains the clause  $\neg x \vee y_i$  for all  $i$ . But then a hyper-resolution step with the clause  $\neg y_1 \vee \dots \vee \neg y_k$  would yield the unit clause  $\neg x$ , which by construction does not occur in  $\mathcal{D}$ . Therefore at least one  $y_i$  is neither in  $X$  nor part of a clause  $\neg x \vee y_i$  with  $x \in X$ , i.e.,  $m(y_i) = 0$ .

## 5.2 Hard Cases

**Two Solution Satisfiability.** In this section we study the feasibility problem of  $\text{MSD}(\Gamma)$  which is, given a  $\Gamma$ -formula  $\varphi$ , to decide if  $\varphi$  has two distinct solutions.

**Problem:**  $2\text{SolutionSAT}(\Gamma)$

*Input:* Conjunctive formula  $\varphi$  over the relations from  $\Gamma$ .

*Question:* Are there two satisfying assignments  $m \neq m'$  of  $\varphi$ ?

A priori it is not clear that the tractability of  $2\text{SolutionSAT}$  is fully characterized by co-clones. The problem is that the implementation of relations of some language  $\Gamma$  by another language  $\Gamma'$  might not be parsimonious, that is, in the implementation one solution to a constraint might be blown up into several ones in the implementation. Fortunately we can still determine the tractability frontier for  $2\text{SolutionSAT}$  by combining the corresponding results for  $\text{SAT}$  and  $\text{AnotherSAT}$ .

**Lemma 9.** For  $\Gamma \in \mathcal{L}$  where  $\text{SAT}(\Gamma)$  is NP-hard,  $2\text{SolutionSAT}(\Gamma)$  is NP-hard.

*Proof.* Since  $\text{SAT}(\Gamma)$  is NP-hard, there must be a relation  $R$  in  $\Gamma$  having more than one tuple, because every relation containing only one tuple is at the same time Horn, dual Horn, bijunctive, and affine. Given an instance  $\varphi$  for  $\text{SAT}(\Gamma)$ ,

construct  $\varphi'$  as  $\varphi \wedge R(y_1, \dots, y_\ell)$  where  $\ell$  is the arity of  $R$  and  $y_1, \dots, y_\ell$  are new variables not appearing in  $\varphi$ . Obviously,  $\varphi$  has a solution if and only if  $\varphi'$  has at least two solutions. Hence, we have proved  $\text{SAT}(\Gamma) \leq_m \text{2SolutionSAT}(\Gamma)$ .  $\square$

**Lemma 10.** *If  $\Gamma \in \mathfrak{L}$  and  $\text{AnotherSAT}(\Gamma)$  is NP-hard,  $\text{2SolutionSAT}(\Gamma)$  is NP-hard.*

*Proof.* Let a formula  $\varphi$  and a satisfying assignment  $m$  be an instance of  $\text{AnotherSAT}(\Gamma)$ . Then  $\varphi$  has a solution other than  $m$  if and only if it has two distinct solutions.  $\square$

**Lemma 11.** *If  $\text{SAT}(\Gamma)$  and  $\text{AnotherSAT}(\Gamma)$  are in P for  $\Gamma \in \mathfrak{L}$ , then the same holds for  $\text{2SolutionSAT}(\Gamma)$ .*

*Proof.* Let  $\varphi$  be an instance of  $\text{2SolutionSAT}(\Gamma)$ . All polynomial-time decidable cases of  $\text{SAT}(\Gamma)$  are constructive, i.e., whenever that problem is polynomial-time decidable, there exists a polynomial-time algorithm computing a satisfying assignment. Thus we can compute in polynomial time a satisfying assignment  $m$  of  $\varphi$ . Now use the algorithm for  $\text{AnotherSAT}(\Gamma)$  on the instance  $(\varphi, m)$  to decide if there is a second solution to  $\varphi$ .  $\square$

**Corollary 12.** *For  $\Gamma \in \mathfrak{L}$ , the problem  $\text{2SolutionSAT}(\Gamma)$  is polynomial-time decidable if both  $\text{SAT}(\Gamma)$  and  $\text{AnotherSAT}(\Gamma)$  are polynomial-time decidable. Otherwise,  $\text{2SolutionSAT}(\Gamma)$  is NP-hard.*

**Proposition 13.** *For  $\Gamma \in \mathfrak{L}$  such that  $\text{2SolutionSAT}(\Gamma)$  is in P, there is a polynomial-time  $n$ -approximation algorithm for  $\text{MSD}(\Gamma)$ , where  $n$  is the number of variables.*

*Proof.* Since  $\text{2SolutionSAT}(\Gamma)$  is in P, both  $\text{SAT}(\Gamma)$  and  $\text{AnotherSAT}(\Gamma)$  must be in P by Corollary 12. Since  $\text{SAT}(\Gamma)$  is in P, we can compute a model  $m$  of the input  $\varphi$  in polynomial time if it exists. Now we check the  $\text{AnotherSAT}(\Gamma)$ -instance  $(\varphi, m)$ . If it has a solution  $m' \neq m$ , it is also polynomial time computable, and we return  $(m, m')$ . If we fail somewhere in this process, then  $\text{MSD}(\Gamma)$  does not have feasible solutions; otherwise,  $\text{hd}(m, m') \leq n \leq n \cdot \text{OPT}(\varphi)$ .  $\square$

**MinDistance-Equivalent Cases.** In this section we show that, as for the Nearest Other Solution problem (see [4, 5]), the affine cases of MSD are MinDistance-complete.

**Proposition 14.**  *$\text{MSD}(\Gamma)$  is MinDistance-complete if  $\Gamma \in \mathfrak{L}$  satisfies  $i\mathbb{L} \subseteq \langle \Gamma \rangle \subseteq i\mathbb{L}_2$ .*

*Proof.* We prove  $\text{MSD}(\Gamma) \equiv_{\text{AP}} \text{NearestOtherSolution}(\Gamma)$ , which is MinDistance-complete by [4]. As  $\Gamma \subseteq i\mathbb{L}_2 = \langle \{\text{even}^4, [x], [-x]\} \rangle$ , any  $\Gamma$ -formula  $\psi$  is expressible as  $\exists y(A_1x + A_2y = c)$ . The projection of the affine solution space is again an affine space, so it can be understood as solutions of a system  $Ax = b$ . If  $(\psi, m_0)$  is an instance of  $\text{NSol}(\Gamma)$ , then  $\psi$  is an  $\text{MSD}(\Gamma)$ -instance, and a feasible solution  $m_1 \neq m_2$  satisfying  $\psi$  gives a feasible solution  $m_3 := m_0 + (m_2 - m_1)$  for  $(\psi, m_0)$ ,

where  $\text{hd}(m_0, m_3) = \text{hd}(m_2, m_1)$ . Conversely, a solution  $m_3 \neq m_0$  to  $(\psi, m_0)$  yields a feasible answer to the MSD-instance  $\psi$ . Thus,  $\text{OPT}(\psi) = \text{OPT}(\psi, m_0)$  and so  $\text{NSol}(\Gamma) \leq_{\text{AP}} \text{MSD}(\Gamma)$ . The other way round, if  $\psi$  is an MSD-instance, then solve the system  $Ax = b$  defined by it; let  $m_0$  be a model of  $\psi$ . As above we conclude  $\text{OPT}(\psi) = \text{OPT}(\psi, m_0)$ , and therefore,  $\text{MSD}(\Gamma) \leq_{\text{AP}} \text{NSol}(\Gamma)$ .  $\square$

**Tightness Results.** It will be convenient to consider the following decision problem, already studied in [5].

**Problem: AnotherSAT<sub><n</sub>(Γ)**

*Input:* A conjunctive formula  $\varphi$  over relations from  $\Gamma$  and an assignment  $m$  satisfying  $\varphi$ .

*Question:* Is there another satisfying assignment  $m'$  of  $\varphi$ , different from  $m$ , such that  $\text{hd}(m, m') < n$ , where  $n$  is the number of variables of  $\varphi$ ?

Note that **AnotherSAT<sub><n</sub>(Γ)** is not compatible with existential quantification. Let  $\varphi(y, x_1, \dots, x_n)$  with the satisfying assignment  $m$  be an instance of **AnotherSAT<sub><n</sub>(Γ)** and  $m'$  its solution satisfying  $\text{hd}(m, m') < n + 1$ . Let  $m_1$  and  $m'_1$  be the corresponding vectors to  $m$  and  $m'$ , respectively, with the first coordinate truncated. When we existentially quantify the variable  $y$  in  $\varphi$ , producing  $\varphi_1(x_1, \dots, x_n) = \exists y \varphi(y, x_1, \dots, x_n)$ , then both  $m_1$  and  $m'_1$  are solutions of  $\varphi'$ , but we cannot guarantee that  $\text{hd}(m_1, m'_1) < n$ . Hence we need the equivalent of Proposition 4 for this problem, whose proof is analogous.

**Proposition 15 (Behrisch et al. [4, 5]).** *For  $\Gamma, \Gamma' \in \mathcal{L}$  with  $\Gamma' \subseteq \langle \Gamma \rangle_\wedge$  we have the reduction  $\text{AnotherSAT}_{<n}(\Gamma') \leq_{\text{m}} \text{AnotherSAT}_{<n}(\Gamma)$ .*

The following proposition presents only a partial result for **AnotherSAT<sub><n</sub>(Γ)** already proved in [5]. An exhaustive complexity classification of the problem **AnotherSAT<sub><n</sub>(Γ)** has been performed in [4].

**Proposition 16 (Behrisch et al. [4, 5]).** *AnotherSAT<sub><n</sub>(Γ) is NP-complete for  $\Gamma \in \mathcal{L}$  such that  $\text{iN} \subseteq \langle \Gamma \rangle \subseteq \text{iI}$ .*

*Remark 17.* It is easy to see that **AnotherSAT<sub><n</sub>(Γ)** is NP-complete for  $\text{iI}_0 \subseteq \langle \Gamma \rangle$  and  $\text{iI}_1 \subseteq \langle \Gamma \rangle$ , since already **AnotherSAT(Γ)** is NP-complete for these cases, as it was proved in [12]. It is also clear that **AnotherSAT<sub><n</sub>(Γ)** is polynomial-time decidable if  $\Gamma$  is Horn ( $\Gamma \subseteq \text{iE}_2$ ), dual Horn ( $\Gamma \subseteq \text{iV}_2$ ), bijunctive ( $\Gamma \subseteq \text{iD}_2$ ), or affine ( $\Gamma \subseteq \text{iL}_2$ ), just for the same reason as for **AnotherSAT(Γ)**. In all these four Schaefer cases, for each variable  $x_i$  we flip the value of  $m[i]$ , substitute  $\bar{m}(x_i)$  for  $x_i$ , and construct another satisfying assignment if it exists. Consider now the solutions which we get for every variable  $x_i$ . Either there is no solution for any variable, then **AnotherSAT<sub><n</sub>(Γ)** has no solution; or there are only the solutions which are the complement of  $m$ , then **AnotherSAT<sub><n</sub>(Γ)** has no solution as well; or else we get a solution  $m'$  with  $\text{hd}(m, m') < n$ , then **AnotherSAT<sub><n</sub>(Γ)** also has a solution. Hence, there is an easy to prove dichotomy result also for **AnotherSAT<sub><n</sub>(Γ)**.

We prove that Proposition 13 is essentially tight.

**Proposition 18.** *For  $\Gamma \in \mathfrak{L}$  such that  $iN \subseteq \langle \Gamma \rangle \subseteq iI$  and any  $\varepsilon > 0$  there is no polynomial-time  $n^{1-\varepsilon}$ -approximation algorithm for  $\text{MSD}(\Gamma)$ , unless  $P = NP$ .*

*Proof.* We show that any polynomial time  $n^{1-\varepsilon}$ -approximation algorithm for  $\text{MSD}(\Gamma)$  would also allow to decide  $\text{AnotherSAT}_{<n}(\Gamma)$ , being NP-complete by Proposition 16, in polynomial time.

The algorithm works as follows. Given an instance  $(\varphi, m)$  for  $\text{AnotherSAT}_{<n}(\Gamma)$ , the algorithm accepts if  $m$  is not a constant assignment. Since  $\Gamma$  is 0-valid (and 1-valid), this output is correct. If  $\varphi$  has only one variable, reject because  $\varphi$  has only two models; otherwise, proceed as follows.

For each variable  $x$  of  $\varphi$ , we construct a new formula  $\varphi'_x$  as follows. Let  $k$  be the smallest integer greater than  $1/\varepsilon$ . Introduce  $n^k - n$  new variables  $x^i$  for  $i = 1, \dots, n^k - n$ . For every  $i \in \{1, \dots, n^k - n\}$  and every constraint  $R(y_1, \dots, y_\ell)$  in  $\varphi$ , such that  $x \in \{y_1, \dots, y_\ell\}$ , construct a new constraint  $R(z_1^i, \dots, z_\ell^i)$  by  $z_j^i = x^i$  if  $y_j = x$  and  $z_j^i = y_j$  otherwise; add all the newly constructed constraints to  $\varphi$  in order to get  $\varphi'_x$ . Note, that we can extend models  $s$  of  $\varphi$  to models  $s'$  of  $\varphi'_x$  by setting  $s'(x^i) = s(x)$ . Now run the  $n^{1-\varepsilon}$ -approximation algorithm for  $\text{MSD}(\Gamma)$  on  $\varphi'_x$ . If for every  $x$  the answer is a pair  $(m_1, m_2)$  with  $m_2 = \overline{m_1}$ , then reject, otherwise accept.

This procedure is a correct polynomial-time algorithm for  $\text{AnotherSAT}_{<n}(\Gamma)$ . For polynomial runtime is clear, it remains to show correctness. If  $\varphi$  has only constant models, then the same is true for every  $\varphi'_x$ . Thus each approximation must result in a pair of complementary constant assignments, and the output is correct. Assume now that there is a model  $s$  of  $\varphi$  different from  $\mathbf{0}$  and  $\mathbf{1}$ . Hence, there exists a variable  $x$  such that  $s(x) = m(x)$ . It follows that  $\varphi'_x$  has a model  $s'$  for which  $\text{hd}(s', m') < n$  holds, where  $n$  is the number of variables of  $\varphi$ . But then the approximation algorithm must find two distinct models  $m_1 \neq m_2$  of  $\varphi'_x$  satisfying  $\text{hd}(m_1, m_2) < n \cdot (n^k)^{1-\varepsilon} = n^{k(1-\varepsilon)+1}$ . Since the inequality  $k > 1/\varepsilon$  holds, it follows that  $\text{hd}(m_1, m_2) < n^k$ . Consequently, we have  $m_2 \neq \overline{m_1}$  and the output of our algorithm is again correct.  $\square$

## 6 Concluding Remarks

Our problem is in PO for constraints, which are bijunctive, or Horn, or dual Horn. The next complexity stage of the solution structure is characterized by affine constraints. In fact, these constraints represent the error correcting codes used in real-word applications. If we search for arbitrary two satisfying assignments with minimum distance, we can apply standard linear algebra techniques and perform an affine transformation, where we can enforce one of the assignments to be the zero-vector. This is not surprising, since in linear algebra many problems in an affine space can be transformed to the same problems in the corresponding vector space. The penultimate stage of solution structure complexity is represented by constraints, for which the existence of a solution is guaranteed by their definition, but we do not have any other exploitable information. For MSD we need a guarantee of at least two solutions. Our problem belongs to the

class poly-APX for these constraints. We can exactly pinpoint the polynomial ( $n$ , i.e. arity of the formula) for which we can get a polynomial-time approximation. Our complexity results indicate moreover that we cannot get a suitable approximation for these types of the considered optimization problem. All other cases are not polynomial-time approximable at all. It is interesting to see that our results differ considerably from those of [9] for `MaxHammingDistance`, asking to produce two satisfying assignments having maximal Hamming distance, even if the two problems are dual.

## References

1. Aspvall, B., Plass, M.R., Tarjan, R.E.: A linear-time algorithm for testing the truth of certain quantified Boolean formulas. *Inf. Process. Lett.* **8**(3), 121–123 (1979)
2. Ausiello, G., Crescenzi, P., Gambosi, G., Kann, V., Marchetti-Spaccamela, A., Prota, M.: *Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties*. Springer, New York (1999)
3. Baker, K.A., Pixley, A.F.: Polynomial interpolation and the Chinese Remainder Theorem for algebraic systems. *Mathematische Zeitschrift* **143**(2), 165–174 (1975)
4. Behrisch, M., Hermann, M., Mengel, S., Salzer, G.: Minimal distance of propositional models (2015). [abs/1502.06761](https://arxiv.org/abs/1502.06761)
5. Behrisch, M., Hermann, M., Mengel, S., Salzer, G.: Give me another one!. In: Elbassioni, K., Makino, K. (eds.) *ISAAC 2015*. LNCS, vol. 9472, pp. 664–676. Springer, Heidelberg (2015). doi:[10.1007/978-3-662-48971-0\\_56](https://doi.org/10.1007/978-3-662-48971-0_56)
6. Böhler, E., Creignou, N., Reith, S., Vollmer, H.: Playing with Boolean blocks, part II: constraint satisfaction problems. *SIGACT News* **35**(1), 22–35 (2004)
7. Böhler, E., Reith, S., Schnoor, H., Vollmer, H.: Bases for Boolean co-clones. *Inf. Process. Lett.* **96**(2), 59–66 (2005)
8. Creignou, N., Khanna, S., Sudan, M.: *Complexity Classifications of Boolean Constraint Satisfaction Problems*. SIAM Monographs on Discrete Mathematics and Applications, vol. 7. SIAM, Philadelphia (2001)
9. Crescenzi, P., Rossi, G.: On the Hamming distance of constraint satisfaction problems. *Theor. Comput. Sci.* **288**(1), 85–100 (2002)
10. Dumer, I., Micciancio, D., Sudan, M.: Hardness of approximating the minimum distance of a linear code. *IEEE Trans. Inf. Theory* **49**(1), 22–37 (2003)
11. Jeavons, P., Cohen, D., Gyssens, M.: Closure properties of constraints. *J. Assoc. Comput. Mach.* **44**(4), 527–548 (1997)
12. Juban, L.: Dichotomy theorem for the generalized unique satisfiability problem. In: Ciobanu, G., Păun, G. (eds.) *FCT 1999*. LNCS, vol. 1684, pp. 327–337. Springer, Heidelberg (1999)
13. Lagerkvist, V.: Weak bases of Boolean co-clones. *Inf. Process. Lett.* **114**(9), 462–468 (2014)
14. Schaefer, T.J.: The complexity of satisfiability problems. In: *Proceedings of the Tenth Annual ACM Symposium on Theory of Computing, STOC 1978*, San Diego, California, pp. 216–226. ACM, New York (1978). <http://dx.doi.org/10.1145/800133.804350>
15. Schnoor, H., Schnoor, I.: Partial polymorphisms and constraint satisfaction problems. In: Creignou, N., Kolaitis, P.G., Vollmer, H. (eds.) *Complexity of Constraints*. LNCS, vol. 5250, pp. 229–254. Springer, Heidelberg (2008)
16. Warshall, S.: A theorem on Boolean matrices. *J. Assoc. Comput. Mach.* **9**(1), 11–12 (1962)