On the Complexity of MM_1SNP

Manuel Bodirsky* Hubie Chen† Tomás Feder‡ two other authors

September 11, 2009

Abstract

Monotone monadic strict NP ($MMSNP$) is a class of computational problems that is closely related to the class of constraint satisfaction problems for constraint languages over finite domains. It is known that the dichotomy problem for both classes are equivalent. Whereas the dichotomy conjecture has been verified for several subclasses of constraint satisfaction problems, little is known about the the computational complexity for subclasses of $MMSNP$.

In this project we present complexity classifications for several subclasses of $MMSNP$. In particular we study MM_1SNP, which is the subclass of $MMSNP$ where only a single monadic predicate is allowed, and we show several partial results.

Problems in $MMSNP$ can also be seen as structure partition problems with forbidden homomorphic obstructions, and they have natural combinatorial descriptions. We completely classify the complexity of $MMSNP$ for the case where the obstructions are monochromatic and where loops in the input are forbidden.

1 Introduction

The class $MMSNP$ can be described in terms of forbidden patterns. Let V be a finite set of values. Let S be a structure signature. Let X be a set of structures over the signature S whose elements are labeled by elements of the set V of finite values. The elements of X are called *forbidden patterns*. An instance of $MMSNP(X)$ is a structure R over the signature S. The instance R is accepted if the elements of R can be labeled by values of V in such a way that no labeled structure $P \in X$ has a label preserving homomorphism to the labeled structure R.

A subclass of the class $MMSNP$ of problems $MMSNP(X)$ is the class MM_1SNP of problems $MM_1SNP(X)$ for which the set of values is $V = \{0, 1\}$. An instance R is *loopless* if no relation R_i in R contains a tuple with all its arguments the same. A set of forbidden patterns X is *monochromatic* if each pattern $P \in X$ has all its elements labeled by the same $a_P \in V$.

*Ecole Polytechnique. Email: bodirsky@informatik.hu-berlin.de.

†Universitat Pompeu Fabra. Email: hubiechen@gmail.com.

‡268 Waverley St., Palo Alto, CA 94301. E-mail: tomas@theory.stanford.edu.

A *clique of size* $k \geq 1$ *over the signature* S is a structure Q over S with k elements where each relation R_i of arity r_i contains all r_i-tuples of the k elements other than those for which the r_i arguments are all the same element. Given a loopless structure R over S, the *chromatic number* $c(R)$ is the least $k \geq 1$ such that there is a homomorphism from R to the clique of size k. If X is a monochromatic set of forbidden patterns and $a \in V$, then $c(X, a)$ is the least $k \geq 1$ such that some $P \in X$ with labels a has $c(P) = k$.

The *core* of a structure R is the smallest substructure R' of R such that there is a homomorphism form R to R'. A structure R is a *core* if the core of R is R.

2 Classification of $MM_1SNP(X)$ over Loopless Input Structures R and Monochromatic Forbidden Patterns

Theorem 1 *Let X be a monochromatic set of forbidden patterns over signature S with values $V = \{0, 1\}$. Then $MM_1SNP(X)$ is either polynomial time solvable or NP-complete for loopless input structures R.*

PROOF: We may assume without loss of generality that the structures in X are cores, and no structure of X maps homomorphically to a proper substructure of another structure of X with the same label.

If for some value $a \in V$ there is no forbidden pattern in X with label a, then we may label all elements of R by a, so in this case the problem is polynomial time solvable.

If for some value $a \in V$ there is a forbidden pattern $P \in X$ with $c(P) = 1$, then no element of R may be assigned label a, so it suffices to test the assignment of label $b \in V$ with $b \neq a$ to all elements of R to verify whether a solution exists, so in this case the problem is polynomial time solvable.

In the remaining cases we thus have $c(X, 0) \geq 2, c(X, 1) \geq 2$. If every structure in X has 2 elements, then the occurrences of the structures $P \in X$ can be viewed as 2-satisfiability clauses over $V = \{0, 1\}$ forbidding that two elements be both assigned values 0 or both assigned values 1. Since 2-satisfiability can be solved in polynomial time, in this case the problem is polynomial time solvable.

Suppose $c(X, 0) = c(X, 1) = 2$ and X has a structure P with $k \geq 3$ elements and $c(P) = 2$, say with label 0. Choose P with the smallest connected component P'. If P is not connected, we shall include in every instance a connected component consisting of two copies of P' joined by a complete bipartite graph with edges that are cliques of size two. The two copies shall be assigned labels that are 0 and 1 respectively, thus the pattern P' will apear both with labels 0 and 1, thus reducing the problem to a smaller X, by induction. We may thus assume P is connected. Then assigning to all k elements of P value 0 if P occurs in an instance R is not allowed, but assigning to $k-1$ elements of P value 0 and to the remaining element of P value 1. We infer that P can be used to define a Boolean constraint that gives rise to an NP-complete problem from Schaefer's classification. P cannot be majority since the majority of $100\cdots, 010\cdots, 001\cdots$ is $000\cdots$ which is forbidden. P cannot be Horn since the corresponding commutative conservative function applied to $100\cdots, 010\cdots$ gives

$000\cdots$ which is forbidden. If we attach to each of the k elements of P a clique of size two, the endpoints of each of these k cliques must be assigned different values over V, since $c(X,0)=c(X,1)=2$. P cannot thus be dual Horn since the corresponding commutative conservative function applied to $011\cdots, 101\cdots$ gives $111\cdots$ which is forbidden, at the other end of the cliques of size 2. Finally P cannot be linear because if we can assign to two elements of P the value 1 and to the remaining ones the value 0, then the linear function applied to $100\cdots, 010\cdots, 110\cdots$ gives $000\cdots$ which is forbidden. So at least two ones in P is forbidden, yet the linear function applied to $100\cdots, 010\cdots, 001\cdots$ gives $111\cdots$ which is forbidden. The occurrences of P or of P extende by cliques of size 2 in an instance R may be joined together by paths of length two forcing complementation of the value in V twice, thus representing an instance of an NP-complete problem in Schafer's classification.

In the remaining case $3 \leq c(X,0) \geq c(X,1) \geq 2$ or $3 \leq c(X,1) \geq c(X,0) \geq 2$, say $3 \leq c(X,0) \geq c(X,1) \geq 2$. Let $k=c(X,0)$. Then an occurrence of a clique Q of size k in an instance R may not be labeled all 0 of all 1, and we must use some number t of 1s with $k-t$ 0s, in any order, with $1 \leq t < c(X,1)$. This clique corresponds to a relation in Schaefer'c classification such as 1-in-3 SAT or not-all-equal SAT, which are NP-complete. The NP-completeness carries over for instances of girth greater than any chosen constant d, as shown by Feder and Vardi with randomized reductions derandomized by Kun. We may choose d greater than the maximum size of any $P \in X$. Then any occurrence of such a P would not involve a cycle of copies of Q, and thus the chromatic number involving 0 would be at most $c(X,0)-1$ and the chromatic number involving 1 would be at most $c(X,1)-1$, so no $P \in X$ will occur in a solution to the NP-complete problem. So the problem in this last case is also NP-complete. ■

Corollary 1 *The assumption that instances are loopless in Theorem 1 may be dropped if the signature S contains a single relation.*

PROOF: Any forbidden $P \in X$ will forbid the corresponding value from being assigned to a loop at an element. ■

3 From $MM_1SNP(X)$ to $MMSNP(X)$

Theorem 2 *Let X be a monochromatic set of forbidden patterns over signature S with values V. Then $MMSNP(X)$ is either polynomial time solvable or NP-complete for loopless input structures R.*

PROOF: If $|V|=1$ then the problem is polynomial time solvable. The case $|V|=2$ was classified in Theorem 1.

In the remaining case $|V| \geq 3$. If some label a does not occur in any pattern in X then we may label all elements of R by a, so in this case the problem is polynomial time solvable. If for some label a there is a pattern $P \in X$ with $c(P)=1$, then no element of R may be assigned label a, so we may replace V by $V \setminus \{a\}$ and proceed by induction.

Finally if for all $a \in V$ we have $c(X,a) \geq a$, then a clique of size $k=\sum_{a\in V}(c(X,a)-1)$ must have $c(X,a)-1$ elements with label a. This relation on k elements gives an NP-complete CSP problem, which may be reduced to piecing together cliques of size k, with high girth as in the proof of Theorem 1 by the results of Feder and Vardi, and Kun enforcing high girth. ■

Corollary 2 *The assumption that instances are loopless in Theorem 2 ay be dropped if the signature S contains a single relation.*

PROOF: Any forbidden $P \in X$ will forbid the corresponding value from being assigned to a loop at an element. ■

4 A Mixed NP-complete Case for $MM_1SNP(X)$

We note that every polynomial problem in Schaefer's classification can be encoded as a polynomial problem on digraphs by representing relations of arity $k_i \geq 3$, one for each k_i chosen, with a directed cycle of length k_i with one edge reversed and the corresponding assignments of values, with high girth enforced by paths of the equality relation.

The rest of this section is devoted to the following problem, the *path problem.* Let $V = \{0, 1\}$, let S be the signature of digraphs (a single binary relation). Let k be an integer. Let B be a set of integers $0 \leq i \leq k$. Let X consist of paths on k vertices directed forward with any combination having i 1s and $k - i$ 0s on the path is forbidden, for $i \notin B$. The problem $MM_1SNP(X)$ is the path problem.

If $0 \in B$ or $k \in B$, then the problem is trivially polynomial, as we can assign all 0s or all 1s respectively. We thus assume $0 \notin B$ and $1 \notin B$. For $0 \leq a < b \leq k$, we say that a, b is a *skip* if $a \notin B$, $b \notin B$, and for all $a < i < b$, $i \in B$. The *gap* of the skip is $b - a - 1$.

Theorem 3 *Assume $k \geq 3$, and B contains a skip a, b with gap $b - a - 1 \geq 2$ and $(a + b)/2 = k/2$. Then the path problem is NP-complete.*

PROOF: Consider a gadget defined by a path $x_1, x_2, \ldots, x_{b-a}$ and bidirected edges $x_i y_i$ for each $1 \leq i \leq b-a$. This gadget just forbids the assignments to the x_i that are either all 0 or all 1, as x_i, y_i must be complementary 0,1 for each i. This gadget defines thus on the x_i an NP-complete relation from Schaefer's classification, and we may join in an instance multiple occurrences of the same variable in different relational tuples by a path of even length at least k of bidirected edges. ■

Theorem 4 *Assume $k \geq 3$, and B contains a skip a, b with $a < k/2$ and $b > k/2$ having $b - a - 1 \geq 2$. Then the path problem is NP-complete.*

PROOF: Assume by symmetry that $(a+b)/2 \leq k/2$. Let b' be such that $(a+b')/2 = k/2$. Consider a gadget defined by a path $x_1, x_2, \ldots, x_{b'-a}$ and bidirected edges $x_i y_i$ for each $1 \leq i \leq b' - a$. This gadget just forbids the assignments to the x_i that are either all 0s or exactly $b - a$ 1s or all 1s, but allows at least the assignments that are exactly one 1 and the assignment $1010^{b'-a-3}$. This gadget defines thus on the x_i an NP-complete relation from Schaefer's classification, since the problem cannot be (1) linear, as adding $1000^{b'-a-3}, 0010^{b'-a-3}, 1010^{b'-a-3}$ gives $0^{b'-a}$ which is forbidden, (2) 2-satisfiability, as no pair 00 among the $b - a$ x_i is forbidden, so all 0 would be allowed, (3) Horn, as combining

$100^{b'-a-2}, 010^{b'-a-2}$ gives $0^{b'-a}$ which is forbidden, or (3) dual Horn, since combining all $x0^{b'-b}$ where x is sequences of length $b-a$ containing exactly one 1 gives $1^{b-a}0^{b'-a}$ which is forbidden. We may join in an instance multiple occurrences of the same variable in different relational tuples by a path of even length at least k of bidirected edges. ■

Theorem 5 *Assume $k \geq 3$, and the maximum skip of a gap in B is 1. Then the path problem is NP-complete.*

PROOF: Let T consist of a path $x_1, x_2, \ldots, x_k$ combined with paths $x_k, y_1^i, y_2^i, \ldots, y_i^i$ for $1 \leq i \leq k$. For each $1 \leq i \leq k$, we may show by induction on j that y_j^i gets the same value as x_j from V, using the fact that valid numbers of 1s on paths of length k differ by at least 2. Then $y_1^1, y_2^2, \ldots, y_k^k$, at the end of the paths, have the values of $x_1, x_2, \ldots, x_k$, so the number of 1s among these must be in B, giving an NP-complete relation of Schaefer's proof that can be used in an instance R. ■

To handle the case where the maximum skip has gap $b-a-1 > 1$ (as gap of 1 was shown in Theorem 5) not in the middle (as gap in the middle was show in Theorem 3, we notice the following. If we can construct a gadget that will give different values to different source vertices in a digraph $x_0 \neq x_1$, then joining x_0 and x_1 with a forward edge to a gadjet for B replaces B with B' having $k' = k-1$ and $i \in B'$ if and only if $i \in B$ and $i+1 \in B$. Thus the maximum gap is increased by 1. We shall carry this process for B having maximum gap $b-a-1 \leq 4$.

Theorem 6 *Assume the maximum skip of a gap in B is 2. Then the path problem is NP-complete.*

PROOF: Consider skips with gap $b-a-1 = 2$. Assume $b-1 < k/2$ (the symmetric case $a+1 > k/2$ is similar) Construct a cycle C of length $k-2$ given by $x_1 x_2 \cdots x_{k-2}$, join all vertices of C with backward edges to u, v, w, and add the path yvw. If this cycle has c zeros, then it will contain consecutive 00 and 01 (or consecutive 11 and 01 similarly). Therefore $a < c-1, c < b$, so u, v, w get assigned value 0 and y gets assigned value 1. (Symmetrically for consecutive 11 and 01 we assign 1 to u, v, w and 0 to y.) Thus the sources u, y get different values giving the reduction to gap 1 and NP-completeness. ■

The case of skips with larger gaps $b-a-1 \geq 3$ and no middle skip with gap at least 2 remains open. We solve three more cases.

Theorem 7 *Assume $k = 7$ and there is a skip a, b having $a = 0, b = 4$ and gap $b-a-1 = 3$. Then the path problem is NP-complete.*

PROOF: Consider a gadget consisting of two directed cycles C, C' of length 3, all edges from C to C', all edges from C' to u, v_1, and a directed path $v_1 v_2 v_3$. Then the cycles C, C' will be assigned exactly one 1 and two 0s, u, v_1, v_2 will be assigned 0, and v_3 will be assigned 1. So u, v_3 are sink vertices assigned 0 and 1 respectively, thus giving a reduction as before from the case $k' = 6$ having a skip a', b' with $a' = 0, b' = 3$ and gap $b' - a' - 1 = 2$. ■

This completes the cases with $k \leq 7$.

Theorem 8 *Assume $k = 8$ and there is a skip a, b having $a = 0, b = 4$ and gap $b-a-1 = 3$. Then the path problem is NP-complete.*

PROOF: Assume first there is no skip a', b' having $a' = 4, b' = 8$ and gap $b' - a' - 1 = 3$. Consider a gadget consisting of two directed cycles C, C' of length 5, and if $C = (v_1, \ldots, v_5)$, $C' = (v'_1, \ldots, v'_5)$, all edges $v_i v'_j$ except for $v_1 v'_2$, $v_2 v'_1$, and all edges from C' to vertices u, w_1 plus an edge $w_1 w_2$. Then the cycles C, C' will be assigned $(0,1,0,0,0), (1,0,0,0,0)$ respectively, u, w_1 will be assigned 1 and w_2 will be assigned 0. So u, w_2 are sink vertices assigned 1 and 0 respectively, thus giving a reduction as before from the case $k' = 7$ having a skip a'', b'' with $a'' = 0, b'' = 3$ and gap $b'' - a'' - 1 = 2$.

Assume last there is a skip a', b' having $a' = 4, b' = 8$ and gap $b' - a' - 1 = 3$. Then either C, C' are both assigned a single 1 or they are both assigned a single 0, so u, w_2 are sink vertices assigned either $1, 0$ respectively or $0, 1$ respectively, thus giving a reduction as before from the case $k' = 7$ having skips a'', b'' with $a'' = 0, b'' = 3$ and a''', b''' with $a''' = 4, b''' = 7$, where the gaps are $b'' - a'' - 1 = 2$ and $b''' - a''' - 1 = 2$. ■

This completes the cases with $k \leq 8$.

Theorem 9 *Assume $k = 9$ and there is a skip a, b having either $a = 0, b = 5$ and gap $b - a - 1 = 4$, or $a = 1, b = 5$ and gap $b - a - 1 = 3$, $a = 0, b = 4$ and gap $b - a - 1 = 3$. Then the path problem is NP-complete.*

PROOF: Suppose first there is a skip a, b having $a = 0, b = 5$ and gap $b - a - 1 = 4$ (the case $a = 4, b = 9$ is similar). Consider a gadget consisting of two cycles C, C' of length 5, all edges from C to C', a vertex u and an edge vw, and all edges from C' to u, v. Then C, C' will each be assigned a single 1, u, v will be assigned 1, and w will be assigned 0. Thus u, w are sinks assigned 1,0 respectively and the induction to $k' = 8$ with skip a', b' having $a' = 0, b' = 4$ and gap $b' - a' - 1 = 3$ follows.

Suppose next there is a skip a, b having $a = 1, b = 5$ and gap $b - a - 1 = 3$ (the case $a = 4, b = 8$ is similar). Consider a gadget consisting of two cycles C, C' of length 3, all edges from C to C', a vertex u and an edge vw, and all edges from C' to u, v. Then C, C' will each be assigned a single 1, u, v will be assigned 0, and w will be assigned 1. Thus u, w are sinks assigned 0,1 respectively and the induction to $k' = 8$ with skip a', b' having $a' = 1, b' = 4$ and gap $b' - a' - 1 = 2$ follows.

Suppose finally there is a skip a, b having $a = 0, b = 4$ and gap $b - a - 1 = 3$ (the case $a = 5, b = 9$ is similar). Consider first the case where there is no skip a', b' having $a' = 5, b' = 9$ and gap $b' - a' - 1 = 3$. Consider a gadget consisting of two cycles C, C' of length 6 given by $(v_1, \ldots, v_6)$, $(v'_1, \ldots, v'_6)$ respectively, all edges from C to C' except for $v_3 v'_1$, $v_1 v'_2$, $v_2 v'_2$, $v_2 v'_3$, a vertex u and an edge vw, and all edges from C' to u, v. Then C, C' will be assigned $(0,0,1,0,0,0)$, $(1,0,0,0,0,0)$ respectively, u, v will be assigned 1, and w will be assigned 0. Thus u, w are sinks assigned 1,0 respectively and the induction to $k' = 8$ with skip a'', b'' having $a'' = 0, b'' = 3$ and gap $b'' - a'' - 1 = 2$ follows. Finally if there is also a skip a', b' having $a' = 5, b' = 9$ and gap $b' - a' - 1 = 3$, then either C, C' are each assigned a single 1 and u, w are sinks assigned $1, 0$ respectively, or C, C' are each assigned a single 0 and u, w are sinks assigned $0, 1$ respectively. Thus the induction to $k' = 8$ with skips a'', b'' having $a'' = 0, b'' = 3$ and a''', b''' having $a''' = 5, b''' = 8$, gaps $b'' - a'' - 1 = b''' - a''' - 1 = 2$, follows. ■

This completes the cases with $k \leq 9$.