

Complexity of the homomorphism extension problem in the random case

Alexandr Kazda*

Received: October 30, 2011; revised: July 31, 2012, and July 14, 2013; published: July 18, 2013.

Abstract: We prove that if \mathbb{A} is a large random relational structure (with at least one relation of arity at least 2) then the homomorphism extension problem $\text{EXT}(\mathbb{A})$ is almost surely NP-complete.

Key words and phrases: homomorphism, constraint satisfaction problem, random digraph

1 Introduction

The complexity of the constraint satisfaction problem (CSP) with a fixed target structure is a well established field of study in combinatorics and computer science (see [4] for an overview). In the last decade, we have seen algebraic tools brought to bear on the question of CSP complexity, yielding major new results (see e.g. [2], [3], [1]).

In the algebraic approach, it is customary to study relational structures that contain all possible constants. If \mathbb{A} is such a structure and we are to decide the existence of a homomorphism $f : \mathbb{B} \rightarrow \mathbb{A}$ then the constant constraints prescribe values for f at some vertices of \mathbb{B} . We are thus deciding if some partial homomorphism $f_c : \mathbb{B} \rightarrow \mathbb{A}$ can be extended to the whole \mathbb{B} . Therefore, $\text{CSP}(\mathbb{A})$ becomes the *homomorphism extension problem* with target structure \mathbb{A} , denoted by $\text{EXT}(\mathbb{A})$. It is easy to see that $\text{CSP}(\mathbb{A})$ reduces to $\text{EXT}(\mathbb{A})$, since in CSP we extend the empty partial homomorphism.

In [5], the authors prove that $\text{CSP}(\mathbb{A})$ is almost surely NP-complete when \mathbb{A} is a large random relational structure with at least one at least binary relation and without loops. We show, by a different method, that the same hardness result holds for $\text{EXT}(\mathbb{A})$ even if we allow loops.

*The author would like to thank Libor Barto for suggesting the problem. Supported by the Czech-Polish cooperation grant 7AMB13PL013 “General algebra and applications”, by the Charles University project GA UK 558313, the GAČR project 13-01832S, and the SVV-2011-263317 grant.

2 Preliminaries

A *relational structure* \mathbb{A} is any set A together with a family of relations $\{R_i : i \in I\}$ where $R_i \subset A^{n_i}$. We call the number n_i the *arity* of R_i . The sequence $(n_i : i \in I)$ determines the *similarity type* of \mathbb{A} . We consider only finite structures (and finitary relations) in this paper. We use the notation $[n] = \{1, 2, \dots, n\}$.

Let $\mathbb{A} = (A, \{R_i : i \in I\})$ and $\mathbb{B} = (B, \{S_i : i \in I\})$ be two relational structures of the same similarity type. A mapping $f : A \rightarrow B$ is a homomorphism if for every $i \in I$ and every $(a_1, \dots, a_{n_i}) \in R_i$ we have $(f(a_1), \dots, f(a_{n_i})) \in S_i$.

Let us fix some $p \in (0, 1)$ and let A be a set. The relation $S \subset A^l$ is an *l -ary random relation* on A if every possible l -tuple belongs to S with probability p (independently of other l -tuples). We will call any relational structure with one or more random relations a *random relational structure*. In particular, a random relational structure with just one binary relation is a *random digraph*.

The *Constraint Satisfaction Problem* with the target structure \mathbb{A} , denoted by $\text{CSP}(\mathbb{A})$, consists of deciding whether a given input relational structure \mathbb{B} of the same similarity type as \mathbb{A} can be homomorphically mapped to \mathbb{A} . It is easy to come up with examples of \mathbb{A} such that $\text{CSP}(\mathbb{A})$ is NP-complete and this is in a sense typical behavior as proved in [5]: If $R(n, k)$ is a k -ary random relation on the set $[n]$ (with $p = 1/2$) that does not contain any elements of the form (a, a, \dots, a) for $a \in A$ then

$$\forall k \geq 2, \lim_{n \rightarrow \infty} \text{Prob}(\text{CSP}([n], R(n, k)) \text{ is NP-complete}) = 1, \quad (2.1)$$

$$\forall n \geq 2, \lim_{k \rightarrow \infty} \text{Prob}(\text{CSP}([n], R(n, k)) \text{ is NP-complete}) = 1. \quad (2.2)$$

There is a reason why the authors of [5] disallow loops: If \mathbb{A} has only one relation R and R contains a loop (a, a, \dots, a) then every \mathbb{B} of the same similarity type as \mathbb{A} can be homomorphically mapped to \mathbb{A} simply by sending everything to a , so $\text{CSP}(\mathbb{A})$ is very simple to solve.

Given a target structure \mathbb{A} , the *Homomorphism Extension Problem* for \mathbb{A} , denoted by $\text{EXT}(\mathbb{A})$, consists of deciding whether a given input structure \mathbb{B} and a given partial mapping $f : \mathbb{B} \rightarrow \mathbb{A}$ can be extended to a homomorphism from \mathbb{B} to \mathbb{A} .

Let \mathbb{A} be a relational structure and $a \in A$. The *constant relation* c_a is the unary relation consisting only of a , i.e. $c_a = \{(a)\}$. When searching for a homomorphism to \mathbb{A} , the relation c_a prescribes a set of elements of B that must be mapped to a . It is an easy observation that if \mathbb{A} contains constant relations for each of its elements (as is usual in the algebraic treatment of CSP) then $\text{CSP}(\mathbb{A})$ and $\text{EXT}(\mathbb{A})$ are essentially the same problem.

Since the homomorphism extension problem is quite important to algebraists, it makes sense to ask what is the typical complexity of $\text{EXT}(\mathbb{A})$. We will use the phrase “ $\text{EXT}(\mathbb{A})$ is almost surely NP-complete for n large” as abbreviation for “For each $n \in \mathbb{N}$, there exists a random relational structure \mathbb{A}_n (whose precise definition is obvious from the context) such that we have

$$\lim_{n \rightarrow \infty} \text{Prob}(\text{EXT}(\mathbb{A}_n) \text{ is NP-complete}) = 1.”$$

Because additional relations do not make CSP easier to solve, the limit (2.1) gives us that that $\text{EXT}(\mathbb{A})$ is almost surely NP-complete if \mathbb{A} is a large random relational structure with no loops and at least one relation of arity greater than one. In the remainder of the paper we show that we can allow loops without making the problem any easier.

3 The problem EXT for random digraphs

We will begin by investigating random digraphs and then generalize our findings to all relational structures.

Theorem 3.1. *Let G be a random digraph on n vertices. Then $\text{EXT}(G)$ is almost surely NP-complete for n large.*

Proof. Let $G = (V, E)$ be a digraph. Consider G as a relational structure and add to G every constant relation possible. Let $v_1, \dots, v_l \in V(G)$. Consider the set

$$F_{v_1, \dots, v_l} = \{u \in V(G) : \forall i, (v_i, u) \in E(G)\}$$

We will call this set a *subalgebra* of G .

Note that the sets F_{v_1, \dots, v_l} are subalgebras, in the sense of Universal Algebra, and our technique can be greatly generalized to all primitive positive definitions (see [2]). For our proof, however, we need a lot less: Assume that for some choice of v_1, \dots, v_l the subalgebra F_{v_1, \dots, v_l} induces a loopless triangle in G . We claim that we can then reduce graph 3-colorability to $\text{EXT}(G)$, making $\text{EXT}(G)$ NP-complete.

Let H be a graph whose 3-colorability we wish to test. We then consider H as a symmetric digraph and add to H new vertices w_1, \dots, w_l and new edges (w_i, u) for each $i \in \{1, \dots, l\}$ and all $u \in V(H)$, obtaining the digraph H' . Our $\text{EXT}(G)$ instance will then consist of the digraph H' along with the partial map f which maps each w_i to v_i . Now f can be extended to a homomorphism if and only if H can be homomorphically mapped into the triangle induced by F_{v_1, \dots, v_l} which happens if and only if H is 3-colorable.

All we need to do now is to show that G almost surely contains a subalgebra that induces a triangle. Our aim, roughly speaking, is to show that G almost surely contains many three element subalgebras because then there is a large chance that at least one of these subalgebras will be a triangle.

We will partition $V(G)$ into two sets $A = \{1, \dots, \lfloor n/2 \rfloor\}$ and $B = \{\lfloor n/2 \rfloor + 1, \dots, n\}$. We will now use points of A to define subalgebras lying in B . Denote by S_k the event “ G contains at least k disjoint three-element subalgebras of the form $F_{v_1, \dots, v_l} \subset B$ for some $v_1, \dots, v_l \in A$.” We can write

$$S_k = \bigcup_{\substack{C_1, \dots, C_k \subset B \\ \forall i \neq j, C_i \cap C_j = \emptyset \\ \forall i, |C_i| = 3}} S_{C_1, \dots, C_k},$$

where S_{C_1, \dots, C_k} is the event “The sets C_1, \dots, C_k are subalgebras of G ”. Finally, denote by T_{C_1, \dots, C_k} the event “There exists an $i \in \{1, 2, \dots, k\}$ such that the set C_i induces a triangle subgraph of G .”

Since a probability that a fixed C_i induces a triangle is $p^6(1-p^3)$, the probability of the event T_{C_1, \dots, C_k} is (for C_1, \dots, C_k pairwise disjoint three element sets)

$$\text{Prob}(T_{C_1, \dots, C_k}) = 1 - (1 - p^6(1 - p^3))^k,$$

which tends to 1 when k goes to infinity.

Observe that the event S_{C_1, \dots, C_k} is independent from the event T_{C_1, \dots, C_k} for each choice of $C_1, \dots, C_k \subset B$ since both events talk about disjoint sets of edges of G .

Assume for a moment that for all $k \in \mathbb{N}$ the value of $\text{Prob}(S_k)$ tends to 1 as n tends to infinity. Then, given an $\varepsilon > 0$, we choose k so that $\text{Prob}(T_{C_1, \dots, C_k}) \geq 1 - \varepsilon$. When n is large enough, the digraph G contains some k pairwise disjoint three element subalgebras C_1, \dots, C_k with probability at least $1 - \varepsilon$. The probability that one of the sets C_1, \dots, C_k induces a triangle is $T_{C_1, \dots, C_k} \geq 1 - \varepsilon$. Thus we get an NP-complete CSP problem with probability at least $(1 - \varepsilon)^2 > 1 - 2\varepsilon$ and since ε was arbitrary, we see that for large n the homomorphism extension problem is almost surely NP-complete.

It remains to show $\lim_{n \rightarrow \infty} \text{Prob}(S_k) = 1$ for all k . Fix the value of k . For each value of n , let l be the integer satisfying $np^l \geq 1 > np^{l+1}$. We will now search for the three element subalgebras of B for n large. We proceed in steps: Assume that after i steps we have already found m such subalgebras C_1, \dots, C_m . In the $(i + 1)$ -th step, we take the vertices $1 + il, 2 + il, \dots, l + il$ of A and consider the subalgebra $F_{1+il, 2+il, \dots, l+il}$. If this subalgebra lies in B , has size three and is disjoint with all the sets C_1, \dots, C_m , we let $C_{m+1} = F_{1+il, 2+il, \dots, l+il}$, increase m by one and continue with the next step. Otherwise, $F_{1+il, 2+il, \dots, l+il}$ is not a good candidate for C_{m+1} , so we leave m unchanged and continue with the next step.

What is the probability that we find the $(m + 1)$ -th subalgebra in a given step? Every vertex of G is in $F_{1+il, 2+il, \dots, l+il}$ with the probability p^l . The probability that $F_{1+il, 2+il, \dots, l+il}$ consists of three yet-unused vertices of B is then equal to

$$q = \binom{|B| - 3 \cdot m}{3} p^{3l} (1 - p^l)^{n-3} \geq \frac{(n/2 - 3m - 3)^3}{6} p^{3l} (1 - p^l)^n$$

If $m \geq k$, we have already won, so assume $m < k$:

$$q \geq \frac{(n/2 - 3k)^3}{6} p^{3l} (1 - p^l)^n = \frac{(1/2 - 3k/n)^3}{6} n^3 p^{3l} (1 - p^l)^n$$

Now let $r = \frac{(1/2 - 3k/n)^3}{6}$ and observe that $r > 0$ for n large enough. Using the the inequalities $np^l \geq 1 > np^{l+1}$ we have:

$$q \geq rn^3 p^{3l} (1 - p^l)^n \geq r(1 - p^l)^n > r \left(1 - \frac{1}{pn}\right)^n.$$

The lower bound on q tends to $r/e^{1/p}$ as n tends to infinity, so there exists a δ such that $q > \delta > 0$ for all n large enough.

Therefore, the probability of producing a new three-element subalgebra in a given step is at least $\delta > 0$ and this bound does not depend on the number of subalgebras we have already found. Now observe that l is approximately $\log_{1/p} n$ and therefore we have enough vertices in A for approximately $s = \frac{n}{2 \log_{1/p} n}$ steps. If we choose n large enough, we can have s as large as we want and so the probability of finding at least k subalgebras can be arbitrarily close to 1. Therefore, $\lim_{n \rightarrow \infty} \text{Prob}(S_k) = 1$, concluding the proof. \square

4 Random relational structures

It is easy to see that if \mathbb{A} is a relational structure with unary relations only then $\text{EXT}(\mathbb{A})$ is always polynomial. We would now like to investigate the case of relations of arity greater than two. Intuition tells us that greater arity means greater complexity. The intuition is right.

Lemma 4.1. *Let $k \geq 2$, n be large and let $\mathbb{A} = ([n], S)$ be a relational structure with S a random k -ary relation. Then the homomorphism extension problem $\text{CSP}(\mathbb{A})$ is almost surely NP-complete.*

Proof. We have proven the result for $k = 2$. If $k > 2$, consider the binary relational structure $\mathbb{B} = ([n], R)$ where $R = \{(x, y) \in [n]^2 : (x, y, 1, 1, \dots, 1) \in S\}$. It is easy to see that if S is a random k -ary relation then \mathbb{B} is a random digraph where each edge exists with probability p . From Theorem 3.1 we see that $\text{EXT}(\mathbb{B})$ is almost surely NP-complete. We will now show how to reduce $\text{EXT}(\mathbb{B})$ to $\text{EXT}(\mathbb{A})$ in polynomial time, proving that $\text{EXT}(\mathbb{A})$ is almost surely NP-complete.

Using algebraic tools, the reduction of $\text{EXT}(\mathbb{B})$ to $\text{EXT}(\mathbb{A})$ follows from the fact that R is defined by a primitive positive formula that uses only S and the constant 1. However, we will provide an elementary reduction here: Let $\mathbb{C} = (C, T)$ be a relational structure with a single binary relation T and let $f : C \rightarrow [n]$ be a partial mapping. We add to C a new element e , construct the k -ary relation $U = \{(x, y, e, e, \dots, e) : (x, y) \in T\}$ and the partial mapping $g : C \cup \{e\} \rightarrow [n]$ so that $g|_C = f$ and $g(e) = 1$. It is straightforward to show that g can be extended to a homomorphism $(C \cup \{e\}, U) \rightarrow \mathbb{A}$ if and only if f can be extended to a homomorphism $(C, T) \rightarrow \mathbb{B}$, concluding the proof. \square

Additional relations in \mathbb{A} do not make $\text{EXT}(\mathbb{A})$ easier, so we have the most general version of our NP-completeness result:

Corollary 4.2. *Let \mathbb{A} be the relational structure $([n], \{R_i : i \in I\})$ where at least one R_i is a random relation of arity greater than one. Then $\text{EXT}(\mathbb{A})$ is almost surely NP-complete for n large.*

As a final note, we will now prove the analogue of the limit (2.2) for EXT .

Corollary 4.3. *Let us fix a set A of at least two elements and let $\mathbb{A} = (A, R)$ be a relational structure with R random k -ary relation. Then $\text{EXT}(\mathbb{A})$ is almost surely NP-complete for k large.*

Proof. Assume first that k is even and let $m = k/2$.

Consider the relational structure $\mathbb{B} = (A^m, S)$ with

$$S = \{((a_1, \dots, a_m), (a_{m+1}, \dots, a_{2m})) : (a_1, \dots, a_{2m}) \in R\}.$$

It is straightforward to prove that S is a binary random relation on A^m and therefore $\text{EXT}(\mathbb{B})$ is almost surely NP-complete for large even k . What is more, $\text{EXT}(\mathbb{B})$ can be easily reduced to $\text{EXT}(\mathbb{A})$: If $\mathbb{C} = (C, T)$ is a relational structure with T binary and $f : C \rightarrow A^m$ is a partial mapping, we construct the structure $\mathbb{C}' = (C', T')$ with

$$\begin{aligned} C' &= \{(c, i) : c \in C, i \in \{1, \dots, m\}\}, \\ T' &= \{((c, 1), \dots, (c, m), (d, 1), \dots, (d, m)) : (c, d) \in T\} \end{aligned}$$

and a partial mapping $g : C' \rightarrow A$ such that $g(c, i) = a_i$ whenever $f(c)$ is defined and equal to (a_1, \dots, a_m) .

It is easy to see that g can be extended to a homomorphism from \mathbb{C}' to \mathbb{A} if and only if f can be extended to a homomorphism from \mathbb{C} to \mathbb{A} .

In the case that $k = 2m + 1$, we fix an $e \in A$, choose $\mathbb{B} = (A^m, S)$ with

$$S = \{((a_1, \dots, a_m), (a_{m+1}, \dots, a_{2m})) : (a_1, \dots, a_{2m}, e) \in R\}$$

and proceed similarly to the previous case.

We see that for a large enough k , no matter if it is odd or even, the problem $\text{EXT}(\mathbb{B})$ is almost surely NP-complete. \square

5 Conclusions

We have shown that the homomorphism extension problem is almost surely NP-complete for large relational structures (assuming we have at least one non-unary relation). In a sense, our result is not surprising since the relational structures we consider are very dense, so it stands to a reason that we can find hard instances most of the time.

It might therefore be interesting to see what is the complexity of CSP or EXT for large structures obtained by other random processes, particularly when relations are sparse. Such structures might better correspond to “typical” cases of CSP or EXT encountered in practice. Some such results already exist; see [6] for a criterion on the random graph process to almost surely produce projective graphs (if G is projective then $\text{EXT}(G)$ is NP-complete, see [2]). Our guess is that both CSP and EXT will remain to be almost surely NP-complete in all the nontrivial cases.

References

- [1] LIBOR BARTO AND MARCIN KOZIK: Constraint satisfaction problems of bounded width. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science, FOCS'09*, pp. 595–603, Atlanta, 2009. IEEE. [doi:10.1109/FOCS.2009.32] 1
- [2] ANDREI BULATOV, PETER JEAVONS, AND ANDREI KROKHIN: Classifying the complexity of constraints using finite algebras. *SIAM Journal on Computing*, 34(3):720–742, 2005. [doi:10.1137/S0097539700376676] 1, 3, 6
- [3] BENOIT LAROSE AND PASCAL TESSON: Universal algebra and hardness results for constraint satisfaction problems. *Theoretical Computer Science*, 410(18):1629–1647, 2009. [doi:10.1016/j.tcs.2008.12.048] 1
- [4] JAROSLAV NEŠETŘIL AND PAVOL HELL: *Graphs and homomorphisms*. Oxford Lecture Series in Mathematics and Its Applications. Oxford University Press, New York, 2004. 1
- [5] JAROSLAV NEŠETŘIL AND TOMASZ ŁUCZAK: A probabilistic approach to the dichotomy problem. *SIAM Journal on Computing*, 36:835–843, 2006. [doi:10.1137/S0097539703435492] 1, 2
- [6] JAROSLAV NEŠETŘIL AND TOMASZ ŁUCZAK: When is a random graph projective? *European Journal of Combinatorics*, 27(7):1147–1154, 2006. [doi:10.1016/j.ejc.2006.06.010] 6

AUTHOR

Alexandr Kazda
 Researcher
 Charles University, Prague, Czech Republic
 alexak@atrey.karlin.mff.cuni.cz
<http://atrey.karlin.mff.cuni.cz/~alexak/english.html>