On Approximate Horn Formula Minimization

Amitava Bhattacharya[1], Bhaskar DasGupta[2,⋆], Dhruv Mubayi[3,⋆⋆], and György Turán[3,4,⋆⋆⋆]

[1] School of Mathematics, Tata Institute of Fundamental Research, Colaba, Mumbai 400 005, INDIA, amitava@math.tifr.res.in
[2] Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7053, dasgupta@cs.uic.edu
[3] Department of Mathematics, Statistics & Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7045, mubayi@math.uic.edu gyt@uic.edu
[4] Research Group on Artificial Intelligence, University of Szeged, Hungary.

Abstract

The minimization problem for Horn formulas is to find a Horn formula equivalent to a given Horn formula, using a minimum number of clauses. A $2^{\log^{1-\epsilon}(n)}$-inapproximability result is proven, which is the first inapproximability result for this problem. We also consider several other versions of Horn minimization. The more general version which allows for the introduction of new variables is known to be too difficult as its equivalence problem is co-NP-complete. Therefore, we propose a variant called *Steiner-minimization*, which allows for the introduction of new variables in a restricted manner. Steiner-minimization of Horn formulas is shown to be MAX-SNP-hard. In the positive direction, a $o(n)$, namely, $O(n \log\log n/(\log n)^{1/4})$-approximation algorithm is given for the Steiner-minimization of definite Horn formulas. The algorithm is based on a new result in algorithmic extremal graph theory, on partitioning bipartite graphs into complete bipartite graphs, which may be of independent interest. Inapproximability results and approximation algorithms are also given for restricted versions of Horn minimization, where only clauses present in the original formula may be used.

1 Introduction

The CNF minimization problem is to find a shortest CNF expression equivalent to a given expression. This problem has been studied in different versions for many decades in switching theory, computer science and engineering, and it is still a topic of active research, both in complexity theory and circuit design. Umans [40, 41] showed Σ_p^2-completeness and a $O(n^{1-\varepsilon})$-inapproximability result for this problem. Horn minimization is the special case of CNF minimization for Horn formulas. Horn formulas are conjunctions of *Horn clauses*, *i.e.*, of disjunctions containing *at most one* unnegated variable. Horn clauses can also be written as implications. For instance, $\bar{a} \vee \bar{b} \vee c$ is a Horn clause which can also be written as $a, b \to c$.

⋆ Supported by NSF Grant IIS-0346973 and DIMACS special focus on Computational and Mathematical Epidemiology.
⋆⋆ Supported by NSF grant DMS-0653946.
⋆⋆⋆ Supported by NSF grant CCF-0916708.

Horn formulas are an expressive and tractable fragment of propositional logic, and therefore provide a basic framework for knowledge representation and reasoning [35]. Horn formulas are, for example, a natural framework for representing systems of rules for expert systems. An interesting potential new application area for Horn formulas is the automated, interactive development of large-scale knowledge bases of commonsense knowledge (see [36] for the description of such a project). This application has algorithmic aspects involving knowledge representation, reasoning, learning and knowledge update. A model incorporating these aspects, called *Knowledge Base Learning (KnowBLe)* is formulated in [29] (see also [28, 30]) for related work). Efficient algorithms for approximate Horn minimization would be useful in these applications.

Satisfiability of Horn formulas can be decided in linear time and the equivalence of Horn formulas can be decided in polynomial time [25]. Thus Horn minimization is expected to be easier than CNF minimization. Horn minimization was shown to be NP-complete by Hammer and Kogan [21] if the number of literals is to be minimized, and by Ausiello *et al.* [4] and Boros and Čepek [8] if the number of clauses is to be minimized. On the positive side, Hammer and Kogan [22] gave a polynomial algorithm for minimizing quasi-acyclic Horn formulas, which include both acyclic and 2-Horn formulas. It was also shown in [21] that there is an efficient $(n-1)$-approximation algorithm for general Horn minimization, where n is the *number of different variables* in the formula (not the number of variable occurrences). As noted in [18], such an algorithm is also provided by the Horn formula learning algorithm of [3].

1.1 Contributions of this paper

First, in Theorem 3 we prove a $2^{\log^{1-\epsilon}(n)}$-inapproximability result for Horn minimization assuming NP$\not\subseteq$DTIME$(n^{\mathrm{polylog(n)}})$ via a reduction from the MINREP problem [26]. This seems to be the first inapproximability result for this problem. We next consider several other versions of the Horn minimization problem. Depending on the application, different versions may be relevant and thus of interest for exploration of their approximability properties.

It may be possible to *add new variables* in order to compress the formula. For example, the formula

$$\varphi = \bigwedge_{i=1}^{n} \bigwedge_{j=1}^{n} (x_i \to y_j) \tag{1}$$

having n^2 clauses can be compressed to the $2n$ clause formula

$$\psi = \bigwedge_{i=1}^{n} (x_i \to z) \wedge \bigwedge_{j=1}^{n} (z \to y_j), \tag{2}$$

where z is a new variable. Note that φ and ψ are clearly *not* equivalent, *e.g.*, φ does not depend on z, while ψ does. On the other hand, φ and ψ *are* equivalent in the sense that they both imply the same set of clauses over the original variables $x_1, \ldots, x_n, y_1, \ldots, y_n$. Thus, in terms of the knowledge base application, the new variable z can be thought of as being *internal* to the knowledge base and invisible to the user. Flögel *et al.* [17] showed that deciding the equivalence of such extended Horn formulas is co-NP-complete. This is bad news as it shows that the extended version is too expressive and therefore intractable[5].

On the other hand, notice that in the example above the new variable is added in a rather *restricted manner*. Formula (1) can be thought of as a complete directed bipartite graph with parts $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_n\}$. Formula (2), then, represents the operation of adding a new node z in the middle, with edges from the x_i's to z and from z to the y_j's. The two graphs have the same reachability relations as far as the original nodes are concerned. Using the similarity to Steiner problems where new points may be added [23], we refer to this operation as a *Steiner extension* (a formal definition appears in Section 5). As we observe, in contrast to general extensions, the equivalence of Steiner extensions *can* be decided efficiently (Corollary 1). Thus this type of extension could be considered as a tractable alternative in the applications mentioned. The *Steiner minimization* problem for Horn formulas is then to find an equivalent Steiner-extended Horn formula, for a given Horn formula, with a minimum number of clauses. We show in Theorem 5 that this problem is MAX-SNP-hard. On the other hand, in Theorem 6 we prove that there is an efficient $O(n \log\log n/(\log n)^{1/4})$-approximation algorithm for this problem, where n is the number of variables in the original formula. This is the first approximation algorithm for Horn minimization with a $o(n)$ approximation guarantee.

The algorithm for Steiner minimization makes use of an algorithmic result on the *partition of bipartite graphs into complete bipartite graphs (*i.e.*, bicliques)*, which may be of interest on its own. It was shown by Chung, Erdős and Spencer [13] and Bublitz [10] that the edges of every n-vertex graph can be partitioned into complete bipartite graphs such that the sum of the number of vertices in these bipartite graphs[6] is $O(n^2/\log n)$, and this is asymptotically the best possible. Tuza [39] gave an analogous result for bipartite graphs. These results are based on a counting argument due to Kővári, Sós and Turán [27], which shows that sufficiently dense graphs contain large complete bipartite subgraphs, and thus are non-constructive. Kirchner [24] considered the problem of finding an algorithmic version, and gave an efficient algorithm to find complete balanced bipartite graphs of size $\Omega(\sqrt{\log n})$ in dense graphs. In a previous paper [33] we improved this to the optimal $\Omega(\log n)$, and as a corollary, showed that partitions proved to *exist* in [10, 13] can also be *found efficiently*[7].

[5] Introducing new variables has been considered earlier, going back to Tseitin [38]. While there are many exponential lower bounds for resolution proofs (see, *e.g.*, [14]), complexity-theoretic results suggest that proving such results for extended resolution is much harder [15].

[6] Note that the complexity of a partition is not measured by the number of graphs in it, but by a different measure, which comes from circuit complexity [37].

[7] Extremal combinatorics provides results on the existence of substructures. The results of [24] and [33] can be viewed as *algorithmic extremal combinatorics* as they also give efficient algorithms to actually find such substructures. Previous results in this direction are given in [2]. The results of [2] apply to dense graphs and find substructures of constant size, while here we have to handle sparser graphs as well and to find substructures of nonconstant size.

In this paper we give an algorithmic version for the bipartite case. We show in Theorem 1 that the edges of every bipartite graph with sides a and b, where $a \geq b$, can be partitioned into complete bipartite graphs such that the sum of the number of vertices of these graphs is $O((ab/\log a) + a \log b + a)$, and we give an efficient algorithm to find such a partition.

We also consider *restricted* versions of the Horn minimization problem, where one is restricted to use clauses from the original formula. Such a restriction may be justified in applications where the particular rules, provided by an expert, are supposed to be meaningful and thus cannot be replaced. The goal is to eliminate redundant rules. Modifying the construction of Theorem 3, in Theorem 7 we prove $2^{\log^{1-\epsilon}(n)}$-inapproximability for the restricted case, which holds even if the input formula has clauses of size at most 3.

One may want to optimize a Horn formula in the restricted sense either by *minimizing the number of rules left*, or by *maximizing the number of rules removed*. The two versions may differ in their approximability (cf. the maximum independent set and the minimum vertex cover problems for graphs). As (1) suggests, Horn formulas with clauses of the form $x \to y$ correspond to directed graphs. For such formulas, optimization corresponds to *transitive reduction* problems for directed graphs. Thus approximation algorithms for these directed graph problems (in both versions) may be applied for Horn formulas. Examples of this connection are given in Theorem 8.

The rest of the paper is organized as follows. Bipartite graph decompositions are discussed in Section 3. Section 4 contains the inapproximability result for Horn minimization. Horn minimization with new variables is discussed in Section 5, and restricted Horn minimization in Section 6.

2 Preliminaries

A *clause* is a disjunction of literals. A *Horn clause* is a clause with at most one unnegated variable. A *definite* Horn clause has exactly one unnegated variable, called its *head*; the other variables form its *body*. A *negative* clause consists of only negated variables. The *size* of a clause is the number of its literals. A clause of size 1 (resp., 2) is a *unit* (resp., *binary*) clause. A *(definite) Horn formula* is a conjunction of (definite) Horn clauses. The *size* of a formula is the number of its clauses. A *k-Horn formula* is a Horn formula with clauses of size at most k.

A clause C is an *implicate* of a formula φ (also written as $\varphi \models C$) if every truth assignment satisfying φ also satisfies C. An implicate is a *prime* implicate if none of its proper subclauses is an implicate. The *resolution* operation takes two clauses of the form $C_1 \vee x$ and $C_2 \vee \bar{x}$ and produces the clause $C_1 \vee C_2$. For basic properties of resolution, see, e.g. [25].

Deciding whether a *definite Horn* clause C is an implicate of a *definite Horn* formula φ can be decided by a simple and well-known marking procedure often called *forward chaining*. The procedure begins by marking the variables in the body of C. If every variable in the body of a clause in φ is marked then its head is marked as well. This is repeated as long as new variables get marked. Then it holds that C is an implicate of φ iff its head gets marked.

3 Partitioning/covering bipartite graphs using bicliques

Let $G = G(A, B, E)$ be a bipartite graph with parts A, B of sizes a, b, respectively, and edge set E of size m. We assume w.l.o.g. that $a \geq b$. A bipartite graph is *balanced* if $|A| = |B|$. The complete bipartite graph (or biclique) with parts of size p and q is denoted by $K_{p,q}$. We consider bicliques $G_i = (A_i, B_i, E_i)$ for $i = 1, \ldots, t$ such that $A_i \subseteq A, B_i \subseteq B$, and $(E_1, \ldots, E_t)$ is a partition, resp. a cover, of E. The cost of such a decomposition is $\sum_{i=1}^{t}(|A_i| + |B_i|)$. The problem is to find a decomposition of small cost. The trivial decomposition into single edges has a cost of $2m \leq 2ab$.

We consider two versions of the problem. In the first version we are interested in finding a partition such that its size is upper bounded by some function of a and b, independent of m.

Theorem 1. *For every bipartite graph G one can find a partition of cost $O\left(\frac{ab}{\log a} + a \log b + a\right)$ in polynomial time.*

The decomposition is found by iteratively finding large bipartite subgraphs. There are two procedures, depending on a carefully chosen notion of density. Let $3 \leq b \leq a$, $6a \leq m \leq ab$ and $f(a, b, m) = \left\lfloor \frac{\log a}{\log(2eab/m)} \right\rfloor$; note that the case $b \leq 2$ is trivial.

Lemma 1. *Suppose that $m \geq af(a, b, m)$. Then there is a polynomial time algorithm that finds a $K_{q,q}$ in G with $q = f(a, b, m)$.*

Lemma 2. *Suppose that $m < af(a, b, m)$. Then there is a polynomial time algorithm that finds a $K_{q,q}$ in G with $q = \lfloor m/a \rfloor$.*

Remark 1. If G is a star then $b = 1$ and the optimal decomposition has cost $a + 1$, hence the upper bound $ab/\log a$ claimed in [39] does not hold, and an additional term (or some other modification) is needed. It is open whether the quantity $a \log b + a$ can be improved.

In the second version we are interested in finding a partition (resp., cover) of minimal cost. For technical reasons, we use a slightly different cost function here (using this cost function would not change anything in the previous result). The size of $K_{p,q}$ is $p \cdot q$ and the modified cost $\text{cost}'(K_{p,q})$ of $K_{p,q}$ is $p \cdot q$ if $p = 1$ or $q = 1$, and is $p + q$ otherwise. The reason for using the modified cost measure for Horn minimization is that when a set of Horn clauses $\wedge_{i=1}^{p} \wedge_{j=1}^{q} (x_i \rightarrow y_j)$ corresponding to a biclique $K_{p,q}$ is replaced by a set of Horn clauses $\wedge_{i=1}^{p}(x_i \rightarrow z) \bigwedge \wedge_{j=1}^{q}(z \rightarrow y_j)$ by introducing a new variable z if $p, q > 1$, and is left unchanged otherwise, the size of the new formula is $\text{cost}'(p, q)$. We define the Linear-Cost-Biclique-Cover (resp., Linear-Cost-Biclique-Partition) problem as follows: given a bipartite graph $G = (A, B, E)$, cover (resp., partition) its edges with bicliques of minimum total *modified* cost. The minimization of the number of bicliques in a cover was shown to be NP-complete by Orlin [34]. The following result follows by an approximation-preserving reduction from the maximum independent set problem for 3-regular graphs.

Theorem 2. *Assuming* P≠NP, LINEAR-COST-BICLIQUE-COVER *and* LINEAR-COST-BICLIQUE-PARTITION *cannot be approximated in polynomial time within an approximation ratio of* $1 + (1/1138)$ *even if the input graph has no biclique of size more than* 6*.*

4 Inapproximability

Theorem 3. *For any fixed* $0 < \epsilon < 1$*, unless* NP⊆DTIME($n^{polylog(n)}$)*, the Horn minimization problem for definite Horn formulas is* $2^{\log^{1-\epsilon} n}$*-inapproximable.*

The reduction is from the MINREP problem [26]. An instance M is given by a bipartite graph $G = (A, B, E)$ with $|E| = m$, a partition of A into equal-size subsets $A_1, A_2, \ldots, A_\alpha$ and a partition of B into equal-size subsets $B_1, B_2, \ldots, B_\beta$. One can define a natural bipartite super-graph H in the following manner. H has a super-vertex for every A_i and B_j. There is a super-edge between A_i and B_j if and only if there exists $u \in A_i$ and $v \in B_j$ such that (u, v) is an edge of G. Let the number of super-edges be p. A pair of nodes u and v *witnesses* a super-edge (A_i, B_j) provided $u \in A_i$, $v \in B_j$ and the edge (u, v) exists in G. A set of nodes S of G witnesses a super-edge if and only if there exists at least one pair of nodes in S that witnesses the super-edge. The goal of MINREP is to find $A' \subseteq A$ and $B' \subseteq B$ such that $A' \cup B'$ witnesses every super-edge of H and $|A'| + |B'|$ is as small as possible. The *size* of an optimal solution is denoted by $OPT(M)$. Let $s = |A| + |B|$. It is shown in Kortsarz *et al.* [26] that MINREP is $2^{\log^{1-\varepsilon} n}$-inapproximable under the complexity-theoretic assumption NP⊈DTIME($n^{\mathrm{polylog(n)}}$).

Consider an instance M of MINREP . Let t be a sufficiently large positive integer to be fixed later. We construct a definite Horn formula φ. For simplicity and with an abuse of notation, some variables in φ are denoted as the corresponding objects (vertices and super-edges) in the MINREP instance. The formula φ contains *amplification* variables $x_1, \ldots, x_t$, *node* variables u for every vertex u in $A \cup B$ and *super-edge variables* e for every super-edge e in H. The clauses of φ belong to the following groups:

amplification clauses: there is a clause $x_i \to u$ for every $i \in \{1, \ldots, t\}$ and for every $u \in A \cup B$,

witness clauses: there is a clause $u, v \to e$ for every super-edge e of H and for every pair of nodes $u \in A$ and $v \in B$ witnessing e,

feedback clauses: there is a clause $e_1, \ldots, e_p \to u$ for every $u \in A \cup B$, where $e_1, e_2, \ldots, e_p$ are the super-edges of H.

As φ is definite, all its prime implicates are definite. Also, as φ consists of non-unit definite clauses, all its prime implicates are non-unit (the all-zero vector satisfies φ and falsifies all unnegated variables). For a further analysis of the prime implicates of φ, we make use of forward chaining.

Lemma 3. *Let x be an amplification variable. Then the prime implicates containing x are clauses of the form $x \to v$, where v is a node or super-edge variable.*

Lemma 4. *Let U be a set of node variables such that U is not a solution to* MINREP *, and U' be the set of super-edge variables witnessed by U. Then every implicate with body contained in $U \cup U'$ has head in U'.*

Lemma 5. *Let x be an amplification variable and let ψ be a prime and irredundant Horn formula equivalent to φ. Then ψ has at least $OPT(M)/2$ clauses containing x.*

Based on these lemmas one can prove that the reduction is gap-preserving.

Lemma 6 (Gap preserving reduction lemma).

(a) *If $OPT(M) = \alpha + \beta$, then $OPT(\varphi) \leq t \cdot (\alpha + \beta) + m + s$.*
(b) *If $OPT(M) \geq (\alpha + \beta) \cdot 2^{\log^{1-\varepsilon} s}$ then, $OPT(\varphi) \geq t(\alpha + \beta) \cdot 2^{\log^{1-\varepsilon} s}/2$.*

Our result now follows from the inapproximability result for MINREP mentioned above.

5 Formulas with new variables

In this section we consider versions of the Horn minimization problem where one can introduce new variables in order to compress the formula.

5.1 General extensions

First we consider the general version where there is no restriction on the way new variables are introduced.

Definition 1 (Generalized equivalence). [17] *Let X be a set of variables. Formulas φ and ψ are X-equivalent if for every clause C involving only variables from X it holds that $\varphi \models C$ iff $\psi \models C$.*

Consider the set of variables $X = \{x_1, \ldots, x_n, y_1, \ldots, y_n, u\}$ and the 2^n-clause Horn formula $\varphi = \bigwedge(v_1, \ldots, v_n \to u)$ where $v_i \in \{x_i, y_i\}$ for $i = 1, \ldots, n$, and the conjunction includes all possible such selections. As no resolutions can be performed, it follows that all the prime implicates of φ are the clauses themselves. Let now $\{z_1, \ldots, z_n\}$ be new variables. Then the $(2n + 1)$-clause Horn formula $\psi = (z_1, \ldots, z_n \to u) \wedge \bigwedge_{i=1}^{n}(x_i \to z_i) \wedge (y_i \to z_i)$ is X-equivalent to φ. Thus the introduction of new variables can lead to an exponential compression in size.

For knowledge representation formalisms it is useful to have an efficient procedure to decide equivalence. Thus the following result of [17] suggests that general extensions of Horn formulas are too general for applications.

Theorem 4. [17] *Generalized equivalence of definite Horn formulas is* co-NP-complete.

5.2 Steiner extension

The proof of Theorem 4 shows that generalized equivalence is already hard if new variables are introduced in a rather restricted manner. This gives a motivation to consider even more stringent restrictions on the introduction of new variables.

Definition 2 (Steiner extension). *Let φ be a Horn formula and X be a subset of its variables. Then φ is a* Steiner extension *over X if every variable* not *in X occurs in φ either as a head, or as a single body variable in a binary definite clause having its head in X.*

The corresponding notion of equivalence is the following.

Definition 3 (Steiner equivalence). *Let X be a set of variables. Horn formulas φ and ψ are* Steiner *X-equivalent if*

- *φ and ψ are X-equivalent,*
- *both φ and ψ are Steiner extensions over X.*

The Horn formulas in (1) and (2) in Section 1.1 are both Steiner extensions over $X = \{x_1, \ldots, x_n, y_1, \ldots, y_n\}$, and they are Steiner X-equivalent. On the other hand, the example in Section 5.1 is not a Steiner extension as additional variables occur in the body of a non-binary clause. In contrast to Theorem 4, Steiner equivalence can be decided efficiently.

Proposition 1. *There is a polynomial algorithm which, given a Steiner extension φ over X, computes a Steiner X-equivalent Horn formula $\psi(X)$ containing only the variables in X such that $size(\psi) = O(size(\varphi)^2)$.*

The Steiner X-equivalence of φ_1 and φ_2 can be decided by using Proposition 1 to produce formulas $\psi_1(X)$ and $\psi_2(X)$ and checking their equivalence.

Corollary 1. *Steiner equivalence of Horn formulas is in P.*

The minimization problem for Steiner equivalence is the following.

Definition 4 (Steiner minimization of Horn formulas). *Given a Horn formula φ over variables X, find a minimal size Horn formula that is Steiner X-equivalent to φ.*

Using the correspondence between bipartite graphs and Horn formulas of binary clauses discussed earlier, Theorem 2 can be used to show the following.

Theorem 5. *Steiner minimization of definite Horn formulas is* MAX-SNP-hard.

We now show that Steiner minimization of definite Horn formulas has an efficient approximation algorithm with performance guarantee $o(n)$.

Remark 2. It may be assumed w.l.o.g. that Horn formulas to be minimized have no unit clause prime implicates. This holds as every prime representation can be partitioned into those unit clauses and a set of clauses not containing any variable that occurs in a unit clause prime implicate. The second half then can be minimized separately.

Theorem 6. *There is a polynomial time algorithm with approximation ratio*

$$O(n \log \log n/(\log n)^{1/4})$$

for Steiner minimization of definite Horn formulas, where n is the number of variables in the original formula.

The algorithm uses several procedures. It uses previous algorithms for listing prime implicates of Horn formulas and for body minimization. It also uses a procedure for the exact minimization of Horn formulas having a short equivalent formula and the bipartite graph partition algorithm of Section 3.

The *prime implicate listing problem* for Horn formulas is to produce a list of all prime implicates of a Horn formula. As the number of prime implicates can be exponential in the size of the original formula, a possible criterion of efficiency is *total polynomial time*, *i.e.*, time polynomial in the combined size of the input and the output. Boros, Crama and Hammer [9] give an algorithm which lists all prime implicates of a Horn formula, in time polynomial in the size of the formula and the number of prime implicates.

Consider the following special case of (standard) Horn minimization.

Problem 1 ($\sqrt{\log n}$-Horn minimization). Given a Horn formula φ over n variables, find an equivalent minimal size Horn formula of size at most $\sqrt{\log n}$ if such a formula exists, or output *'none'*.

Lemma 7. *The $\sqrt{\log n}$-Horn minimization problem is polynomial-time solvable.*

A further ingredient of the algorithm is an efficient procedure for body minimization. The *body minimization* problem for Horn formulas asks for an equivalent Horn formula with the minimal number of distinct bodies. While Horn minimization is hard, there *are* efficient algorithms for body minimization. Such algorithms were found in several different contexts, such as implicational systems [19] (see also [11]), functional dependencies for databases [32], directed hypergraphs [4] and computational learning theory[3].

Given a Horn formula χ over a set of variables X, we now describe a construction of a Steiner extension $\psi = \text{STEINER}(\chi)$ of χ. Let $Bodies(\chi)$ denote the set of bodies in χ, and $Heads(\chi)$ denote the set of heads in χ. Form a bipartite graph $G(\chi)$ with parts $Bodies(\chi)$ and $Heads(\chi)$, adding an edge between a body and a head if the corresponding Horn clause occurs in χ. Let $G_1, \ldots, G_t$ be a decomposition of $G(\chi)$ into bicliques obtained by the graph partition procedure of Theorem 1[8]. Let the bipartite graphs in the decomposition have parts

[8] The bipartite graphs need not be balanced. Also, for this application it would be sufficient to consider coverings instead of partitions. The result of Section 3 is formulated for balanced partitions in order to give a stronger positive result.

$A_i \subseteq Bodies(\chi)$ and $B_i \subseteq Heads(\chi)$ for $i = 1, \ldots, t$. Introduce new variables $y_1, \ldots, y_t$, and let $\psi = STEINER(\chi)$ consist of the clauses $b \to y_i$ and $y_i \to h$ for every $b \in A_i$ and $h \in B_i$, $i = 1, \ldots, t$.

In the following description of the algorithm let $\sqrt{\log n}$–HORN–MIN denote the procedure of Lemma 7 and let MIN-BODY be an efficient body minimization procedure.

Input: a definite Horn formula φ

Algorithm:

if $\psi = \sqrt{\log n}$–HORN–MIN(φ) $\neq$ 'none' **then** return ψ

else return STEINER(MIN-BODY(φ))

The performance bound of the algorithm follows by considering different cases depending on the value of $OPT(\varphi)$ and the relationship between the number of bodies and heads returned by the body minimization procedure.

6 Restricted Horn minimization

A special case of the Horn minimization problem is when only clauses from the original formula may be used in the new formula. Finding an irredundant subset of clauses representing the input function can always be done in polynomial time using the standard Horn formula procedures. However, there may be many irredundant formulas, having different sizes. The inapproximability result in Theorem 7 below shows that in fact it is hard to approximate the shortest one, *even if we assume that the formula to be minimized is* 3*-Horn.*

Theorem 7. *For any fixed* $0 < \epsilon < 1$*, unless NP*$\subseteq$*DTIME*$(n^{polylog(n)})$*, the restricted Horn minimization problem is* $2^{\log^{1-\epsilon} n}$*-inapproximable, even for definite* 3*-Horn formulas.*

As noted in the introduction, in the case of restricted Horn minimization one can also try to *maximize* the number of deleted clauses. We refer to this problem below as *Horn maximization.* In contrast to Theorem 7, for definite 2-Horn formulas constant approximation is possible.

Theorem 8.
(a) *Both the Horn minimization and Horn maximization problems are* MAX-SNP-hard *for definite* 2*-Horn formulas without unit clauses.*
(b) *Restricted Horn minimization for definite* 2*-Horn formulas admits a* 1.5*-approximation.*
(c) *Restricted Horn maximization for definite* 2*-Horn formulas admits a* 2*-approximation.*

In view of Remark 2, these results follow from [5, 42] and the correspondence between Horn formulas with binary clauses and directed graphs.

Remark 3. For **(a)**, the best inapproximability constants can be be obtained by using a randomized construction of a special class of Boolean satisfiability instances by Berman *et al.*[6] giving an inapproximability constant of $1+(1/896)$ for the minimization version and $1 + (1/539)$ for the maximization version.

References

1. R. Albert, B. DasGupta, R. Dondi and E. Sontag. *Inferring (biological) signal transduction networks via transitive reductions of directed graphs*, Algorithmica, 51 (2), 129-159, 2008.
2. N. Alon, R. A. Duke, H. Lefmann, V. Rödl, R. Yuster. *The algorithmic aspects of the regularity lemma*, J. of Algorithms, 16, 80-109, 1994.
3. D. Angluin, M. Frazier, L. Pitt. *Learning conjunctions of Horn clauses*, Machine Learning, 9, 147-164, 1992.
4. G. Ausiello, A. D'atri and D. Saccà. *Minimal representations of directed hypergraphs*, SIAM J. Comput., 15 (2), 418-431, 1986.
5. P. Berman, B. DasGupta and M. Karpinski. *Approximating transitive reduction problems for directed networks*, 11th Algorithms and Data Structures Symposium, F. Dehne, M. Gavrilova, J.-R. Sack and C. D. Tòth (eds.), LNCS 5664, 74-85, August, 2009.
6. P. Berman, M. Karpinski and A. D. Scott. *Approximation hardness of short symmetric instances of MAX-3SAT*, Electronic Colloquium on Computational Complexity, Report TR03-049, 2003.
7. E. Boros. *Horn functions*, in *Boolean Functions: Theory, Algorithms and Applications*, Y. Crama and P. L. Hammer (eds.), Cambridge Univ. Press, 2010 (forthcoming).
8. E. Boros and O. Čepek. *On the complexity of Horn minimization*, Rutcor Research Report 1-94, 1994.
9. E. Boros, Y. Crama, P. L. Hammer. *Polynomial-time inference of all valid implications for Horn and related formulae*, Ann. Math. Artif. Intell., 1, 21-32, 1990.
10. S. Bublitz. *Decomposition of graphs and monotone formula size of homogeneous fuctions*, Acta Informatica, 23, 689-696, 1996.
11. N. Caspard and B. Monjardet. *The lattice of closure systems, closure operators and implications on a finite set: a survey*, Discrete Appl. Math., 127, 241-269, 2003.
12. A. K. Chandra and G. Markowsky. *On the number of prime implicants*, Discrete Math., 24, 7-11, 1978.
13. F. R. K. Chung, P. Erdős and J. Spencer. *On the decomposition of graphs into complete bipartite graphs*, in *Studies in Pure Mathematics, To the Memory of Paul Turán*, 95-101. Akadémiai Kiadó, 1983.
14. P. Clote and E. Kranakis. *Boolean Functions and Models of Computation*, Springer, 2003.
15. S. A. Cook. *Feasibly constructive proofs and the propositional calculus*, STOC, 83-97, 1975.
16. U. Feige and S. Kogan. *Hardness of approximation of the balanced complete bipartite subgraph problem*, Tech. Rep. MCS04-04, Dept. of Comp. Sci. and Appl. Math., The Weizmann Inst. of Science, 2004.
17. A. Flögel, H. Kleine Büning and T. Lettmann. *On the restricted equivalence subclasses of propositional logic*, ITA, 27, 327-340, 1993.
18. M. D. Frazier. *Matters Horn and other features in the computational learning theory landscape: the notion of membership*, Ph. D. thesis, Univ. of Illinois at Urbana-Champaign, 1994.
19. J. L. Guigues and V. Duquenne. *Familles minimales d'implications informatives résultant d'un tableau de données binaires*, Math. Sci. Humaines, 95, 5-18, 1986.

20. P. L. Hammer and A. Kogan. *Horn functions and their DNFs*, Information Processing Letters, 44, 23-29, 1992.
21. P. L. Hammer and A. Kogan. *Optimal compression of propositional Horn knowledge bases: complexity and approximation*, Artificial Intelligence, 64, 131-145, 1993.
22. P. L. Hammer, A. Kogan: *Quasi-acyclic Horn knowledge bases: optimal compression*, IEEE Trans. on Knowledge and Data Engineering 7, 751-762, 1995.
23. F. K. Hwang, D. S. Richards and P. Winter. *The Steiner Tree Problem*, North-Holland, 1992.
24. S. Kirchner. *Lower bounds for Steiner tree algorithms and the construction of bicliques in dense graphs*, Ph.D. Dissertation, Humboldt-Universität zu Berlin, 2008. (In German.)
25. H. Kleine Büning and T. Lettmann. *Propositional Logic: Deduction and Algorithms*, Cambridge Univ. Press, 1999.
26. G. Kortsarz, R. Krauthgamer and J. R. Lee. *Hardness of approximating vertex-connectivity network design problems*, SIAM J. Comput., 33 (3), 704-720, 2004.
27. T. Kővári, V. T. Sós and P. Turán. *On a problem of K. Zarankiewicz*, Colloq. Math., 3, 50-57, 1954.
28. M. Langlois, R. H. Sloan and Gy. Turán. *Horn upper bounds and renaming*, J. Satisfiability, Boolean Modelling and Computation, 7, 1-15, 2009.
29. M. Langlois, R. H. Sloan, B. Szörényi and Gy. Turán. *Horn complements: towards Horn-to-Horn belief revision*, AAAI, 466-471, 2008.
30. M. Langlois, D. Mubayi, R. H. Sloan and Gy. Turán. *Combinatorial problems for Horn formulas*, in *Graph Theory, Computational Intelligence and Thought*, Springer LNCS 5420, 54-65, 2009.
31. L. Lovász. *Coverings and colorings of hypergraphs*, Proc. 4th Southeastern Conf. on Combinatorics, Graph Theory and Computing, Utilitas Math., 3-12, 1973.
32. D. Maier. *The Theory of Relational Databases*, Comp. Sci. Press, 1983.
33. D. Mubayi and Gy. Turán. *Finding bipartite subgraphs efficiently*, Inf. Proc. Lett. 110, 174-177, 2010.
34. J. Orlin. *Contentment in graph theory: covering graphs with cliques*, Indag. Math. 80, 406-424, 1977.
35. S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*, 2nd Ed. Prentice Hall, 2002.
36. P. Singh, T. Lin, E. T. Mueller, G. Lim, T. Perkins and W. L. Zhu. *Open Mind Common Sense: knowledge acquisition from the general public*, CoopIS/DOA/ODBASE, 1223-1237, 2002.
37. T. Tarján. *Complexity of lattice-configurations*, Studia Sci. Math. Hung., 10, 203-211, 1975.
38. G. S. Tseitin. *On the complexity of derivation in propositional calculus*, Seminars in Mathematics, V. A. Steklov Mathematical Institut, Leningrad, 8, 1968.
39. Zs. Tuza. *Covering of graphs by complete bipartite subgraphs; complexity of 0-1 matrices*, Combinatorica, 4, 111-116, 1984.
40. C. Umans. *Hardness of approximating Σ_2^p minimization problems*, FOCS, 465-474, 1999.
41. C. Umans. *The Minimum Equivalent DNF problem and shortest implicants*, J. Comput. Syst. Sci., 63, 597-611, 2001.
42. A. Vetta. *Approximating the minimum strongly connected subgraph via a matching lower bound*, 12th ACM-SIAM Symposium on Discrete Algorithms, 417-426, 2001.