Tractable Classes for Directional Resolution

Alvaro del Val
E.T.S. Informática
Universidad Autónoma de Madrid
delval@ii.uam.es
http://www.ii.uam.es/∼delval

Abstract

The original, resolution-based Davis-Putnam satisfiability algorithm (Davis & Putnam 1960) was recently revived by (Dechter & Rish 1994) under the name "directional resolution" (DR). We provide new positive complexity results for DR. First, we identify a class of theories (ACT, Acyclic Component Theories), which includes many real-world theories, for which DR takes polynomial time. Second, we present an improved analysis of the complexity of directional resolution through refined notions of induced width, which yields new tractable classes for DR, and much better predictions of its space and time requirements under various atom orderings. These estimates can be used for heuristically choosing among various orderings before running DR.

Introduction

The original, resolution-based Davis-Putnam satisfiability algorithm (Davis & Putnam 1960) was recently revived by (Dechter & Rish 1994; Rish & Dechter 2000) under the name "directional resolution" (DR). DR was shown to be a relatively efficient method for certain kinds of semi-structured problems, on which it often outperforms the backtracking-based Davis-Putnam algorithm (Davis, Logemann, & Loveland 1962) by orders of magnitude. While DR was also consistently outperformed by backjumping in these experiments, DR provides more information than backtracking procedures, as pointed out by (Dechter & Rish 1994), in that any model can be found backtrack-free after running DR.

In addition, DR lies at a very interesting theoretical crossroad. First, as shown in (Dechter 1999), DR is one of a family of "bucket elimination" (BE) procedures for logical, Bayesian, and constraint reasoning, as well as for linear and dynamic programming. The complexity of most of these BE procedures can be analysed in terms of a single structural parameter, induced width. Second, DR is, specifically, the bucket elimination version of propositional ordered resolution (in the sense of atom ordering, see e.g. (Fermüller *et al.* 1993; Bachmair & Ganzinger 1999)), itself a rich source of complexity results (Basin & Ganzinger 1996; Fermüller *et al.* 1993). Finally, (del Val 1999) has shown how to extend ordered resolution (with or without BE) to consequence-finding tasks, such as finding consequences that contain only certain literals, or with bounded length, or simply finding all consequences (the prime implicate task). DR provides a "bottom line" of computational effort for these consequence-finding tasks, as all the BE methods of (del Val 1999) perform at least as much work as DR. We extend our analysis of DR in this paper to consequence-finding in (del Val 2000).

(Dechter & Rish 1994) show that DR is tractable for a few classes of theories, including binary theories, and theories with bounded induced width or induced diversity. This paper makes two contributions. First, we introduce a wide class of structured, realistic theories for which DR takes polynomial time, even though it appears this cannot be predicted by induced width. Second, we refine the complexity analysis of DR in terms of induced width (Dechter & Rish 1994) by introducing new structural parameters to estimate the space and time requirements of DR. This allows us to identify new tractable classes, and also to obtain more accurate *empirical predictions* of complexity for any atom ordering. By comparing estimates for various orderings, we may heuristically choose among them before running DR. We will in fact go to some length to try to improve the accuracy of these predictions.

We assume familiarity with the standard terminology of propositional reasoning and resolution. The algorithm DR is very simple. Given a set of clauses Σ, fix some ordering $o = x_1, \ldots, x_n$ of the propositional variables. Associate to each variable a bucket $b[x_i]$ of clauses whose smallest variable, according to o, is x_i. Then process buckets in ascending order:[1]

Algorithm DR

for $i = 1$ to n do:
  compute all non-tautologous resolvents on x_i of clauses in $b[x_i]$, adding them to their corresponding buckets.

Example 1 *(Dechter & Rish 1994, example 2)* Consider the theory $\Sigma_1 = \{\overline{x_1}x_2, x_1\overline{x_3}, \overline{x_2}x_4, x_3x_4x_5\}$. Along the natural ordering, DR generates the clauses $x_2\overline{x_3}$, $\overline{x_3}x_4$, and x_4x_5. The buckets' final contents are $b[x_1] = \{\overline{x_1}x_2, x_1\overline{x_3}\}$, $b[x_2] = \{\overline{x_2}x_4, x_2\overline{x_3}\}$, $b[x_3] = \{x_3x_4x_5, \overline{x_3}x_4\}$, $b[x_4] = \{x_4x_5\}$, $b[x_5] = \emptyset$. □

DR is a complete SAT method. It is compatible with deletion of tautologies and subsumed clauses (del Val 1999),

[1] As in (del Val 1999), we use order of processing as our primary ordering, and speak of the variables processed first as the "earliest" or "smallest" in the ordering. (Dechter & Rish 1994; Rish & Dechter 2000) use as primary ordering reverse order of processing, speaking of processing buckets by resolving on "largest" literals. It is trivial to map from one representation to another.

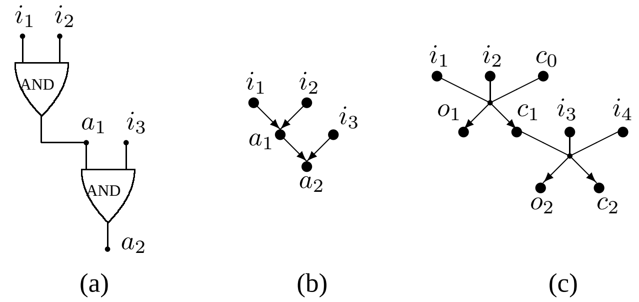

Figure 1: Sample ACTs.

though we will mostly ignore this for simplicity. In the above example, this would remove $x_3x_4x_5$.

For an ordering o, $DR_o(\Sigma)$ denotes the set of clauses obtained by processing Σ with DR along o. We use $b[x_i]^+$ (respectively $b[x_i]^-$) for the set of clauses of $b[x_i]$ in which x_i occurs positively (respectively, negatively). For any expression E, $Var(E)$ denotes its variables.

Tractable classes: ACNs

We next define acyclic component networks. They are in the spirit of other "structured system descriptions," e.g. (Darwiche & Pearl 1994; Dechter & Dechter 1996). Our building elements are *components*, whose associated "microtheories" describe their input/output behavior. This, together with an acyclic graphical structure (a DAG) representing flow from "inputs" to "outputs", ensures tractability results.

Formally, a component C_i is a tuple $C_i = \langle I_i, O_i, \Gamma_i \rangle$, where I_i and O_i are disjoint sets of variables (respectively, the "input" and "output" variables of C_i), and Γ_i is a set of clauses over $I_i \cup O_i$, the "component theory." The "component micrograph" $G_i = \langle V_i, E_i \rangle$ is defined by vertices $V_i = I_i \cup O_i$ and *directed* edges E_i, where $(x, y) \in E_i$ iff $x \in I_i$ and $y \in O_i$.

Components may be those of some device (e.g. logical gates in a circuit, engine parts), or represent events, or some other causal or logical relationship. We restrict how components can be linked together, which is what makes interesting structures appear:

Definition 1 *An Acyclic Component Network (ACN) is defined in terms of a set $COMP = \{C_1, \ldots, C_m\}$ of components. We require:*

1. *$O_i \cap O_j = \emptyset$, for any distinct i and j (i.e. a variable can be output variable of at most one component).*
2. *The component graph $G_C = \langle \bigcup_i V_i, \bigcup_i E_i \rangle$ is acyclic.*
3. *For any i and inverse topological order o of G_C, (a) $DR_o(\Gamma_i)$ can be computed in polynomial time, and (b) any clause $C \in DR_o(\Gamma_i)$ satisfies $Var(C) \cap O_i \neq \emptyset$.*

A theory Σ is an Acyclic Component Theory (ACT) just in case it can be partitioned into a set of disjoint subsets as $\Sigma = \bigcup_i \Gamma_i$ such that $Var(\Gamma_i)$ can in each case be partitioned into two sets I_i and O_i as above, and all other required properties are satisfied.

Example 2 Figure 1 displays a simple logical circuit (a) side by side with its component graph (b). The associated theory $\Sigma_2 = \Gamma_1 \cup \Gamma_2$ consists of microtheories $\Gamma_1 = \{\overline{i_1 i_2} a_1, \overline{a_1} i_1, \overline{a_1} i_2\}$, and $\Gamma_2 = \{\overline{a_1 i_3} a_2, \overline{a_2} a_1, \overline{a_2} i_3\}$.

Figure 1.c illustrates multiple outputs with a possible component graph for a 2-bit adder, composed of two full adders, with carries c, input bits i and output bits o. For convenience, all arcs for a single component are summarized as a single, star-shaped hyperedge.

Notice that one can easily "read" in the component graph an associated graph where nodes are components rather than propositional variables. $\Box$

Theorem 1 *Let Σ be an ACT. There is an order o such that $DR_o(\Sigma) = \bigcup_i DR_o(\Gamma_i)$, for which computing $DR_o(\Sigma)$ takes polynomial time.*

Proof sketch: Fix o to be any inverse topological sort of G_C. For any x_i, let $O(x_i)$ be the unique j, if any, such that $x_i \in O_j$. (Uniqueness follows from condition 1). We can show, by induction on the number k of variables processed that, for any x_i, if $O(x_i)$ is defined, then $b[x_i] \subseteq DR_o(\Gamma_{O(x_i)})$, else $b[x_i] = \emptyset$. In particular, we can show that every resolvent R generated by processing x_k is in $DR_o(\Gamma_{O(x_k)})$, and can be indexed only in a bucket $b[x_j]$ such that $x_j \in O_{O(x_k)} = O_{O(x_j)}$, thus preserving $b[x_j] \subseteq DR_o(\Gamma_{O(x_j)})$.

It now easily follows that $DR_o(\Sigma) = \bigcup_i DR_o(\Gamma_i)$, and since each component can be computed in polynomial time, the total cost of $DR_o(\Sigma)$ is also polynomial. $\Box$

Example 3 DR generates no non-tautologous resolvents on the theory Σ_2 of Example 2, using for example the order of processing a_2, a_1, i_1, i_2, i_3. $\Box$

Condition (3.b) implies that each Γ_i is satisfiable; thus by theorem 1, any ACT is satisfiable, as DR is complete for satisfiability. As in (Darwiche & Pearl 1994), consistency of components guarantees global consistency. What DR adds is the ability to generate any model of an ACT backtrack free (Dechter & Rish 1994) from $DR_o(\Sigma)$, using reverse order of processing —i.e. in topological order, with inputs assigned before outputs, if we follow Theorem 1.

ACNs capture a wide class of real-world theories, in particular those describing any combinatorial circuit. In this case the components would be the gates, and their simple microtheories, illustrated above, satisfy the given requirements. Condition 2 follows from the absence of feedback loops in combinatorial circuits. Condition 3 follows from the fact that these microtheories are (or can be easily put into) prime implicate form, so that (a) no new unsubsumed resolvents can be generated by any form of resolution on a gate's microtheory Γ_i, hence $DR_o(\Gamma_i) = \Gamma_i$ for any o; and (b) is satisfied initially by Γ_i, and by (a) this does not change. Thus applying Theorem 1 we obtain that in this case $DR_o(\Sigma) = \Sigma$ for any inverse topological sort o of G_C.

Theorem 1 is in fact a generalization of an *empirical observation*, the behavior of DR with ISCAS logic circuit benchmarks. It was not obvious at all that DR could do well on these benchmarks, until we tried them with an ordering compatible with this theorem. ACTs are however more general than theories of circuits, as in the latter inputs actually determine outputs; whereas Theorem 1 only suggests the weaker restriction that any assignment to a component's inputs can be *consistently extended* to its outputs; it need not *fix* them.

Other complexity results for structured descriptions, e.g. (El Fattah & Dechter 1995; Darwiche 1998) address tasks different from model-finding, e.g. abduction or diagnosis. It is interesting to note though that they come up with induced width analysis of circuits which fail to predict tractability except for tree-structured circuits. This suggests that the induced width analysis that we ourselves will advocate in the next section would fail to predict the tractability of DR on ACTs. For, as said, all circuits fit in the ACT framework.

Even though the tasks addressed are different, it is worth comparing the expressive power of ACTs and Symbolic Causal Networks (SCNs) (Darwiche & Pearl 1994; Darwiche 1998), since there are many similarities. It appears that every SCN is an ACN such that: (a) there is a single output per component; (b) "direct causes" are our inputs; (c) the SCN's "exogenous" or "assumable" propositions are treated as additional ACN outputs;[2] (d) outputs are determined by inputs;[3] (e) microtheories are required to have bounded size.

It can be seen, in particular from (a), (d) and (e), that SCNs are a relatively simple special case of ACNs.

Topological parameters for DR

We now turn to a more general analysis of the complexity of DR. We first introduce the concepts of induced width and diversity from (Dechter & Rish 1994). For application of these concepts to other areas of AI and CS, see (Dechter 1999; Bodlaender 1993). The intuitive idea is very simple. We capture the input theory Σ with a graph that represents cooccurrence of literals in clauses of Σ. We then "simulate" DR in polynomial time by processing the graph to generate a new "induced" graph. Finally, we recover from the induced graph information about all relevant complexity parameters for the hypothetical execution of DR for the given theory and ordering. Example 5 below will illustrate this idea of polynomial simulation of resolution.

As before, we use as our primary ordering the order of processing, as opposed to the inverse order used in (Dechter & Rish 1994). This means that *our induced width along an ordering corresponds to Dechter and Rish's induced width along the reverse order.*[4] The following definitions, though somewhat dense, should not be hard to parse for readers familiar with the notion of induced width.

Definition 2 *Let G be an undirected graph, and $o = v_1, \ldots, v_n$ an ordering of its vertices.*

The downward set $D(v_i)$ of v_i along o *is the set of vertices v_j such that there is an edge (v_i, v_j) in G and $i < j$. The* (downward) width of v_i along o *is the cardinality $|D(v_i)|$ of its downward set.*

The (downward) width of a graph G along o *is the maximum downward width among the nodes of G.*

Definition 3 *(Dechter & Rish 1994, induced width) Let Σ be a clausal theory, and $o = x_1, \ldots, x_n$ an ordering of its propositional variables $Var(\Sigma)$.*

1. The interaction graph *of Σ is an undirected graph $GI(\Sigma) = \langle V_I, E_I \rangle$, with vertices $V_I = Var(\Sigma)$, and edges $E_I = \{(x_i, x_j) \mid x_i, x_j \in Var(C)$ for some $C \in \Sigma\}$.*

[2]They are "private" to each component in (Darwiche & Pearl 1994), in line with our weaker restriction $O_i \cap O_j = \emptyset$.

[3]This follows from the syntactic restrictions on SCN microtheories imposed in (Darwiche & Pearl 1994).

[4]See also footnote 1. After all, the induced graphs below are also generated in DR's order of processing.

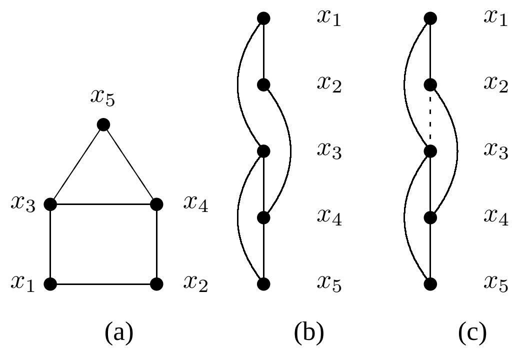

Figure 2: The interaction graph of Σ_1.

2. The induced interaction graph of Σ along ordering o *is the graph $I_o(GI(\Sigma)) = \langle V_I, E_I^o \rangle$ obtained from $GI(\Sigma)$ as follows: initially, $E_I^o = E_I$; then for $i = 1$ to n do: add edges (x_j, x_k) to E_I^o for all x_j and x_k in the current $D(x_i)$.*

3. The width $w_o(x_i)$ of a variable x_i along o *is the width of x_i in $GI(\Sigma)$ along o. The* induced width $w_o^*(x_i)$ of x_i along o *is the width of x_i in $I_o(GI(\Sigma))$ along o. When the ordering o is fixed, we abbreviate $w_o(x_i)$ as w_i, and $w_o^*(x_i)$ as w_i*.*

4. The width $w(o)$ of Σ along o *is the width of $GI(\Sigma)$ along o, and the* width w of Σ *is the minimal width of Σ over all orderings. The* induced width $w*(o)$ of Σ along o *is the width of $I_o(GI(\Sigma))$ along o, and the* induced width $w*$ *of Σ is the minimal induced width over all possible orderings.*

Example 4 Figure 2 illustrates induced width for the theory Σ_1 of Example 1: (a) the interaction graph of Σ_1; (b) the same graph, with nodes ordered from top to bottom along the given (natural) order; (c) the induced interaction graph along this ordering. The width and induced width of Σ_1 with this ordering are both 2. □

Definition 4 *(Dechter & Rish 1994, diversity) The* induced diversity of a variable x_i along order o *is $div_o^*(x_i) = |b[x_i]^+| \times |b[x_i]^-|$, where the product is taken after running $DR_o(\Sigma)$. The* induced diversity $div*(o)$ of Σ along o *is the maximum induced diversity of its variables along o. The* induced diversity $div*$ of Σ *is its minimal induced diversity over all orderings.*

Clearly, the induced diversity of an ordering provides a bound on the number of resolution steps performed by DR. Since each resolution step is $O(n)$, the time complexity of DR along an ordering o is $O(n^2 \cdot div*(o))$.

(Dechter & Rish 1994) show that the size of $DR_o(\Sigma)$ is bounded by $n \cdot 2 \cdot 3^{w*(o)}$, and the number of resolution steps by $n \cdot (2 \cdot 3^{w*(o)})^2 = n \cdot 4 \cdot 3^{2w*(o)}$.

We can greatly improve these estimates by introducing some new concepts. The basic idea is to split literals according to their sign, and use this to define more fine-grained concepts of width and to estimate diversity.

Definition 5 (split interaction graph) *Let Σ be a clausal theory, and $o = x_1, \ldots, x_n$ an ordering of $Var(\Sigma)$, extended in the obvious way to literals (e.g. $\overline{x_i} < x_j$ iff $i < j$).*[5]

[5]Note that the extended ordering is a partial order.

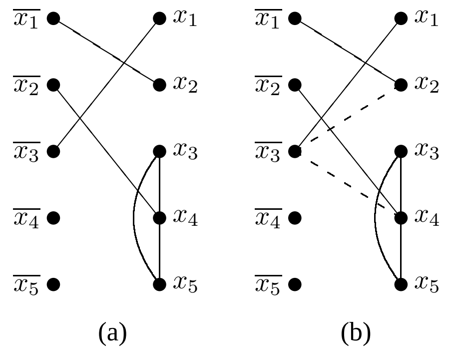

Figure 3: The split interaction graph of Σ_1.

1. The split interaction graph *of Σ is an undirected graph $GS(\Sigma) = \langle V_S, E_S \rangle$, whose vertices V_S are the set of all literals of Σ, and E_S is the set of edges (l_i, l_j) such that literals l_i and l_j occur together in some clause.*

2. The induced split interaction graph of Σ along ordering o *is the graph $I_o(GS(\Sigma)) = \langle V_S, E_S^o \rangle$ obtained by augmenting $GS(\Sigma)$ as follows: initially, $E_S^o = E_S$; then for $i = 1$ to n do: add edges (l_j, l_k) to E_S^o for each $l_j \in D(x_i)$ and $l_k \in D(\overline{x_i})$.*

3. The literal width of a literal l_i along o *is the width of l_i in $GS(\Sigma)$ along o. The* induced literal width of l_i in Σ along o *is the width of l_i in $I_o(GS(\Sigma))$ along o. When o is fixed, we write w_i^+ and w_i^- for the literal widths along o of x_i and $\overline{x_i}$, respectively; similarly, we use w_i^+* and w_i^-* for the respective induced literal widths.*

4. The s-width of a variable x_i along o *is $sw_i = max(w_i^+, w_i^-)$. The* t-width of x_i along o *is $tw_i = w_i^+ + w_i^-$. The* induced s-width sw_i* *and* induced t-width tw_i* *are defined similarly from w_i^+* and w_i^-*.*

5. The s-width $sw(o)$ of Σ along o *is the maximal s-width of its variables along o. The* t-width $tw(o)$ of Σ along o *is the maximal t-width of its variables along o.*

6. The induced s-width $sw*(o)$ of Σ along o *is the maximal induced s-width of its variables along o. The* induced t-width $tw*(o)$ of Σ along o *is the maximal induced t-width of variables along o. The* induced s-width $sw*$ *and* induced t-width $tw*$ *of Σ are respectively the minimal induced s-width and t-width over all orderings.*

Induced width is basically an instrument to predict that certain literals will occur together in certain clauses; induced s-width and t-width are more fine-grained instruments that take into account the sign of literals to predict, respectively, the space (s) and time (t) requirements of DR. Note that both induced graphs can be computed in polynomial time.

Example 5 Figure 3 illustrates induced s-width and t-width for the theory Σ_1 of Example 1: (a) the split interaction graph of Σ_1, with nodes ordered from top to bottom along the given ordering; (b) the induced interaction graph along this ordering. Σ_1 has s-width, t-width, and induced s-width 2, and induced t-width 3.

Note that the generation of the induced graph closely matches the generation of resolvents by DR. Since x_1 cooccurs with $\overline{x_3}$ and $\overline{x_1}$ with x_2 we can "predict" that as a result of resolving upon x_1 we will obtain a resolvent in which x_2 and $\overline{x_3}$ occur together. This corresponds to adding the edge $(x_2, \overline{x_3})$ to the induced graph; and similarly with other added edges. □

All these definitions, incidentally, can be adjusted to apply to Σ under any fixed set of polynomial time simplifications, along the lines of the "adjusted width" of (Dechter 1999). Our next lemma shows that the induced graph simulates DR as far as the split interaction graph is concerned.

Lemma 2 *$GS(DR_o(\Sigma))$ is a subgraph of $I_o(GS(\Sigma))$.*

Proof: We show by induction along the ordering that if $l_i, l_j \in C$ for some clause C obtained when DR processes bucket $b[x_k]$ then the edge (l_i, l_j) is in $I_o(GS(\Sigma))$ after processing x_k when generating the induced graph. (Assume $l_i \in \{x_i, \overline{x_i}\}$ and $l_j \in \{x_j, \overline{x_j}\}$.)

For the base case, edges corresponding to clauses of Σ are obviously in both graphs. Inductively, consider any edge (l_i, l_j) such that $l_i, l_j \in R$ for some resolvent R of clauses $C, D \in b[x_k]$. Note that $k < i$, and $k < j$. If l_i and l_j occur together in some parent then the claim follows directly from the inductive hypothesis, as both parents were generated before DR processes x_k. Otherwise, say $x_k, l_i \in C$, and $\overline{x_k}, l_j \in D$. By inductive hypothesis, the edges (x_k, l_i) and $(\overline{x_k}, l_j)$ were added to $I_o(GS(\Sigma))$ before processing x_k in the generation of the induced graph. Since $k < i$ and $k < j$, the edge (l_i, l_j) is added to $I_o(GS(\Sigma))$ when processing x_k. □

We can now bound the space and time complexity of DR.

Lemma 3 *$|b[x_i]^+| \le 2^{w_i^+}$, and $|b[x_i]^-| \le 2^{w_i^-}$.*

Proof: $b[x_i]^+$ contains all clauses whose smallest literal is x_i. There are w_i^+ literals that cooccur with the positive literal x_i which are later in the ordering than x_i, hence there are $\sum_{0 \le i \le w_i^+} \binom{w_i^+}{i} = 2^{w_i^+}$ subsets of those w_i^+ literals. Only clauses formed by adding x_i to one such subset can be in $b[x_i]^+$. □

Theorem 4 *For any ordering o, the size of $DR_o(\Sigma)$ is bounded by $\sum_{1 \le i \le n} (2^{w_i^+*} + 2^{w_i^-*}) \le n \cdot 2^{sw*(o)+1}$. The number of resolution operations performed by DR is bounded by $\sum_{1 \le i \le n} 2^{tw_i*} \le n \cdot 2^{tw*(o)}$, hence the time complexity is $O(n^2 \cdot 2^{tw*(o)})$.*

Proof: Use lemmas 2 and 3 to bound bucket sizes in $DR_o(\Sigma)$. From this obtain $div_o^*(x_i) \le 2^{w_i^+*} \cdot 2^{w_i^-*} = 2^{tw_i*}$. As each resolution step is $O(n)$, time follows. □

This yields our second tractable class:

Corollary 5 *If the induced s-width or t-width of Σ is bounded by a constant then for some ordering o computing $DR_o(\Sigma)$ takes polynomial time and space.*

Finding the minimal induced s- or t-width of a theory is NP-hard,[6] though we conjecture that techniques to recognize in time $exp(k)$ theories with induced width bounded by k can be adapted to recognize theories with bounded induced s-width. And we can always use the polynomially computable bounds for any given ordering to choose among candidate orderings.

(Dechter & Rish 1994) prove a result similar to Corollary 5 for bounded *standard* induced width. But it is easy to find theories with constant induced s-width yet unbounded (i.e. $\Omega(n)$) standard induced width:

[6] As proven with a trivial reduction from the problem of finding the minimal induced width.

Example 6 Let $\Sigma_6 = \{x_i x_{2i}, x_i x_{2i+1} \mid 1 \le i \le m\}$. Along the natural order, each x_i gets linked to x_{i+1} through x_{2i+1} in the *standard* induced graph, whereas no new edges are added in the *split* induced graph, since negative literals have no edges. Thus $sw*(o) = 2$ is constant, while the standard induced width is the unbounded $w*(o) = w_m* = m + 1$.

For a non-binary theory with identical induced graphs, consider $\Sigma_6^* = \{x_i x_{2i} x_{2i+1} \mid 1 \le i \le m\}$. There is little point however in providing non-binary examples to exhibit properties of $sw*$ and $w*$, even though binary theories are tractable. This is because, for any non-binary theory Σ, there exists a binary theory Σ_B with exactly identical initial and induced graphs. □

Example 7 Let us compare the estimates derivable from induced width, t-width, and s-width with the theory Σ_1 of Example 1, using figures 2 and 3.

We can obtain two estimates from induced width, using the results of (Dechter & Rish 1994). The looser prediction yields a size estimate of $n \cdot 2 \cdot 3^{w*(o)} = 5 \cdot 2 \cdot 3^2 = 90$ clauses for $DR_o(\Sigma_1)$, or more precisely (summing over the induced widths of each variable) $3 \cdot (2 \cdot 3^2) + 2 \cdot 3^1 + 2 \cdot 3^0 = 62$ clauses. The estimated number of resolutions is, loosely, $n \cdot (2 \cdot 3^{w*(o)})^2 = 5 \cdot (2 \cdot 3^2)^2 = 1620$; and more precisely, $3 \cdot (2 \cdot 3^2)^2 + (2 \cdot 3^1)^2 + (2 \cdot 3^0)^2 = 1012$.

The estimates derived from s-width and t-width are significantly better. For size, these are $n \cdot 2^{sw*(o)+1} = 5 \cdot 2^3 = 40$ or more precisely $(2^1 + 2^1) + (2^1 + 2^1) + (2^1 + 2^2) + (2^0 + 2^1) + (2^0 + 2^0) = 19$ clauses. For time, the loose estimate is that there are $n \cdot 2^{tw*(o)} = 5 \cdot 2^3 = 40$ resolution steps; the precise estimate, doing the summation, is that there are only 19 steps. □

More generally, we can compare the *rough* estimates provided by these parameters as follows:

Theorem 6 *If (l_j, l_k) is an edge of $I_o(GS(\Sigma))$ then (x_j, x_k) is an edge of $I_o(GI(\Sigma))$. Hence $sw_i* \le w_i*$ for any i and fixed ordering.*

It easily follows that the size estimate derived from induced width is at least $(3/2)^{w*(o)}$ times larger than the estimate derived from induced s-width, and the time estimate at least $4 \cdot (3/2)^{2w*(o)}$ times larger. Both ratios hold when $sw_i* = w_i*$, but as Example 6 illustrates, often sw_i* is much smaller. And, in fact, we can greatly improve even the summation form of our estimates.

Relative width

A clause C is captured by the interaction graph as a *clique* of all its literals, i.e. all literals of C are pairwise linked. The size bound $2^{w_i^+}$ of Lemma 3 can be read as an estimate of the number of cliques in which x_i is the smallest literal, an estimate which is tight only if its downward set $D(x_i)$ is itself a clique. Furthermore, it follows from the proof of Lemma 2 that the clique corresponding to a resolvent obtained when DR processes $b[x_k]$ is added to $I_o(GS(\Sigma))$ when processing x_k as well. Thus, resolvents whose smallest literal is x_i become cliques *before* processing x_i, since they are generated by processing earlier buckets.[7]

[7]This is important because processing x_i may link together nodes of $D(x_i)$. Indeed, in the *standard* induced graph processing x_i means making $D(x_i)$ a clique.

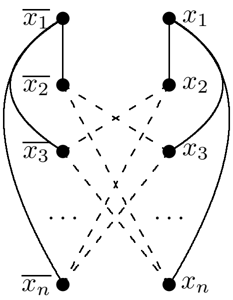

Figure 4: Induced split interaction graph for Σ_8.

Counting cliques over $D(x_i)$ is NP-hard, but there are a number of ways to approximate them, including polynomial randomized algorithms. The following simple upper bound suffices to obtain new tractable classes:

Lemma 7 *Let $D_i(l)$ be the downward set of l in $I_o(GS(\Sigma))$ right before processing x_i when generating the induced graph. Let $d_i^+(l) = |D_i(x_i) \cap D_i(l)|$, $d_i^-(l) = |D_i(\overline{x_i}) \cap D_i(l)|$. After running DR, $|b[x_i]^+| \le 1 + \sum_{l \in D(x_i)} 2^{d_i^+(l)}$, and $|b[x_i]^-| \le 1 + \sum_{l \in D(\overline{x_i})} 2^{d_i^-(l)}$.*

We can see $d_i^+(l)$ as the *relative width* of l with respect to x_i. Intuitively, each term in the sum estimates the number of cliques (whose smallest literal is x_i) containing l but no earlier literals; the estimate is tight only when $D(x_i) \cap D_i(l)$ is itself a clique right before processing x_i. Our next result is now straightforward:

Theorem 8

1. *Space: The size of $DR_o(\Sigma)$ is bounded by $2n + \sum_{1 \le i \le n} \left(\sum_{l \in D(x_i)} 2^{d_i^+(l)} + \sum_{l \in D(\overline{x_i})} 2^{d_i^-(l)} \right)$.*
2. *Time: The number of resolution steps performed by DR along ordering o is bounded by $\sum_{1 \le i \le n} \left[\left(1 + \sum_{l \in D(x_i)} 2^{d_i^+(l)}\right)\left(1 + \sum_{l \in D(\overline{x_i})} 2^{d_i^-(l)}\right) \right]$.*

As a special case of this theorem, suppose that for every literal l_i, the restriction of $I_o(GS(\Sigma))$ to its downward set $D(l_i)$ at the time x_i is processed contains no edges. (This is compatible with unbounded s-width, see Example 8.) Then the size of each $b[x_i]$ is only $2 + w_i^+ * + w_i^- *$, and its diversity $(1 + w_i^+ *)(1 + w_i^- *)$. And now the exponents are out! Generalizing this observation yields a new tractable class:

Corollary 9 *Suppose the $d_i(l)$'s are bounded by a constant for some ordering o. Then $|DR_o(\Sigma)| = O(n \cdot sw*(o)) = O(n^2)$. The number of resolution operations is $O(n \cdot (sw*(o))^2) = O(n^3)$, and thus the time complexity is $O(n^4)$.*

Proof: We have $|b[x_i]^+| \le 1 + \sum_{l \in D(x_i)} k$, for some constant k, hence $|b[x_i]^+| = O(w_i^+ *) = O(sw*(o)) = O(n)$. The rest is obvious. □

Example 8 $\Sigma_8 = \{x_1x_2, \ldots, x_1x_n, \overline{x_1x_2}, \ldots, \overline{x_1x_n}\}$ illustrates the principle of bounded $d_i(l)$'s with unbounded s-width. The induced split interaction graph $I_o(GS(\Sigma_8))$ is depicted in Figure 4, with induced edges dashed. Despite the high s-width of the graph ($w_i^+* = w_i^-* = n - i$ for each i), the $d_i(l)$'s always equal 0. □

We can also use relative widths to improve the estimates derived from the standard induced graph $I_o(GI(.))$. However, this would not allow us to derive tractability results such as corollary 9. This is because bounded $d_i(l)$'s in $I_o(GI(.))$ imply bounded induced width.[8] In other words, the use of the *split* induced graph is essential in obtaining this new tractable class.

The effect of subsumption

A clique of size k over $D(x_i)$ has 2^k subcliques, and all our estimates of bucket size so far take all of them as representing legitimate clauses. However, it is a well-known result in combinatorics, known as Sperner's theorem (Anderson 1987), that the maximum number of *unsubsumed* subsets of a set of k literals is $C(k, k/2) = \binom{k}{k/2}$. If we delete subsumed clauses, therefore, any term of the form 2^k in our previous estimates can be replaced by $C(k, k/2)$. In the theory Σ_1 of Example 1 this improves our size estimate to 11 clauses, and the number of resolution steps to 6.

This is unlikely to yield new tractable classes. It does allow us to obtain better predictions of complexity, though at a significant computational cost. This cost can be easily reduced with some loss of accuracy, e.g. by building a table of upper bounds in terms of powers of two (e.g. for $9 \leq k \leq 40$, $C(k, k/2) \leq 2^{k-2}$), so that all calculations are by powers of two. Even so, if our goal is only to compare various orderings in order to choose which one to use with DR, it is unclear whether this greater accuracy will help discriminate better among them.

Discussion

We have introduced new tractable classes for DR, and tightened the space and time bounds provided by (Dechter & Rish 1994) by means of a more refined analysis of the structure of theories. These bounds can be used to choose in polynomial time among different orderings before running DR.

As mentioned, the generation of the induced graph can be seen as a "polynomial time simulation" of resolution, which need not be limited to DR. In (del Val 2000), we extend this kind of analysis to the consequence-finding task, which for example allows us to estimate the number of prime implicates of any theory and identify tractable abduction classes.

References

Anderson, I. 1987. *Combinatorics of Finite Sets*. Oxford: Oxford University Press.

Bachmair, L., and Ganzinger, H. 1999. A theory of resolution. In Robinson, J., and Voronkov, A., eds., *Handbook of Automated Reasoning*. Elsevier.

[8]Suppose $w_i* = m = \Omega(n)$, and let x_j and x_k be the two smallest variables of $D(x_i)$. Processing x_i links together all edges of $D(x_i)$, hence $|D(x_j) \cap D_j(x_k)| \geq m - 2 = \Omega(n)$.

Basin, D., and Ganzinger, H. 1996. Complexity analysis based on ordered resolution. In *LCS'96, Proc. 11th IEEE Symposium on Logic in Computer Science*.

Bodlaender, H. L. 1993. A tourist guide through treewidth. *Acta Cybernetica* 11:1–21.

Darwiche, A., and Pearl, J. 1994. Symbolic causal networks. In *AAAI'94, Proc. 12th National Conference on Artificial Intelligence*, 238–244. Menlo Park, CA, USA: AAAI Press.

Darwiche, A. 1998. Model-based diagnosis using structured system descriptions. *J. of Artificial Intelligence Research* 8:165–222.

Davis, M., and Putnam, H. 1960. A computing procedure for quantification theory. *Journal of the ACM* 7(3):201–215.

Davis, M.; Logemann, G.; and Loveland, D. 1962. A machine program for theorem proving. *Communications of the ACM* 5:394–397.

Dechter, R., and Dechter, A. 1996. Structure driven algorithms for truth maintenance. *Artificial Intelligence* 82:1–20.

Dechter, R., and Rish, I. 1994. Directional resolution: The Davis-Putnam procedure, revisited. In *KR'94, Proc. 4th Int. Conf. on Principles of Knowledge Representation and Reasoning*, 134–145. Morgan Kaufmann.

Dechter, R. 1999. Bucket elimination: A unifying framework for reasoning. *Artificial Intelligence* 113:41–85.

del Val, A. 1999. A new method for consequence finding and compilation in restricted languages. In *AAAI'99, Proc. 16th (U.S.) National Conference on Artificial Intelligence*, 259–264. AAAI Press/MIT Press.

del Val, A. 2000. The complexity of restricted consequence finding and abduction. In *AAAI'2000, Proc. 17th (U.S.) National Conference on Artificial Intelligence*. AAAI Press/MIT Press.

El Fattah, Y., and Dechter, R. 1995. Diagnosing tree-decomposable circuits. In *IJCAI'95, International Joint Conference on Artificial Intelligence*, 1742–1748.

Fermüller, C.; Leitsch, A.; Tammet, T.; and Zamov, N. 1993. *Resolution Methods for the Decision Problem*. Springer-Verlag.

Rish, I., and Dechter, R. 2000. Resolution versus search: Two strategies for SAT. *Journal of Automated Reasoning* 24:225–275.