Transforming quantified boolean formulas using biclique covers

Oliver Kullmann Ankit Shukla

Computer Science Department
Swansea University

The Art of SAT (Shonan 2023)
October 4, 2023

A (very) simple example

A DQCNF:

$$F := \forall x_1, x_2, x_3, x_4 \, \exists y_1(x_1, x_2, x_3) \, \exists y_2(x_1, x_2, x_3, x_4) :$$
$$(y_1 \vee x_2 \vee x_3) \wedge (\neg y_1 \vee x_1 \vee \neg x_2) \wedge (\neg y_2 \vee \neg x_1 \vee \neg x_2 \vee \neg x_3 \vee x_4) \wedge (y_2 \vee \neg x_4)$$

A **solution**

- is a substitution of boolean functions into the y_i
- using only the universal variables given by the dependencies,
- such that one obtains a tautology (in the universal variables).

For example

$$y_1 = \neg x_2, \quad y_2 = x_4$$

is a solution.

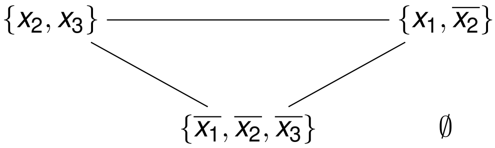
The global conflict graph of the example

A **global variable** is a universal variable
such that all existential variables depend on it.

- The global variables of F are x_1, x_2, x_3.
- The sub-clauses given by the global variables yield the

  global slice

  of F.
- The **conflict graph** of the global slice is the **global conflict graph**:

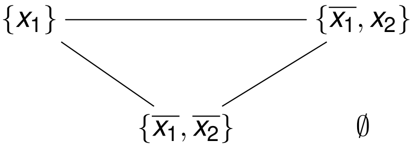


For the global slice, only the conflict graph matters

The main observation:

The global slice can be replaced SAT-equivalently
by anything with the same conflict graph.

The triangle can be realised with just two variables x_1, x_2:



This triangle-realisation is Horn, minimally unsatisfiable, with one clause more than variables (and indeed this all is always possible).

The replacement result

$$F = \forall \boxed{x_1, x_2, x_3,}\, x_4\, \exists y_1(x_1, x_2, x_3)\, \exists y_2(x_1, x_2, x_3, x_4) :$$

$$(y_1 \vee \underline{x_2 \vee x_3}) \wedge (\neg y_1 \vee \underline{x_1 \vee \neg x_2}) \wedge (\neg y_2 \vee \underline{\neg x_1 \vee \neg x_2 \vee \neg x_3} \vee x_4) \wedge (y_2 \vee \neg x_4)$$

$$F' = \forall \boxed{x_1, x_2,}\, x_4\, \exists y_1(x_1, x_2)\, \exists y_2(x_1, x_2, x_4) :$$

$$(y_1 \vee \underline{x_1}) \wedge (\neg y_1 \vee \underline{\neg x_1 \vee x_2}) \wedge (\neg y_2 \vee \underline{\neg x_1 \vee \neg x_2} \vee x_4) \wedge (y_2 \vee \neg x_4).$$

The solutions change by this process; a new solution is

$$y_1 = \neg x_1, \quad y_2 = x_4.$$

Investigating structure

The general problem is to

understand the structure
of (classes of) problem instances.

We investigate the **global slice** and its manipulations.

- The global variables are the most accessible part of a DQCNF.
- Their "meaning" is only in their conflict patterns.

Definition of DQCNF

Formally a DQCNF is a 4-tuple (A, E, F, D) where

1. A is the set of universal variables
2. E is the set of existential variables ($A \cap E = \emptyset$)
3. F is a clause-set over $A \cup E$
4. $D : E \to \mathbb{P}(A)$ gives the dependencies of existential variables.

A satisfying assignment is a map Φ, which maps an existential variable v to a boolean function $\Phi(v)$ over $D(v)$, such that substitution into F yields a tautology (over A).

- CNFs are the cases with $A = \emptyset$.
- QCNFs are the cases where the set $\{D(v) : v \in E\}$ of dependency-sets is linearly ordered (by subsumption).

Global variables

A **global variable** of (A, E, F, D) is a universal variable $x \in A$ such that

$$\forall y \in E : x \in D(y).$$

- A QCNF has global variables iff the outermost quantifier block is universal,
- in which case the global variables are the variables of this block.
- So for a 2QCNF (quantifier-prefix $\forall\exists$) all universal variables are global variables.

Global expansion

Consider a DQCNF F and a universal variable x.

1. If F is satisfiable then $\langle x \to 0 \rangle * F$ and $\langle x \to 1 \rangle * F$ are satisfiable.
2. Here for a partial assignment φ of boolean values to universal variables, by $\varphi * F$ we denote the result of applying φ to the clause-set F and removing the variables of φ.
3. However the reverse direction is not true in general!
4. If for an existential variable y with $x \notin D(y)$ we get boolean functions Φ_0, Φ_1 in the solutions for $\langle x \to 0/1 \rangle * F$, and we have $\Phi_0 \neq \Phi_1$, then we can't lift the solution to F.

Exactly for the global variables we never have this problem.

The conflict graph of labelled clause-sets

A labelled clause-set is a pair (L, F), where F is a map which maps a label $l \in L$ to a clause $F(l)$.

- The **conflict graph** has vertex set L.
- And there is an edge between vertices l_1, l_2 iff there is a literal $x \in F(l_1)$ with $\overline{x} \in F(l_2)$.

Note that we have "simple graphs":

1. no (self)-loops (since no tautological clauses)
2. no parallel edges (since we ignore them!).

The global conflict graph

For a DQCNF $F = (A, E, F, D)$, the **global slice**

- is the labelled clause-set with label-set F,
- mapping every $C \in F$ to the sub-clause given by the global variables in C.

The **global conflict graph** of F is the conflict graph of the global slice of F.

Independent sets of the global conflict graph

Theorem 3.1

A DQCNF F is unsatisfiable iff there is a maximal independent subset F' of the global conflict graph of F, which as a sub-DQCNF of F is unsatisfiable.

Note that in F' all global variables of F are pure (occur only in one sign), and thus can be eliminated.

Corollary 3.2

Let k be the number of global variables of F. Then 2^k is an upper bound on the number of maximal independent subsets of the global conflict graph of F.

Replacement of the global slice

So if two DQCNF's F, F' are

- the same up to the global slice, and
- have the same maximal independent subsets of their global conflict graphs,

then they are SAT-equivalent.

A sufficient condition for this is:

Corollary 3.3

Two DQCNF's F, F', which are the same up to the global slice, and have the same global conflict graph, are SAT-equivalent.

We call F, F' **gcg-equivalent** if this criterion applies.

Good new global slices?!

So we are now free to choose for a given DQCNF F any gcg-equivalent F':

- Which are better for solving?
- First guess: minimise the number of variables.
- The non-constructive proof of "gcg-equivalence $\Rightarrow$ SAT-equivalence" suggests that such changes can drastically alter complexity.

GSP and GSM

Global Slice Preprocessing could be a useful tool:

- **Global Slice Minimisation** (GSM), that is, replace the global slice by another one with a minimum number of variables, is the most natural first approach.
- GSM is naturally equivalent to the **Biclique Cover Problem**, the well-known NP-complete problem of covering a graph with as few bicliques as possible.
- In this first study we concentrated on general results.

Horn MUs

The connected components of the global conflict graph are to be processed independent, and so we can assume w.l.o.g. that the global conflict graph is connected:

Theorem 4.1

*Every connected global slice with m clauses can be replaced by another one, called **Horn-MU**, with the following properties:*

1. *it is minimally unsatisfiable;*
2. *it is Horn;*
3. *it has $m - 1$ variables*
4. *there is never more than one conflict between clauses.*

An example for an optimal realisation

Assume that the global slice has 256 clauses, and the conflict graph is complete — what is an optimum replacement?

An example for an optimal realisation

Assume that the global slice has 256 clauses, and the conflict graph is complete — what is an optimum replacement?

The canonical normal form with 8 variables
and 2^8 clauses of length 8.

1. Here it is important that we only **cover** the conflict graph, that is, we can have more conflicts (between 1 and 8 here).
2. If we would need to **partition** a complete graph with m vertices, then by the Graham-Pollak Theorem the previous upper bound of $256 - 1 = 255$ variables is optimal.

First experiments

We created random 2QCNFs, where the components of the global conflict graphs were complete graphs, together with three realisations:

1. the **trivial realisation**, with one variable per edge;
2. the **Horn-MU realisation**;
3. the **logarithmic realisation.**

For these benchmarks, “smaller is better” seemed (mostly) true.

Future research

1. The effects of GSP on proof-complexity and certificate-extraction needs to be explored.
2. We are currently investigating how to perform GSM efficiently, and what it yields.
3. Can the results be generalised to more general types of variables?!?

Future research

1. The effects of GSP on proof-complexity and certificate-extraction needs to be explored.
2. We are currently investigating how to perform GSM efficiently, and what it yields.
3. Can the results be generalised to more general types of variables?!?

End