

# Partial Regularization of First-Order Resolution Proofs (Experimental/Tool Paper)

Jan Gorzny<sup>1</sup> \* and Bruno Woltzenlogel Paleo<sup>2,3</sup>

<sup>1</sup> jgorzny@uvic.ca, University of Victoria, Canada

<sup>2</sup> bruno@logic.at, Vienna University of Technology, Austria

<sup>3</sup> Australian National University

**Abstract.** This paper describes the generalization of the proof compression algorithm `RecyclePivotsWithIntersection` from propositional to first-order logic. The generalized algorithm performs partial regularization of resolution proofs containing resolution and factoring inferences with *unification*, as generated by many automated theorem provers. An empirical evaluation of the generalized algorithm and its combinations with `GreedyLinearFirstOrderLowerUnits` is also presented.

## 1 Introduction

First-order automated theorem provers, commonly based on resolution and superposition calculi, have recently achieved a high degree of maturity. Proof production is a key feature that has been gaining importance, since proofs are crucial for applications that require certification of a prover’s answers or information extractable from proofs (e.g. unsat cores, interpolants, instances of quantified variables). Nevertheless, proof production is non-trivial [11], and the best, most efficient provers do not necessarily generate the best, least redundant proofs.

For proofs using propositional resolution generated by SAT- and SMT-solvers, there is a wide variety of proof compression techniques. Algebraic properties of the resolution operation that might be useful for compression were investigated in [5]. Compression algorithms based on rearranging and sharing chains of resolution inferences have been developed in [2] and [12]. Cotton [4] proposed an algorithm that compresses a refutation by repeatedly splitting it into a proof of a heuristically chosen literal  $\ell$  and a proof of  $\bar{\ell}$ , and then resolving them to form a new refutation. The `Reduce&Reconstruct` algorithm [10] searches for locally redundant subproofs that can be rewritten into subproofs of stronger clauses and with fewer resolution steps. A linear time proof compression algorithm based on partial regularization was proposed in [3] and improved in [6].

In contrast, there has been much less work on simplifying first-order proofs. For tree-like sequent calculus proofs, algorithms based on cut-introduction [9, 8] have been proposed. However, converting a DAG-like resolution or superposition

---

\* Supported by the Google Summer of Code 2014 program.

proof, as usually generated by current provers, into a tree-like sequent calculus proof may increase the size of the proof. For arbitrary proofs in the TPTP [13] format (including DAG-like first-order resolution proofs), there is a simple algorithm [15] that looks for terms that occur often in any TSTP [13] proof and introduces abbreviations for these terms.

The work reported in this paper is part of a new trend that aims at lifting successful propositional proof compression algorithms to first-order logic. Our first target was the propositional **LowerUnits** algorithm, which delays resolution steps with unit clauses, resulting in the **GreedyLinearFirstOrderLowerUnits** (**GFOLU**) algorithm [7]. Here we continue this line of research by lifting the **RecyclePivotsWithIntersection** (**RPI**) algorithm [6], which is an improvement of the **RecyclePivots** (**RP**) algorithm [3], providing better compression on proofs where nodes have several children.

Section 2 introduces the first-order resolution calculus and the notations used in this paper. Section 4 discusses the challenges that arise in the first-order case (mainly due to unification), which are not present in the propositional case. Section 5 describes an algorithm that overcomes these challenges. Section 6 concludes the paper by presenting experimental results obtained by applying this algorithm, and also its combinations with **GFOLU**, on hundreds of proofs generated with the SPASS theorem prover.

## 2 The Resolution Calculus

We assume that there are infinitely many variable symbols (e.g.  $X, Y, Z, X_1, X_2, \dots$ ), constant symbols (e.g.  $a, b, c, a_1, a_2, \dots$ ), function symbols of every arity (e.g.  $f, g, f_1, f_2, \dots$ ) and predicate symbols of every arity (e.g.  $p, q, p_1, p_2, \dots$ ). A *term* is any variable, constant or the application of an  $n$ -ary function symbol to  $n$  terms. An *atomic formula* (*atom*) is the application of an  $n$ -ary predicate symbol to  $n$  terms. A *literal* is an atom or the negation of an atom. The *complement* of a literal  $\ell$  is denoted  $\bar{\ell}$  (i.e. for any atom  $p$ ,  $\bar{p} = \neg p$  and  $\neg \bar{p} = p$ ). The set of all literals is denoted  $\mathcal{L}$ . A *clause* is a multiset of literals.  $\perp$  denotes the *empty clause*. A *unit clause* is a clause with a single literal. Sequent notation is used for clauses (i.e.  $p_1, \dots, p_n \vdash q_1, \dots, q_m$  denotes the clause  $\{\neg p_1, \dots, \neg p_n, q_1, \dots, q_m\}$ ).  $\text{FV}(t)$  (resp.  $\text{FV}(\ell)$ ,  $\text{FV}(\Gamma)$ ) denotes the set of variables in the term  $t$  (resp. in the literal  $\ell$  and in the clause  $\Gamma$ ). A *substitution*  $\{X_1 \setminus t_1, X_2 \setminus t_2, \dots\}$  is a mapping from variables  $\{X_1, X_2, \dots\}$  to, respectively, terms  $\{t_1, t_2, \dots\}$ . The application of a substitution  $\sigma$  to a term  $t$ , a literal  $\ell$  or a clause  $\Gamma$  results in, respectively, the term  $t\sigma$ , the literal  $\ell\sigma$  or the clause  $\Gamma\sigma$ , obtained from  $t$ ,  $\ell$  and  $\Gamma$  by replacing all occurrences of the variables in  $\sigma$  by the corresponding terms in  $\sigma$ . The set of all substitutions is denoted  $\mathcal{S}$ . A *unifier* of a set of literals is a substitution that makes all literals in the set equal. A *resolution proof* is a directed acyclic graph of clauses where the edges correspond to the inference rules of resolution and contraction (as explained in detail in Definition 1). A *resolution refutation* is a resolution proof with root  $\perp$ .

**Definition 1 (First-Order Resolution Proof).**

A directed acyclic graph  $\langle V, E, \Gamma \rangle$ , where  $V$  is a set of nodes and  $E$  is a set of edges labeled by literals and substitutions (i.e.  $E \subset V \times 2^{\mathcal{L}} \times \mathcal{S} \times V$  and  $v_1 \xrightarrow[\sigma]{\ell} v_2$  denotes an edge from node  $v_1$  to node  $v_2$  labeled by the literal  $\ell$  and the substitution  $\sigma$ ), is a proof of a clause  $\Gamma$  iff it is inductively constructible according to the following cases:

- **Axiom:** If  $\Gamma$  is a clause,  $\hat{\Gamma}$  denotes some proof  $\langle \{v\}, \emptyset, \Gamma \rangle$ , where  $v$  is a new (axiom) node.
- **Resolution:** If  $\psi_L$  is a proof  $\langle V_L, E_L, \Gamma_L \rangle$  with  $\ell_L \in \Gamma_L$  and  $\psi_R$  is a proof  $\langle V_R, E_R, \Gamma_R \rangle$  with  $\ell_R \in \Gamma_R$ , and  $\sigma_L$  and  $\sigma_R$  are substitutions such that  $\ell_L \sigma_L = \bar{\ell}_R \sigma_R$  and  $\text{FV}((\Gamma_L \setminus \{\ell_L\}) \sigma_L) \cap \text{FV}((\Gamma_R \setminus \{\ell_R\}) \sigma_R) = \emptyset$ , then  $\psi_L \odot_{\ell_L \ell_R}^{\sigma_L \sigma_R} \psi_R$  denotes a proof  $\langle V, E, \Gamma \rangle$  s.t.

$$\begin{aligned} V &= V_L \cup V_R \cup \{v\} \\ E &= E_L \cup E_R \cup \left\{ \rho(\psi_L) \xrightarrow[\sigma_L]{\ell_L} v, \rho(\psi_R) \xrightarrow[\sigma_R]{\ell_R} v \right\} \\ \Gamma &= (\Gamma_L \setminus \{\ell_L\}) \sigma_L \cup (\Gamma_R \setminus \{\ell_R\}) \sigma_R \end{aligned}$$

where  $v$  is a new (resolution) node and  $\rho(\varphi)$  denotes the root node of  $\varphi$ . The resolved atom  $\ell$  is such that  $\ell = \ell_L \sigma_L = \bar{\ell}_R \sigma_R$  or  $\ell = \bar{\ell}_L \sigma_L = \ell_R \sigma_R$ .

- **Contraction:** If  $\psi'$  is a proof  $\langle V', E', \Gamma' \rangle$  and  $\sigma$  is a unifier of  $\{\ell_1, \dots, \ell_n\}$  with  $\{\ell_1, \dots, \ell_n\} \subseteq \Gamma'$ , then  $[\psi]_{\{\ell_1, \dots, \ell_n\}}^\sigma$  denotes a proof  $\langle V, E, \Gamma \rangle$  s.t.

$$\begin{aligned} V &= V' \cup \{v\} \\ E &= E' \cup \left\{ \rho(\psi') \xrightarrow[\sigma]{\{\ell_1, \dots, \ell_n\}} v \right\} \\ \Gamma &= (\Gamma' \setminus \{\ell_1, \dots, \ell_n\}) \sigma \cup \{\ell\} \end{aligned}$$

where  $v$  is a new (contraction) node,  $\ell = \ell_k \sigma$  (for any  $k \in \{1, \dots, n\}$ ) and  $\rho(\varphi)$  denotes the root node of  $\varphi$ .  $\square$

### 3 The Propositional Algorithm

RPI (formally defined in Appendix A) removes *irregularities*, which are resolution inferences with a node  $\eta$  when the resolved literal (a.k.a. *pivot*) occurs as the pivot of another inference located below in the path from  $\eta$  to the root of the proof. In the worst case, regular resolution proofs can be exponentially bigger than irregular ones, but RPI takes care of regularizing the proof only partially, removing inferences only when this does not enlarge the proof.

RPI traverses the proof twice. On the first traversal (bottom-up), it stores for each node a set of *safe literals* that are resolved in all paths below it in the proof or that occur in the root clause of the proof. If one of the node's resolved literals belongs to the set of safe literals, then it is possible to *regularize* the node by replacing it by the parent containing the safe literal. To do this replacement

efficiently, the replacement is postponed by marking the other parent as to be **deleted**. Then, on a single second traversal (top-down), regularization is performed: any node that has a parent node marked to be **deleted** is replaced by its other parent.

The RPI and the RP algorithms differ from each other mainly in the computation of the safe literals of a node that has many children. While the former returns the intersection as shown in Algorithm 6, the latter returns the empty set. Moreover, while in RPI the safe literals of the root node contain all the literals of the root clause, in RP the root node is always assigned an empty set of literals.

## 4 First-Order Challenges

In this section, we describe challenges that have to be overcome in order to successfully adapt RPI to the first-order case. The first example illustrates the need to take unification into account. The other two examples discuss complex issues that can arise when unification is taken into account in a naive way.

*Example 1.* Consider the following proof  $\psi$ . When computed as in the propositional case, the safe literals for  $\eta_3$  are  $\{\vdash q(c), p(a, X)\}$ .

$$\frac{\frac{\eta_1: \vdash p(W, X) \quad \eta_2: p(W, X) \vdash q(c)}{\eta_3: \vdash q(c)} \quad \eta_4: q(c) \vdash p(a, X)}{\frac{\eta_5: \vdash p(a, X) \quad \eta_6: p(Y, b) \vdash}{\psi: \perp}}$$

As neither of  $\eta_3$ 's pivots is syntactically equal to a safe literal, the propositional RPI algorithm would not change  $\psi$ . However,  $\eta_3$ 's left pivot  $p(W, X) \in \eta_1$  is unifiable with the safe literal  $p(a, X)$ . Regularizing  $\eta_3$ , by deleting the edge between  $\eta_2$  and  $\eta_3$  and replacing  $\eta_3$  by  $\eta_1$ , leads to further deletion of  $\eta_4$  (because it is not resolvable with  $\eta_1$ ) and finally to the much shorter proof below.

$$\frac{\eta_1: \vdash p(W, X) \quad \eta_6: p(Y, b) \vdash}{\psi': \perp}$$

Unlike in the propositional case, where a pivot must be syntactically equal to a safe literal for regularization to be possible, the example above suggests that, in the first-order case, it might suffice that a pivot be unifiable with a safe literal. However, there are cases, as shown in the example below, where mere unifiability is not enough and greater care is needed.

*Example 2.* Again, the safe literals for  $\eta_3$ , when computed as in the propositional case, are  $\{\vdash q(c), p(a, X)\}$ , and as the pivot  $p(a, c)$  is unifiable with the safe literal  $p(a, X)$ ,  $\eta_3$  appears to be a candidate for regularization.

$$\frac{\frac{\eta_1: \vdash p(a, c) \quad \eta_2: p(a, c) \vdash q(c)}{\eta_3: \vdash q(c)} \quad \eta_4: q(c) \vdash p(a, X)}{\frac{\eta_5: \vdash p(a, X) \quad \eta_6: p(Y, b) \vdash}{\psi: \perp}}$$

```

input : A first-order proof  $\psi$ 
output: A possibly less-irregular first-order proof  $\psi'$ 
1  $\psi' \leftarrow \psi$ ;
2 traverse  $\psi'$  bottom-up and foreach node  $\eta$  in  $\psi'$  do
3   if  $\eta$  is a resolvent node then
4     setSafeLiterals( $\eta$ ) ;
5     regularizeIfPossible( $\eta$ )
6  $\psi' \leftarrow \text{fix}(\psi')$  ;
7 return  $\psi'$ ;

```

**Algorithm 1:** FORPI

However, if we attempt to regularize the proof, the same series of actions as in Example 1 would require resolution between  $\eta_1$  and  $\eta_6$ , which is not possible.

One way to prevent the problem depicted above would be to require the pivot to be not only unifiable but in fact more general than a safe literal. A weaker (and better) requirement is possible, however, as defined below.

**Definition 2.** Let  $\eta$  be a node with pivot  $\ell'$  unifiable with safe literal  $\ell$  which is resolved against literals  $\ell_1, \dots, \ell_n$  in a proof  $\psi$ .  $\eta$  is said to satisfy the pre-regularization unifiability property in  $\psi$  if  $\ell_1, \dots, \ell_n$ , and  $\bar{\ell}'$  are unifiable.

One way to ensure this property is met is to slightly modify the notion of safe literals, by applying the unifier of the resolution step to the each pivot before adding it to the safe literals (cf. algorithm 3, lines 8 and 10). In the case of Example 2, this would result in  $\eta_3$  having the safe literals  $\{\vdash q(c), p(a, b)\}$ , where clearly the pivot  $p(a, c)$  in  $\eta_1$  is not safe.

*Example 3.* Satisfying the pre-regularization unifiability property is not sufficient. Consider the proof  $\psi$  below. After collecting the safe literals,  $\eta_3$ 's safe literals are  $\{q(T, V), p(c, d) \vdash q(f(a, e), c)\}$ .

$$\begin{array}{c}
\frac{\eta_1: p(U, V) \vdash q(f(a, V), U) \quad \eta_2: q(f(a, X), Y), q(T, X) \vdash q(f(a, Z), Y)}{\eta_3: p(U, V), Q(T, V) \vdash q(f(a, Z), U)} \quad \eta_4: \vdash q(R, S) \\
\frac{\eta_6: \vdash p(c, d) \quad \eta_5: p(U, V) \vdash q(f(a, Z), U)}{\eta_8: q(f(a, e), c) \vdash \quad \eta_7: \vdash q(f(a, Z), c)} \\
\psi: \perp
\end{array}$$

$\eta_3$ 's pivot  $q(f(a, V), U)$  is unifiable to (and even more general than) the safe literal  $q(f(a, e), c)$ . Attempting to regularize  $\eta_3$  would lead to the removal of  $\eta_2$ , the replacement of  $\eta_3$  by  $\eta_1$  and the removal of  $\eta_4$  (because  $\eta_1$  does not contain the pivot required by  $\eta_5$ ), with  $\eta_5$  also being replaced by  $\eta_1$ . Then resolution between  $\eta_1$  and  $\eta_6$  results in  $\eta'_7$ , which cannot be resolved with  $\eta_8$ , as shown below.

$$\begin{array}{c}
\eta_8: Q(f(a, e), c) \vdash \quad \frac{\eta_6: \vdash P(c, d) \quad \eta_1: P(U, V) \vdash Q(f(a, V), U)}{\eta'_7: \vdash Q(f(a, d), c)} \\
\psi': ??
\end{array}$$

$\eta_1$ 's literal  $q(f(a, V), U)$ , which would be resolved with  $\eta_8$ 's literal, was changed to  $Q(f(a, d), c)$  due to the resolution between  $\eta_1$  and  $\eta_6$ .

Thus we additionally require the following property be satisfied.

**Definition 3.** *Let  $\eta$  be a node with safe literals  $\phi$  that is marked for regularization with parents  $\eta_1$  and  $\eta_2$ , where  $\eta_2$  is marked **deleted** in a proof  $\psi$ .  $\eta$  is said to satisfy the regularization unifiability property in  $\psi$  if there exists a substitution  $\sigma$  such that  $\sigma\eta_1 \subseteq \phi$ .*

This property ensures that the remainder of the proof does not expect a variable in  $\eta_1$  to be unified to different values simultaneously. This property is not necessary in the propositional case, as the replacement node would not change lower in the proof.

## 5 First-Order RecyclePivotsWithIntersection

This section presents **FirstOrderRecyclePivotsWithIntersection** (FORPI), Algorithm 1, a first order generalization of RPI. FORPI traverses the proof in a bottom-up manner, storing for every node a set of safe literals. The set of safe literals for a node  $\psi$  is computed from the set of safe literals of its children (cf. Algorithm 3), similarly to the propositional case, but additionally applying unifiers to the resolved pivots (cf. Example 2). If one of the node's resolved literals can be unified to a literal in the set of safe literals, then it may be possible to regularize the node by replacing it by one of its parents.

In the first order case, we additionally check for the regularization property (cf. lines 2 and 6 of Algorithm 2). Similarly to RPI, instead of replacing the irregular node by one of its parents immediately, its other parent is marked as a **deletedNode**, as shown in Algorithm 2. As in the propositional case, fixing of the proof is postponed to another (single) traversal, as regularization proceeds bottom up and only nodes below a regularized node may require fixing. During fixing, the irregular node is actually replaced by the parent that is not marked as **deletedNode**. During proof fixing, factoring inferences can be applied, in order to compress the proof further.

## 6 Experiments

A prototype<sup>1</sup> version of FORPI has been implemented in the functional programming language Scala as part of the **Skeptik** library. Evaluation used the same proofs that were used to evaluate GFOLU [7]. The SPASS (<http://www.spass-prover.org/>) theorem prover was executed on 2280 real first-order problems without equality of the TPTP Problem Library (among them, 1032 problems are known to be unsatisfiable). In order to generate pure resolution proofs, the advanced inference rules of SPASS were disabled. The proofs were generated on

<sup>1</sup> Source code available at <https://github.com/jgorzny/Skeptik>

```

input : A node  $\psi = \psi_L \odot_{\ell_L \ell_R}^{\sigma_L \sigma_R} \psi_R$ 
output: nothing (but the proof containing  $\psi$  may be changed)
1 if  $\exists \sigma$  and  $\ell \in \psi.\text{safeLiterals}$  such that  $\sigma\ell = \ell_R$  or  $\ell = \sigma\ell_R$  then
2   if  $\exists \sigma'$  such that  $\sigma'\psi_R \subseteq \psi.\text{safeLiterals}$  then
3     replace  $\psi_L$  by deletedNodeMarker ;
4     mark  $\psi$  as regularized
5 else if  $\exists \sigma$  and  $\ell \in \psi.\text{safeLiterals}$  such that  $\sigma\ell = \ell_L$  or  $\ell = \sigma\ell_L$  then
6   if  $\exists \sigma'$  such that  $\sigma'\psi_L \subseteq \psi.\text{safeLiterals}$  then
7     replace  $\psi_R$  by deletedNodeMarker ;
8     mark  $\psi$  as regularized

```

**Algorithm 2: F0regularizeIfPossible**

```

input : A first order resolution node  $\psi$ 
output: nothing (but the node  $\psi$  gets a set of safe literals)
1 if  $\psi$  is a root node with no children then
2    $\psi.\text{safeLiterals} \leftarrow \psi.\text{clause}$ 
3 else
4   foreach  $\psi' \in \psi.\text{children}$  do
5     if  $\psi'$  is marked as regularized then
6        $\text{safeLiteralsFrom}(\psi') \leftarrow \psi'.\text{safeLiterals}$  ;
7     else if  $\psi' = \psi \odot_{\ell_L \ell_R}^{\sigma_L \sigma_R} \psi_R$  for some  $\psi_R$  then
8        $\text{safeLiteralsFrom}(\psi') \leftarrow \psi'.\text{safeLiterals} \cup \{ \sigma_R \ell_R \}$ 
9     else if  $\psi' = \psi_L \odot_{\ell_L \ell_R}^{\sigma_L \sigma_R} \psi$  for some  $\psi_L$  then
10       $\text{safeLiteralsFrom}(\psi') \leftarrow \psi'.\text{safeLiterals} \cup \{ \sigma_L \ell_L \}$ 
11  $\psi.\text{safeLiterals} \leftarrow \bigcap_{\psi' \in \psi.\text{children}} \text{safeLiteralsFrom}(\psi')$ 

```

**Algorithm 3: F0setSafeLiterals**

the Euler Cluster at the University of Victoria with a time limit of 300 seconds per problem. Under these conditions, SPASS generated 308 proofs.

Figure 1 (a) shows the compression obtained by applying FORPI and GFOLU (as well as their combinations) to the proofs. With this data set, FORPI compresses a few proofs only, and its performance is not as good as that of GFOLU. Furthermore, when FORPI is combined with GFOLU, FORPI provides additional compression to only three proofs already compressed by GFOLU. This is surprising, because in the propositional case, RPI usually compresses up to ten times more than LowerUnits. Nevertheless, this can be easily explained by the fact that all available benchmark proofs have small heights (not more than ToDo); consequently the path from any node to the root is short and unlikely to contain irregularities. In the propositional case, on the other hand, RPI has been tested on proofs that are a thousand times higher.

Figure 1 (b) shows that the order of compression may matter less than it does in the propositional case, since only a few points occur outside the diagonal. However, this might be due to the small size of the proofs.

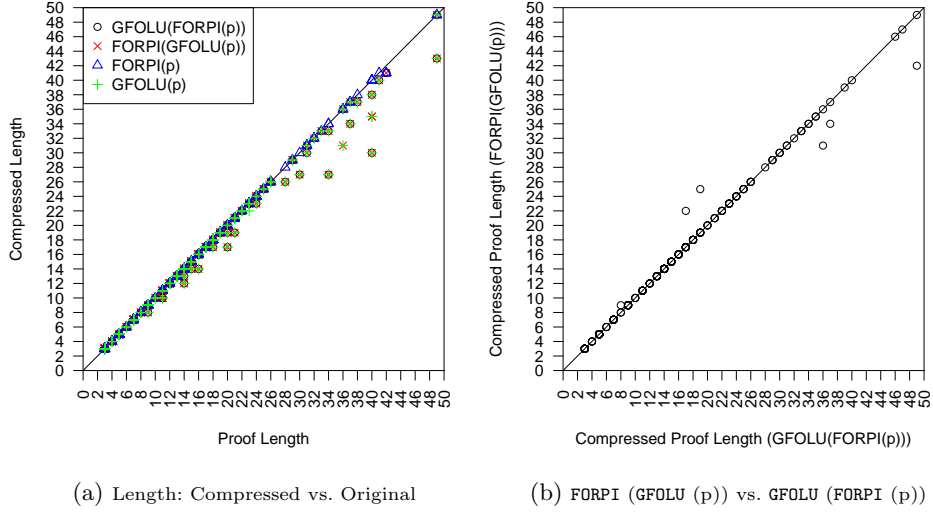


Fig. 1: Scatter Plots of Experimental Results

SPASS required approximately 40 minutes to solve and generate the proofs; the total time for GFOLU and FORPI to be executed on all 308 proofs was just under 8 seconds (both include parsing time).

## 7 Conclusions and Future Work

The main contribution of this paper is the generalization of the propositional proof compression algorithm RPI to the first-order case. As indicated in Section 4, the generalization is challenging, because unification changes the pivots and, consequently, must be taken into account when collecting safe literals and marking nodes for deletion.

Every computational experiment evaluates not only the algorithm but also the data on which it is executed. Although the experimental results are not as promising as expected, this is due to the fact that the 308 proofs currently available are too short to contain a significant amount of irregularities. This is a valuable piece of information, allowing us to conclude that it is not worth to apply FORPI on pure resolution proofs which current state-of-the-art first-order theorem provers seem capable of producing. Nevertheless, based on our positive results for RPI on much longer proofs generated by SAT and SMT solvers [6], FORPI remains a promising option to be revisited in the future, when the performance of first-order theorem provers catch up with advances in SAT and SMT and taller first-order benchmark proofs become available.



## References

1. *Logic for Programming, Artificial Intelligence, and Reasoning - 16th Intl. Conf., Dakar, Senegal, Rev. Selected Papers*, LNCS. Springer, 2010.
2. Hasan Amjad. Compressing propositional refutations. *Electr. Notes Theor. Comput. Sci.*, 185:3–15, 2007.
3. O. Bar-Ilan, O. Fuhrmann, S. Hoory, O. Shacham, and O. Strichman. Linear-time reductions of resolution proofs. In *Haifa Verif. Conf.*, LNCS, pages 114–128. Springer, 2008.
4. Scott Cotton. Two techniques for minimizing resolution proofs. In Ofer Strichman and Stefan Szeider, editors, *Theory and Applications of Satisfiability Testing SAT 2010*, volume 6175 of *LNCS*, pages 306–312. Springer, 2010.
5. P. Fontaine, S. Merz, and B. Woltzenlogel Paleo. Exploring and exploiting algebraic and graphical properties of resolution. In *8th Intl. Wkshp. on SMT*, Edinburgh, 2010.
6. P. Fontaine, S. Merz, and B. Woltzenlogel Paleo. Compression of propositional resolution proofs via partial regularization. In *CADE*, LNCS, pages 237–251. Springer, 2011.
7. J. Gorzny and B. Woltzenlogel Paleo. Towards the compression of first-order resolution proofs by lowering unit clauses. In *CADE*, 2015.
8. S. Hetzl, A. Leitsch, G. Reis, and D. Weller. Algorithmic introduction of quantified cuts. *Theor. Comput. Sci.*, 549:1–16, 2014.
9. B. Woltzenlogel Paleo. Atomic cut introduction by resolution: Proof structuring and compression. In *LPAR-16* [1], pages 463–480.
10. S. F. Rollini, R. Bruttomesso, and N. Sharygina. An efficient and flexible approach to resolution proof reduction. In *Hardware and Software: Verification and Testing*, LNCS, pages 182–196. Springer, 2011.
11. Stephan Schulz and Geoff Sutcliffe. Proof generation for saturating first-order theorem provers. In David Delahaye and Bruno Woltzenlogel Paleo, editors, *All about Proofs, Proofs for All*, volume 55 of *Mathematical Logic and Foundations*. College Publications, London, UK, January 2015.
12. Carsten Sinz. Compressing propositional proofs by common subproof extraction. In Roberto Moreno-Díaz, Franz Pichler, and Alexis Quesada-Arencibia, editors, *EUROCAST*, volume 4739 of *LNCS*, pages 547–555. Springer, 2007.
13. G. Sutcliffe. The TPTP Problem Library and Associated Infrastructure: The FOF and CNF Parts, v3.5.0. *Journal of Automated Reasoning*, 43(4):337–362, 2009.
14. G. S. Tseitin. On the complexity of derivation in propositional calculus. In J. Siekmann and G. Wrightson, editors, *Automation of Reasoning: Classical Papers in Computational Logic 1967-1970*, volume 2. Springer-Verlag, 1983.
15. J. Vyskocil, D. Stanovský, and J. Urban. Automated proof compression by invention of new definitions. In *LPAR-16* [1], pages 447–462.

## A Algorithm RecyclePivotsWithIntersection

**Note:** for the reviewers’ convenience, this appendix summarizes [6].

**RecyclePivotsWithIntersection** (RPI) [6] aims at compressing irregular proofs. It can be seen as a simple but significant modification of the **RP** algorithm described in [3], from which it derives its name. Although in the worst case full

regularization can increase the proof length exponentially [14], these algorithms show that many irregular proofs can have their length decreased if a careful partial regularization is performed.

Consider an irregular proof of the form  $\psi[\eta \odot_p \psi'[\eta' \odot_p \eta'']]$  and assume, without loss of generality, that  $p \in \eta$  and  $p \in \eta'$ . Then, if  $\eta' \odot_p \eta''$  is replaced by  $\eta''$  within the proof-context  $\psi'[\ ]$ , the clause  $\eta \odot_p \psi'[\eta'']$  subsumes the clause  $\eta \odot_p \psi'[\eta' \odot_p \eta'']$ , because even though the literal  $\neg p$  of  $\eta''$  is propagated down, it gets resolved against the literal  $p$  of  $\eta$  later on below in the proof. More precisely, even though it might be the case that  $\neg p \in \psi'[\eta'']$  while  $\neg p \notin \psi'[\eta' \odot_p \eta'']$ , it is necessarily the case that  $\neg p \notin \eta \odot_p \psi'[\eta' \odot_p \eta'']$  and  $\neg p \notin \eta \odot_p \psi'[\eta'']$ .

Although the remarks above suggest that it is safe to replace  $\eta' \odot_p \eta''$  by  $\eta''$  within the proof-context  $\psi'[\ ]$ , this is not always the case. If a node in  $\psi'[\ ]$  has a child in  $\psi[\ ]$ , then the literal  $\neg p$  might be propagated down to the root of the proof, and hence, the clause  $\psi[\eta \odot_p \psi'[\eta'']]$  might not subsume the clause  $\psi[\eta \odot_p \psi'[\eta' \odot_p \eta'']]$ . Therefore, it is only safe to do the replacement if the literal  $\neg p$  gets resolved in all paths from  $\eta''$  to the root or if it already occurs in the root clause of the original proof  $\psi[\eta \odot_p \psi'[\eta' \odot_p \eta'']]$ .

These observations lead to the idea of traversing the proof in a bottom-up manner, storing for every node a set of *safe literals* that get resolved in all paths below it in the proof (or that already occurred in the root clause of the original proof). Moreover, if one of the node's resolved literals belongs to the set of safe literals, then it is possible to regularize the node by replacing it by one of its parents (cf. Algorithm 4).

The regularization of a node should replace a node by one of its parents, and more precisely by the parent whose clause contains the resolved literal that is safe. After regularization, all nodes below the regularized node may have to be fixed. However, since the regularization is done with a bottom-up traversal, and only nodes below the regularized node need to be fixed, it is again possible to postpone fixing and do it with only a single traversal afterwards. Therefore, instead of replacing the irregular node by one of its parents immediately, its other parent is replaced by **deletedNodeMarker**, as shown in Algorithm 5. Only later during fixing, the irregular node is actually replaced by its surviving parent (i.e. the parent that is not **deletedNodeMarker**).

<p><b>input</b> : A proof <math>\psi</math>  <b>output</b>: A possibly less-irregular proof <math>\psi'</math></p> <pre> 1 <math>\psi' \leftarrow \psi</math>; 2 traverse <math>\psi'</math> bottom-up and <b>foreach</b> node <math>\eta</math> in <math>\psi'</math> <b>do</b> 3   <b>if</b> <math>\eta</math> is a resolvent node <b>then</b> 4     setSafeLiterals(<math>\eta</math>) ; 5     regularizeIfPossible(<math>\eta</math>) 6 <math>\psi' \leftarrow \text{fix}(\psi')</math> ; 7 <b>return</b> <math>\psi'</math>; </pre>
--

**Algorithm 4:** RPI

The set of safe literals of a node  $\eta$  can be computed from the set of safe literals of its children (cf. Algorithm 6). In the case when  $\eta$  has a single child  $\varsigma$ , the safe literals of  $\eta$  are simply the safe literals of  $\varsigma$  together with the resolved literal  $p$  of  $\varsigma$  belonging to  $\eta$  ( $p$  is safe for  $\eta$ , because whenever  $p$  is propagated down the proof through  $\eta$ ,  $p$  gets resolved in  $\varsigma$ ). It is important to note, however, that if  $\varsigma$  has been marked as regularized, it will eventually be replaced by  $\eta$ , and hence  $p$  should not be added to the safe literals of  $\eta$ . In this case, the safe literals of  $\eta$  should be exactly the same as the safe literals of  $\varsigma$ . When  $\eta$  has several children, the safe literals of  $\eta$  w.r.t. a child  $\varsigma_i$  contain literals that are safe on all paths that go from  $\eta$  through  $\varsigma_i$  to the root. For a literal to be safe for all paths from  $\eta$  to the root, it should therefore be in the intersection of the sets of safe literals w.r.t. each child.

The RP and the RPI algorithms differ from each other mainly in the computation of the safe literals of a node that has many children. While RPI returns the intersection as shown in Algorithm 6, RP returns the empty set (cf. Algorithm 7). Additionally, while in RPI the safe literals of the root node contain all the literals of the root clause, in RP the root node is always assigned an empty set of literals. (Of course, this makes a difference only when the proof is not a refutation.) Note that during a traversal of the proof, the lines from 5 to 10 in Algorithm 6 are executed as many times as the number of edges in the proof. Since every node has at most two parents, the number of edges is at most twice the number of nodes. Therefore, during a traversal of a proof with  $n$  nodes, lines from 5 to 10 are executed at most  $2n$  times, and the algorithm remains linear. In our prototype implementation, the sets of safe literals are instances of Scala's `mutable.HashSet` class. Being mutable, new elements can be added efficiently. And being HashSets, membership checking is done in constant time in the average case, and set intersection (line 12) can be done in  $O(k.s)$ , where  $k$  is the number of sets and  $s$  is the size of the smallest set.

**input** : A node  $\eta$   
**output**: nothing (but the proof containing  $\eta$  may be changed)

```

1 if  $\eta$ .rightResolvedLiteral  $\in \eta$ .safeLiterals then
2   replace left parent of  $\eta$  by deletedNodeMarker ;
3   mark  $\eta$  as regularized
4 else if  $\eta$ .leftResolvedLiteral  $\in \eta$ .safeLiterals then
5   replace right parent of  $\eta$  by deletedNodeMarker ;
6   mark  $\eta$  as regularized

```

**Algorithm 5:** regularizeIfPossible

**input** : A node  $\eta$   
**output**: nothing (but the node  $\eta$  gets a set of safe literals)

```

1 if  $\eta$  is a root node with no children then
2    $\eta$ .safeLiterals  $\leftarrow \eta$ .clause
3 else
4   foreach  $\eta' \in \eta$ .children do
5     if  $\eta'$  is marked as regularized then
6       safeLiteralsFrom( $\eta'$ )  $\leftarrow \eta'$ .safeLiterals ;
7     else if  $\eta$  is left parent of  $\eta'$  then
8       safeLiteralsFrom( $\eta'$ )  $\leftarrow \eta'$ .safeLiterals  $\cup \{ \eta'.rightResolvedLiteral \}$  ;
9     else if  $\eta$  is right parent of  $\eta'$  then
10      safeLiteralsFrom( $\eta'$ )  $\leftarrow \eta'$ .safeLiterals  $\cup \{ \eta'.leftResolvedLiteral \}$  ;
11  $\eta$ .safeLiterals  $\leftarrow \bigcap_{\eta' \in \eta.children} \text{safeLiteralsFrom}(\eta')$ 

```

**Algorithm 6:** setSafeLiterals

**input** : A node  $\eta$   
**output**: nothing (but the node  $\eta$  gets a set of safe literals)

```

1 if  $\eta$  is a root node with no children then
2    $\eta$ .safeLiterals  $\leftarrow \emptyset$ 
3 else
4   if  $\eta$  has only one child  $\eta'$  then
5     if  $\eta'$  is marked as regularized then
6        $\eta$ .safeLiterals  $\leftarrow \eta'$ .safeLiterals ;
7     else if  $\eta$  is left parent of  $\eta'$  then
8        $\eta$ .safeLiterals  $\leftarrow \eta'$ .safeLiterals  $\cup \{ \eta'.rightResolvedLiteral \}$  ;
9     else if  $\eta$  is right parent of  $\eta'$  then
10       $\eta$ .safeLiterals  $\leftarrow \eta'$ .safeLiterals  $\cup \{ \eta'.leftResolvedLiteral \}$  ;
11   else
12      $\eta$ .safeLiterals  $\leftarrow \emptyset$ 

```

**Algorithm 7:** setSafeLiterals for RP