

# Partial Regularization of First-Order Resolution Proofs

Jan Gorzny<sup>a,\*</sup>, and Bruno Woltzenlogel Paleo<sup>b</sup>

<sup>a</sup> *School of Computer Science, University of Waterloo, 200 University Ave. W., Waterloo, ON N2L 3G1, Canada  
E-mail: jgorzny@uwaterloo.ca*

<sup>b</sup> *College of Engineering and Computer Science, Australian National University, Canberra ACT 0200, Australia  
Vienna University of Technology, Karlsplatz 13, 1040, Vienna, Austria  
E-mail: bruno@logic.at*

**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.

**Keywords:** proof compression, first-order logic, resolution, unification

## 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 that extract extra information from proofs (e.g. unsat cores, interpolants, instances of quantified variables). Nevertheless, proof production is non-trivial [12], and the best, most efficient provers do not necessarily generate the best, least redundant proofs.

For propositional resolution proofs, as those typically generated by SAT- and SMT-solvers, there is a wide variety of proof compression techniques. Algebraic properties of the resolution operation that are potentially useful for compression were investigated in [6]. Compression algorithms based on rearranging and sharing chains of resolution inferences have been developed in [2] and [13]. Cotton [5] proposed an algorithm that compresses a refutation by repeatedly split-

ting 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 [11] 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 [7].

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, 10] have been proposed. However, converting a DAG-like resolution or superposition 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 [14] format (including DAG-like first-order resolution proofs), there is a simple algorithm [16] that looks for terms that occur often in any TSTP [14] proof and abbreviates them.

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` (LU) algorithm, which delays resolution steps with unit clauses, resulting in the

---

\*Corresponding author. E-mail: jgorzny@uwaterloo.ca.

GreedyLinearFirstOrderLowerUnits (GFOLU) algorithm [8]. Here we continue this line of research by lifting the RecyclePivotsWithIntersection (RPI) algorithm [7], which improves the RecyclePivots (RP) algorithm [3] by detecting nodes that can be regularized even when they have several children.

Section 2 introduces the well-known first-order resolution calculus with notations that are suitable for manipulating proofs as first-class objects. 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 presents experimental results obtained by applying this algorithm, and its combinations with GFOLU, on hundreds of proofs generated with the SPASS theorem prover on TPTP benchmarks [14] and other randomly generated proofs. Section 7 concludes the paper.

## 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 *underlying atom* of a literal  $\ell$  is denoted  $|\ell|$  (i.e. for any atom  $p$ ,  $|p| = p$  and  $|\neg p| = p$ ). 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  $\sigma$ . 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 2.1). A *resolution refutation* is a resolution proof with root  $\perp$ .

### Definition 2.1 (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  $\psi_L \odot_{\ell_L \sigma_L}^{\sigma_L \sigma_R} \psi_R$  denotes a proof  $\langle V, E, \Gamma \rangle$  s.t.

$$V = V_L \cup V_R \cup \{v\}$$

$$E = E_L \cup E_R \cup \left\{ \rho(\psi_L) \xrightarrow[\sigma_L]{\ell_L} v, \rho(\psi_R) \xrightarrow[\sigma_R]{\ell_R} v \right\}$$

$$\Gamma = (\Gamma_L \setminus \{\ell_L\}) \sigma_L \cup (\Gamma_R \setminus \{\ell_R\}) \sigma_R$$

where  $v$  is a new (resolution) node and  $\rho(\varphi)$  denotes the root node of  $\varphi$ .  $\ell_L$  and  $\ell_R$  are  $v$ 's *resolved literals*, whereas  $\ell_L \sigma_L$  and  $\ell_R \sigma_R$  are its *instantiated resolved literals*. The *pivot* of  $v$  is the underlying atom of its instantiated resolved literals (i.e.  $|\ell_L \sigma_L|$  or, equivalently,  $|\ell_R \sigma_R|$ ).

- **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.

$$V = V' \cup \{v\}$$

$$E = E' \cup \left\{ \rho(\psi') \xrightarrow[\sigma]{\{\ell_1, \dots, \ell_n\}} v \right\}$$

$$\Gamma = (\Gamma' \setminus \{\ell_1, \dots, \ell_n\}) \sigma \cup \{\ell\}$$

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  $\varphi$ .  $\square$

**Example 2.1.** An example first-order resolution proof is shown below.



Fig. 1. The proof in Example 2.1.

$$\begin{array}{c}
 \frac{\eta_1: q(X), q(a) \vdash p(b) \quad \eta_2: p(b) \vdash}{\eta_3: q(X), q(a) \vdash} \quad \frac{\eta_2 \quad \eta_4: \vdash p(b), q(Y)}{\eta_5: \vdash q(Y)} \\
 \frac{\eta_3: q(X), q(a) \vdash \quad \eta_5: \vdash q(Y)}{\eta'_3: q(a) \vdash} \\
 \frac{\eta'_3: q(a) \vdash \quad \psi: \perp}{\psi: \perp}
 \end{array}$$

The nodes  $\eta_1$ ,  $\eta_2$ , and  $\eta_4$  are axioms. Node  $\eta_3$  is obtained by resolution on  $\eta_1$  and  $\eta_2$  where  $\ell_L = p(b)$ ,  $\ell_R = \neg p(b)$ , and  $\sigma_L = \sigma_R = \emptyset$ . The node  $\eta'_3$  is obtained by a contraction on  $\eta_3$  with  $\sigma = \{X \setminus a\}$ . The node  $\eta_5$  is the result of resolution on  $\eta_2$  and  $\eta_4$  with  $\ell_L = \neg p(b)$ ,  $\ell_R = p(b)$ ,  $\sigma_L = \sigma_R = \emptyset$ . Lastly, the conclusion node  $\psi$  is the result of a resolution of  $\eta'_3$  and  $\eta_5$ , where  $\ell_L = \neg q(a)$ ,  $\ell_R = q(Y)$ ,  $\sigma_L = \emptyset$ , and  $\sigma_R = \{Y \setminus a\}$ . The directed acyclic graph representation of the proof (with edge labels omitted) is shown in Figure 1.

### 3. The Propositional Algorithm

RPI (formally defined in Appendix A) removes *irregularities*, which are resolution inferences with a node  $\eta$  when the resolved literal 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 a *deletedNode*. Then, on a single second traversal (top-down), regularization is performed: any node that has a parent node marked as a *deletedNode* 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 4.1.** Consider the following proof  $\psi$ . When computed as in the propositional case, the safe literals for  $\eta_3$  are  $\{q(c), p(a, X)\}$ .

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

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 4.2.** Again, the safe literals for  $\eta_3$ , when computed as in the propositional case, are  $\{q(c), p(a, X)\}$ , 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.

$$\frac{\eta_1: \vdash p(a, c) \quad \eta_2: p(a, c) \vdash q(c)}{\eta_3: \vdash q(c)} \quad \eta_4: q(c) \vdash p(a, X)$$

$$\frac{\eta_6: p(Y, b) \vdash \quad \eta_5: \vdash p(a, X)}{\psi: \perp}$$

However, if we attempt to regularize the proof, the same series of actions as in Example 4.1 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 4.1.** Let  $\eta$  be a node with safe literals  $\mathcal{S}(\eta)$  and parents  $\eta_1$  and  $\eta_2$ , assuming without loss of generality,  $\eta_2 \xrightarrow[\sigma_2]{\{\ell_2\}} \eta$  such that  $\ell_2$  is unifiable with a safe literal  $\ell \in \mathcal{S}(\eta)$ . Let  $\mathcal{R}(\eta)$  be the set of all resolved literals  $\ell_1$  such that  $\eta_2 \xrightarrow[\sigma_2]{\{\ell_2\}} \eta'$  and  $\eta_1 \xrightarrow[\sigma_1]{\{\ell_1\}} \eta'$  for some  $\eta_1$  and  $\eta'$ . The node  $\eta$  is said to satisfy the *pre-regularization unifiability property* in the proof  $\psi$  if all elements in  $\mathcal{R}(\eta) \cup \{\ell_2\}$  are unifiable.

This property states that a node satisfies the pre-regularization unifiability property if, for a safe resolved literal  $\ell'$  which is resolved against literals  $\ell_1, \dots, \ell_n$  in a proof  $\psi$ ,  $\ell_1, \dots, \ell_n$ , 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 set of safe literals (cf. algorithm 3, lines 8 and 10). In the case of Example 4.2, this would result in  $\eta_3$  having the safe literals  $\{q(c), p(a, b)\}$ , where clearly the pivot  $p(a, c)$  in  $\eta_1$  is not safe.

**Example 4.3.** Satisfying the pre-regularization unifiability property is not sufficient. Consider the proof  $\psi$  in Figure 2. After collecting the safe literals,  $\eta_3$ 's safe literals are  $\{\neg q(T, V), \neg p(c, d), q(f(a, e), c)\}$ .  $\eta_3$ '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.

$$\frac{\eta_8: q(f(a, e), c) \vdash \quad \eta_7': \vdash q(f(a, d), c)}{\psi': ??}$$

$\eta_1$ '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 that the following property be satisfied.

**Definition 4.2.** Let  $\eta$  be a node satisfying the pre-regularization unifiability property, with safe literals  $\mathcal{S}(\eta)$  and parents  $\eta_1$  and  $\eta_2$ , assuming without loss of generality that  $\eta_2 \xrightarrow[\sigma_2]{\{\ell_2\}} \eta$  and  $\bar{\ell}_2 = \ell_1 \in \mathcal{S}(\eta)$ . The node  $\eta$  is said to satisfy the *regularization unifiability property* in  $\psi$  if there exists a substitution  $\sigma$  such that  $\eta_1\sigma \subseteq \mathcal{S}(\eta)$ .

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.

An alternative to the regularization property may also be useful for proof compression in some cases. This alternative relies on knowledge of how literals are changed after the deletion of a node in a proof (similar to the *post-deletion unifiability* property observed for FirstOrderLowerUnits in [8]).

**Definition 4.3.** Let  $\eta_1$  be a node with safe literals  $\mathcal{S}(\eta)$  such that there exists nodes  $\eta_2$  and  $\eta$  and resolved literal  $\ell'$  such that  $\eta_1 \xrightarrow[\sigma']{\{\ell'\}} \eta$  and  $\eta_2 \xrightarrow[\sigma'']{\{\ell'\}} \eta$ . For each literal  $\ell \in \mathcal{S}(\eta_1)$ , let  $\eta_\ell$  be the lowest node on the path from  $\eta_1$  to the root of the proof such that  $\ell$  is used as the pivot. Let  $\mathcal{R}(\eta_\ell)$  be the set of all resolved literals  $\bar{\ell}$  such that  $\eta'_2 \xrightarrow[\sigma_2]{\{\ell\}} \eta_\ell$  and  $\eta'_1 \xrightarrow[\sigma_1]{\{\bar{\ell}\}} \eta_\ell$  for some nodes  $\eta'_2$  and  $\eta'$ . The node  $\eta_1$  is said to satisfy the *post-deletion unifiability property* in  $\psi$  if, for all  $\ell \in \mathcal{S}(\eta_1)$ , all elements in  $\mathcal{R}^\dagger(\eta_\ell) \cup \{\ell^\dagger\}$  are unifiable, where  $\ell^\dagger$  is the literal in  $\psi \setminus \{\eta_2\}$  corresponding to  $\ell$  in  $\psi$  and  $\mathcal{R}^\dagger(\eta_\ell)$  is the set of literals in  $\psi \setminus \{\eta_2\}$  corresponding to literals of  $\mathcal{R}(\eta_\ell)$  in  $\psi$ .

Informally, a node  $\eta$  satisfies the *post-deletion unifiability property* in a proof if it is the parent of a resolution and we can delete the opposite parent of that same resolution, replacing the resolution node with the initial node such that all of its safe literals can still be used as pivots in order to complete the proof.

**Example 4.4.** This example illustrates the usefulness of the post-deletion unifiability property as an alternative to the regularization property. In the proof below, note that  $\eta_8$  does not satisfy the regularization property: there is no unifier  $\sigma$  such that  $\sigma\eta_8 \subseteq \mathcal{S}(\eta_8)$  (where  $\mathcal{S}(\eta_8) = \{\neg q(Y), \neg r(Z), \neg p(W)\}$ ).

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

Fig. 2. An example where pre-regularization unifiability is not sufficient.

$$\begin{array}{c}
\frac{\eta_5: p(Z) \vdash q(Z) \quad \frac{\eta_8: p(X), q(X), r(X) \vdash \quad \eta_7: \vdash p(Y)}{\eta_6: q(Y), r(Y) \vdash}}{\eta_4: p(Z), r(Z) \vdash} \quad \eta_3: \vdash r(W) \\
\frac{\eta_1: \vdash p(U) \quad \eta_2: p(W) \vdash}{\psi: \perp}
\end{array}$$

We show that this proof satisfies the requirements of the post-deletion unifiability property, and that  $\eta_7$  can be removed. To do this, first observe that  $\eta_8$  satisfies the pre-regularization unifiability property, since  $\neg p(X) \in \eta_8$  is in  $\mathcal{S}(\eta_8)$ . Consider the following proof of  $\psi \setminus \{\eta_7\}$ :

$$\begin{array}{c}
\frac{\eta_8: p(X), q(X), r(X) \vdash \quad \eta_5: p(Z) \vdash q(Z)}{\eta_4': p(Z), p(Z), r(Z) \vdash} \\
\frac{\eta_4': p(Z), p(Z), r(Z) \vdash \quad \eta_3: \vdash r(W)}{\eta_4: p(Z), r(Z) \vdash} \quad \eta_2: p(W) \vdash \\
\frac{\eta_1: \vdash p(U) \quad \eta_2: p(W) \vdash}{\psi: \perp}
\end{array}$$

Now observe that for each  $\ell \in \mathcal{S}(\eta_8) = \{\neg q(Y), \neg r(Z), \neg p(W)\}$ , the post-deletion unifiability property is satisfied:

- $\ell = \neg q(Y)$ :  $\ell^\dagger = \neg q(X)$  which is unifiable with  $\bar{\ell}^\dagger = q(Z)$
- $\ell = \neg r(Z)$ :  $\ell^\dagger = \neg r(Z)$  which is unifiable with  $\bar{\ell}^\dagger = r(W)$
- $\ell = \neg p(W)$ :  $\ell^\dagger = \neg p(W)$  which is unifiable with  $\bar{\ell}^\dagger = p(U)$

The post-deletion unifiability property is implied by the previous two properties, as shown in the next theorem.

**Theorem 4.5.** *Let  $\eta$  be a node with parents  $\eta_1$  and  $\eta_2$  in some proof that satisfies the pre-regularization and regularization unifiability properties. Then one of  $\eta_1$  and  $\eta_2$  satisfies the post-deletion unifiability property.*

*Proof.* Let  $\eta$  be a node that satisfies the pre-regularization and regularization unifiability properties with parents  $\eta_1$  and  $\eta_2$ . Assume without loss of generality that  $\eta_1 \sigma \subseteq \mathcal{S}(\eta)$ .

Let  $\ell \in \mathcal{S}(\eta_1)$  be a safe literal of  $\eta_1$ , and let  $\eta_\ell$  be the lowest node on the path from  $\eta_1$  to the root of the proof such that  $\ell$  is used as the pivot. Let  $\mathcal{R}(\eta_\ell)$  be the

set of all resolved literals  $\bar{\ell}$  such that  $\eta'_2 \xrightarrow[\sigma_2]{\{\ell\}} \eta_\ell$  and  $\eta'_1 \xrightarrow[\sigma_1]{\{\bar{\ell}\}} \eta_\ell$  for some nodes  $\eta'_1$  and  $\eta'_2$ . Let  $\mathcal{R}^\dagger(\eta_\ell)$  be defined as in Definition 4.3. Consider  $\bar{\ell} \in \mathcal{R}(\eta_\ell)$ :  $\ell$  and  $\bar{\ell}$  are unifiable in  $\psi$ . We will show that  $\ell^\dagger$  and  $\bar{\ell}^\dagger$  are unifiable in  $\psi \setminus \{\eta_7\}$  (note that  $\bar{\ell}^\dagger \in \mathcal{R}^\dagger(\eta_\ell)$ ).

Let  $\sigma_0$  be a most general unifier of  $\ell$  and  $\bar{\ell}$  in  $\psi$  (note that  $\sigma_0$  exists since  $\eta$  satisfies the pre-regularization property). If  $\bar{\ell} \notin \eta_2$  (i.e.  $\bar{\ell}$  was in some node  $\eta_2 \neq \eta'_2$ ), then we are done, as removing  $\eta_2$  does not apply a substitution to  $\eta_1$ , and thus  $\ell^\dagger = \ell$  and  $\bar{\ell}^\dagger = \bar{\ell}$ , so that  $\ell^\dagger$  and  $\bar{\ell}^\dagger$  are unifiable with  $\sigma_0$ .

Thus we may assume that  $\bar{\ell} \in \eta_2$ . Recall that  $\eta$  satisfies the regularization property and by assumption,  $\sigma\eta_1 \subseteq \mathcal{S}(\eta)$ . Therefore there exists some literal  $\bar{\ell}'$  below  $\eta$  unifiable with  $\ell$ , as otherwise  $\ell \notin \mathcal{S}(\eta)$  (in particular, the resolved literal of node  $\eta_\ell$ ). Note that  $\bar{\ell}' = \bar{\ell}^\dagger$  as it corresponds to the opposite literal of  $\ell$  in  $\psi \setminus \{\eta_2\}$ . Furthermore, as  $\bar{\ell}'$  is in  $\psi$ ,  $\bar{\ell}' \in \mathcal{R}(\eta)$ . Thus, since  $\eta$  satisfies the pre-regularization property  $\bar{\ell}'$  is unifiable with  $\bar{\ell}$  by some  $\sigma'$ . Thus  $\sigma'\bar{\ell}' = \bar{\ell}$ , and we have that  $\sigma_0\ell = \sigma'\bar{\ell}' = \sigma'\bar{\ell}^\dagger$  so that  $\sigma'\sigma_0\ell = \bar{\ell}^\dagger$ , as required.  $\square$

## 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 4.2). 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.

```

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

```

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  $\ell\sigma = \ell_R$  or  $\ell = \ell_R\sigma$  then
2   if  $\exists \sigma'$  such that  $\psi_R\sigma' \subseteq \psi.\text{safeLiterals}$  then
3     mark  $\psi_L$  as deletedNode ;
4     mark  $\psi$  as regularized
5 else if  $\exists \sigma$  and  $\ell \in \psi.\text{safeLiterals}$  such that  $\ell\sigma = \ell_L$  or  $\ell = \ell_L\sigma$  then
6   if  $\exists \sigma'$  such that  $\psi_L\sigma' \subseteq \psi.\text{safeLiterals}$  then
7     mark  $\psi_R$  as deletedNode ;
8     mark  $\psi$  as regularized

```

**Algorithm 2:** FRegularizeIfPossible

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 top-down 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 version of FORPI has been implemented in the functional programming language Scala as part of the Skeptik library. This library includes an implementation of GFOLU [8]. In order to evaluate the algorithm's effectiveness, FORPI was tested on two data sets: proofs generated by a real theorem prover and randomly-generated resolution proofs. The proofs are included in the source code repository, available at <https://github.com/jgorzny/Skeptik>.

```

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       safeLiteralsFrom( $\psi'$ )  $\leftarrow \psi'.$ safeLiterals ;
7     else if  $\psi' = \psi \odot_{\ell_L \ell_R}^{\sigma_L \sigma_R} \psi_R$  for some  $\psi_R$  then
8       safeLiteralsFrom( $\psi'$ )  $\leftarrow \psi'.$ safeLiterals  $\cup \{ \ell_R \sigma_R \}$ 
9     else if  $\psi' = \psi_L \odot_{\ell_L \ell_R}^{\sigma_L \sigma_R} \psi$  for some  $\psi_L$  then
10      safeLiteralsFrom( $\psi'$ )  $\leftarrow \psi'.$ safeLiterals  $\cup \{ \ell_L \sigma_L \}$ 
11  $\psi.\text{safeLiterals} \leftarrow \bigcap_{\psi' \in \psi.\text{children}} \text{safeLiteralsFrom}(\psi')$ 

```

**Algorithm 3:** F0setSafeLiterals

Note that by implementing the algorithms in this library, we are able to guarantee the correctness of the compressed proofs, as in Skeptik every inference rule (e.g. resolution, contraction) is implemented as a small class with a constructor that checks whether the conditions for the application of the rule are met, thereby preventing the creation of objects representing incorrect proof nodes (i.e. unsound inferences). With logical soundness guaranteed by Skeptik's data structures, to ensure the soundness of a proof compression algorithm, we only need to check that the root clause of the compressed proof is equal to or stronger than the root clause of the input proof and that the set of axioms used in the compressed proof is a (possibly non-proper) subset of the set of axioms used in the input proof.

First, FORPI was evaluated on the same proofs used to evaluate GFOLU. This data was generated by executing the SPASS (<http://www.spass-prover.org/>) theorem prover 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 originally generated on the Euler Cluster at the University of Victoria with a time limit of 300 seconds per problem. Under these conditions, SPASS generated 308 proofs. The proofs generated by SPASS were small: proof lengths varied from 3 to 49, and the number of resolutions in a proof ranged from 1 to 32.

In order to test FORPI's effectiveness on larger proofs, a total of 2280 proofs were randomly generated and then used as a second benchmark set. The randomly generated proofs were much larger than those

TODO:  
how  
small?

of the first data set: proof lengths varied from 95 to 700, while the number of resolutions in a proof ranged from 48 to 368. Proofs were generated by the following procedure: start with a root node whose conclusion is  $\perp$ , and make two premises  $\eta_1$  and  $\eta_2$  using a randomly generated literal such that the desired conclusion is the result of resolving  $\eta_1$  and  $\eta_2$ . For each node  $\eta_i$ , determine the inference rule used to make its conclusion: with probability  $p = 0.9$ ,  $\eta_i$  is the result of a resolution, otherwise it is the result of a contraction. Literals are generated by uniformly choosing a number from  $\{1, \dots, k, k+1\}$  where  $k$  is the number of predicates generated so far; if the chosen number  $j$  is between 1 and  $k$ , the  $j$ -th predicate is used; otherwise, if the chosen number is  $k+1$ , a new predicate with a new random arity (at most four) is generated and used. Each argument is a constant with probability  $p = 0.7$  and a complex term (i.e. a function applied to other terms) otherwise; functions are generated similarly to predicates. If a node  $\eta$  should be the result of a resolution, then with probability  $p = 0.2$  we generate a left parent  $\eta_\ell$  and a right parent  $\eta_r$  for  $\eta$  (i.e.  $\eta = \eta_\ell \odot \eta_r$ ) having a common parent  $\eta_c$  (i.e.  $\eta_\ell = (\eta_\ell)_\ell \odot \eta_c$  and  $\eta_r = \eta_c \odot (\eta_r)_r$ ), for some newly generated nodes  $(\eta_\ell)_\ell$  and  $(\eta_r)_r$ . The common parent ensures that also non-tree-like DAG proofs are generated. This procedure is recursively applied to the generated parent nodes. Each parent of a resolution has each of its constants and functions (that are not contained in the pivot literal) replaced by a fresh variable with probability  $p = 0.7$ . At each recursive call, the additional minimum height required for the remainder of the branch is decreased by one with probability  $p = 0.5$ . Thus if each branch always decreases the additional required height, the proof has height equal to the initial minimum value. The process stops when every branch is required to add a subproof of height zero, or after a timeout. The end of each branch is taken as an axiom. The minimum height was set to 7 (which is the minimum number of nodes in an irregular proof plus one) and the timeout was set to 300 seconds (the same timeout allowed for SPASS).

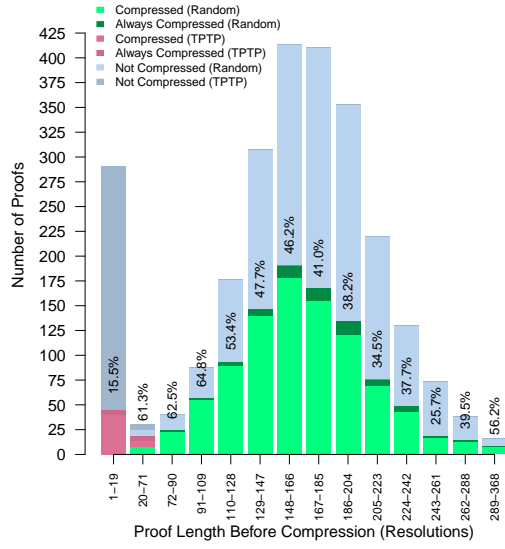
For consistency, the same system and metrics were used. Proof compression and proof generation was performed on a laptop (2.8GHz Intel Core i7 processor with 4GB of RAM (1333MHz DDR3) available to the Java Virtual Machine). For each proof  $\psi$ , we measured the time needed to compress the proof ( $t(\psi)$ ) and the compression ratio  $((|\psi| - |\alpha(\psi)|)/|\psi|)$  where  $|\psi|$  is the number of resolutions in the proof, and  $\alpha(\psi)$  is the result of applying a compression algorithm or some com-

position of FORPI and GFOLU. Note that we consider only the number of resolutions in order to compare the results of these algorithms to their propositional variants (where contractions are implicit). Moreover, contractions could be made implicit within resolution inferences even in the first-order case and we use explicit contractions only for technical convenience.

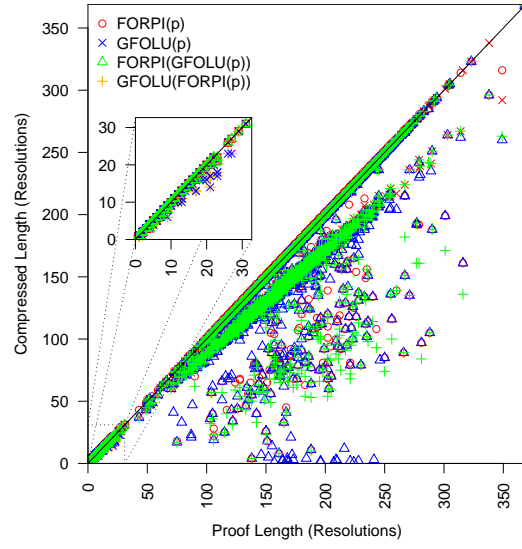
Table 1 summarizes the results of FORPI and its combinations with GFOLU. The first set of columns describes the percentage of proofs that were compressed by each compression algorithm. The algorithm ‘Best’ runs both combinations of GFOLU and FORPI and returns the shortest proof output by either of them. The total number of proofs is  $308 + 2280 = 2588$  and the total number of resolution nodes is  $2,249 + 393,883 = 396,132$ . The percentages in the last three columns are computed by  $(\sum_{\psi \in \Psi} |\psi| - \sum_{\psi \in \Psi} |\alpha(\psi)|) / (\sum_{\psi \in \Psi} |\psi|)$  for each data set  $\Psi$  (TPTP, Random, or Both). The use of FORPI alongside GFOLU allows at least an additional 5% of proofs to be compressed. Furthermore, the use of both algorithms removes more than twice as many nodes than any single algorithm.

Table 2 compares the results of FORPI and its combinations with GFOLU with their propositional variants as evaluated in [4]. The first column describes the mean compression ratio for each algorithm including proofs that were not compressed by the algorithm, while the second column calculates the mean compression ratio considering only compressed proofs. It is unsurprising that the first column is lower than the propositional mean for each algorithm: there are stricter requirements to apply these algorithms to first-order proofs. In particular, additional properties must be satisfied before a unit can be lowered, or before a pivot can be recycled. On the other hand, when first-order proofs are compressed, the levels of compression are on par with or better than their propositional counterparts.

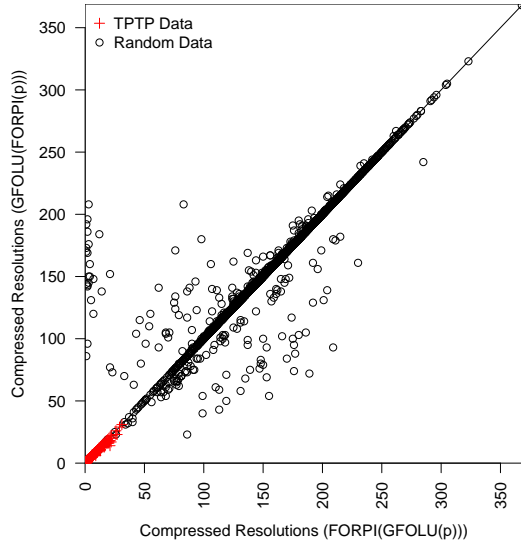
Figure 3 (a) shows the number of proofs (compressed and uncompressed) per grouping based on number of resolutions in the proof. The red (resp. dark grey) data shows the number of compressed (resp. uncompressed) proofs for the TPTP data set, while the green (resp. light grey) data shows the number of compressed (resp. uncompressed) proofs for the random proofs. The number of proofs in each group is the sum of the heights of each coloured bar in to that group. The overall percentage of proofs compressed in a group is indicated on each bar. Dark colors indicate the number of proofs compressed by FORPI, GFOLU, and both compositions of these algorithms; light colors indi-



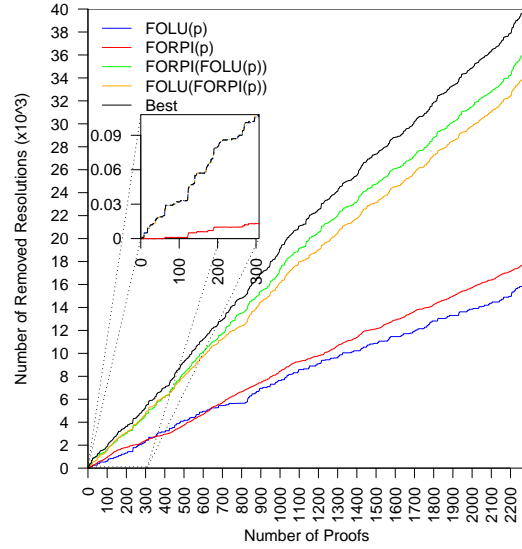
(a) Number of (non)-compressed proofs



(b) Compressed length against input length



(c) FORPI (GFOLU (p)) vs. GFOLU (FORPI (p))



(d) Cumulative proof compression

Fig. 3. GFOLU &amp; FORPI Combination Results



Algorithm	# of Proofs Compressed			# of Removed Nodes		
	TPTP	Random	Both	TPTP	Random	Both
GFOLU(p)	55 (17.9%)	815 (35.7%)	870 (33.6%)	107 (4.8%)	17,730 (4.5%)	17,837 (4.5%)
FORPI(p)	11 (3.6%)	252 (11.1%)	263 (10.2%)	13 (0.6%)	15,913 (4.0%)	15,926 (4.0%)
GFOLU(FORPI(p))	55 (17.9%)	993 (43.6%)	1048 (40.5%)	108 (4.8%)	33,956 (9.1%)	34,064 (9.1%)
FORPI(GFOLU(p))	11 (3.6%)	993 (43.6%)	1004 (38.8%)	108 (4.8%)	36,070 (9.1%)	36,178 (9.1%)
Best	56 (18.2%)	993 (43.6%)	1049 (40.5%)	108 (4.8%)	39,742 (10.1%)	39,850 (10.1%)

Table 1

Number of proofs compressed and number of overall nodes removed

Algorithm	First-Order Compression		Algorithm	Propositional Compression [4]
	All	Compressed Only		
GFOLU(p)	3.4%	33.1%	LU(p)	7.5%
FORPI(p)	4.5%	13.4%	RPI(p)	17.8%
GFOLU(FORPI(p))	7.6%	19.7%	(LU(RPI(p)))	21.7%
FORPI(GFOLU(p))	8.1%	21.0%	(RPI(LU(p)))	22.0%
Best	9.2%	22.8%	Best	22.0%

Table 2

Mean compression results

cate cases were FORPI succeeded, but at least one of GFOLU or a combination of these algorithms achieved zero compression. Given the size of the TPTP proofs, it is unsurprising that few are compressed: small proofs are a priori less likely to contain irregularities. On the other hand, at least 25% of randomly generated proofs in each size group could be compressed.

Figure 3 (b) is a scatter plot comparing the number of resolutions of the input proof against the number of resolutions in the compressed proof for each algorithm. The results on the TPTP data are magnified in the sub-plot. For the randomly generated proofs (points outside of the sub-plot), it is often the case that the compressed proof is significantly shorter than the input proof. Interestingly, GFOLU appears to reduce the number of resolutions by a linear factor in many cases. This is likely due to a linear growth in the number of non-interacting irregularities (i.e. irregularities for which the lowered units share no common literals with any other sub-proofs), which leads to a linear number of nodes removed.

Figure 3 (c) is a scatter plot comparing the size of compression obtained by applying FORPI before GFOLU versus GFOLU before FORPI. Data obtained from the TPTP data set is marked in red; the remaining points are obtained from randomly generated proofs. Points that lie on the diagonal line have the same size after each combination. There are 165 points beneath the line and 258 points above the line. Therefore, as in the propositional case [7], it is not a pri-

ori clear which combination is better. Nevertheless, the distinctly greater number of points above the line suggests that it is more often the case that FORPI should be applied after GFOLU. Not only this combination is more likely to maximize the likelihood of compression, but the achieved compression also tends to be larger.

Figure 3 (d) shows a plot comparing the difference between the cumulative number of resolutions of the first  $x$  input proofs and the cumulative number of resolutions in the first  $x$  proofs after compression (i.e. the cumulative number of *removed* resolutions). The TPTP data is displayed in the sub-plot; note that the lines for everything except FORPI largely overlap (since the values are almost identical; cf. Table 1). Observe that the use of both algorithms is always better than using a single algorithm. The data also shows that using FORPI after GFOLU is normally the preferred order of composition, as it typically results in a greater number of nodes removed than the other combination. An even better approach is to try both combinations and choose the best result (as shown in the ‘Best’ curve).

SPASS required approximately 40 minutes (running on a cluster and including proof generation time for each problem) to generate all the 308 TPTP proofs. The total time to apply both FORPI and GFOLU on all these proofs was just over 8 seconds on a simple laptop computer. The random proofs were generated in 70 minutes, and took approximately 453 seconds (or

7.5 minutes) to compress, both measured on the same computer. All times include parsing time. These compression algorithms continue to be very fast, and may simplify the proof considerably for a relatively small cost in time.

## 7. Conclusions and Future Work

The main contribution of this paper is the lifting of the propositional proof compression algorithm  $\text{RPI}$  to the first-order case. As indicated in Section 4, the generalization is challenging, because unification instantiates literals and, consequently, a node may be regularizable even if its resolved literals are not syntactically equal to any safe literal. Therefore, unification must be taken into account when collecting safe literals and marking nodes for deletion.

We first evaluated the algorithm on all 308 real proofs that the  $\text{SPASS}$  theorem prover (with only standard resolution enabled) was capable of generating when executed on unsatisfiable TPTP problems without equality. Although the compression achieved by the first-order  $\text{FORPI}$  algorithm was not as good as the compression achieved by the propositional  $\text{RPI}$  algorithm on real proofs generated by SAT and SMT solvers [7], this is due to the fact that the 308 proofs were too short (less than 32 resolutions) to contain a significant amount of irregularities. In contrast, the propositional proofs used in the evaluation of the propositional  $\text{RPI}$  algorithm had thousands (and sometimes hundreds of thousands) of resolutions.

Our second evaluation used larger, but randomly generated, proofs. The compression achieved by  $\text{FORPI}$  in a short amount of time on this data set was compatible with our expectations and previous experience in the propositional level. The obtained results indicate that  $\text{FORPI}$  is a promising compression technique to be reconsidered when first-order theorem provers become capable of producing larger proofs. It is important to note that randomly generated proofs may differ from real proofs in shape and may be more or less likely to contain irregularities exploitable by our algorithm. Resolution restrictions and refinements (e.g. ordered resolution [?], hyper-resolution [?], unit-resulting resolution [?]) tend to result in longer chains of resolutions and, therefore, in proofs with a possibly larger height to length ratio. As the number of irregularities increases with height, such proofs would have a higher number of irregularities in relation to length.

In this paper, for the sake of simplicity, we considered a pure resolution calculus without restrictions, refinements or extensions. However, in practice, theorem provers do use restrictions and extensions. It is conceptually easy to adapt the algorithm described here to such variations of resolution. For instance, restricted forms of resolution (e.g. ordered resolution, hyper-resolution, unit-resulting resolution) can be simply regarded as chains of unrestricted resolutions for the purpose of proof compression. The compression process would break the chains and change the structure of the proof, but the compressed proof would still be a correct unrestricted resolution proof, albeit not necessarily satisfying the restrictions that the input proof satisfied. In the case of extensions for equality reasoning using paramodulation-like inferences, it might be necessary to apply the paramodulations to the corresponding safe literals. Another common extension of resolution is the splitting technique [?]. When splitting is used, each split sub-problem is solved by a separate refutation, and the compression algorithm described here could be applied to each refutation independently.

## Acknowledgements

We thank the Google Summer of Code 2014 program for financial support of this research.

## References

- [1] *Logic for Programming, Artificial Intelligence, and Reasoning 16th International Conference, Dakar, Senegal, Revised Selected Papers*, LNCS. Springer, 2010.
- [2] H. Amjad. Compressing propositional refutations. *Electronic Notes in Theoretical Computer Science*, 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 Verification Conference*, LNCS, pages 114–128. Springer, 2008.
- [4] J. Boudou and B. Woltzenlogel Paleo. Compression of propositional resolution proofs by lowering subproofs. In *Automated Reasoning with Analytic Tableaux and Related Methods - 22th International Conference*, LNCS, pages 59–73. Springer, 2013.
- [5] S. Cotton. Two techniques for minimizing resolution proofs. In Ofer Strichman and Stefan Szeider, editors, *SAT 2010*, LNCS, pages 306–312. Springer, 2010.
- [6] P. Fontaine, S. Merz, and B. Woltzenlogel Paleo. Exploring and exploiting algebraic and graphical properties of resolution. In *8th International Workshop on SMT*, 2010.
- [7] P. Fontaine, S. Merz, and B. Woltzenlogel Paleo. Compression of propositional resolution proofs via partial regularization. In *Automated Deduction - CADE-23 - 23rd International Confer-*

- ence on Automated Deduction, Wroclaw, Poland, July 31 - August 5, 2011. *Proceedings*, LNCS, pages 237–251. Springer, 2011.
- [8] J. Gorzny and B. Woltzenlogel Paleo. Towards the compression of first-order resolution proofs by lowering unit clauses. In *Automated Deduction - CADE-25 - 25th International Conference on Automated Deduction, Berlin, Germany, August 1-7, 2015. Proceedings*, 2015.
- [9] S. Hetzl, A. Leitsch, G. Reis, and D. Weller. Algorithmic introduction of quantified cuts. *Theoretical Computer Science*, 549:1–16, 2014.
- [10] B. Woltzenlogel Paleo. Atomic cut introduction by resolution: Proof structuring and compression. In *LPAR-16* [1], pages 463–480.
- [11] 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.
- [12] S. Schulz and G. Sutcliffe. Proof generation for saturating first-order theorem provers. In D. Delahaye and B. Woltzenlogel Paleo, editors, *All about Proofs, Proofs for All*, volume 55 of *Mathematical Logic and Foundations*. College Publications, London, UK, 2015.
- [13] C. Sinz. Compressing propositional proofs by common subproof extraction. In R. Moreno-Díaz, F. Pichler, and A. Quesada-Arencibia, editors, *EUROCAST*, LNCS, pages 547–555. Springer, 2007.
- [14] 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.
- [15] 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*. Springer-Verlag, 1983.
- [16] J. Vyskocil, D. Stanovský, and J. Urban. Automated proof compression by invention of new definitions. In *LPAR* [1], pages 447–462.

## Appendix A. Algorithm

### RecyclePivotsWithIntersection

**Note:** for the reviewers' convenience, this appendix summarizes [7].

RecyclePivotsWithIntersection (RPI) [7] 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 [15], 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 marked as `deletedNode`, as shown in Algorithm 5. Only later

<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     <math>\text{setSafeLiterals}(\eta)</math>; 5     <math>\text{regularizeIfPossible}(\eta)</math> 6 <math>\psi' \leftarrow \text{fix}(\psi')</math>; 7 <b>return</b> <math>\psi'</math>; </pre>
--

**Algorithm 4:** RPI

during fixing, the irregular node is actually replaced by its surviving parent (i.e. the parent that is not marked as `deletedNode`).

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   mark left parent of  $\eta$  as deletedNode ;
3   mark  $\eta$  as regularized
4 else if  $\eta$ .leftResolvedLiteral  $\in \eta$ .safeLiterals then
5   mark right parent of  $\eta$  as deletedNode ;
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