Abstract

This work is about compression of formal proofs. Formal proofs are of great importance to modern computer scienece. They can be used to combine deductive systems. For example SAT- Solvers [9] are heavily used for all kinds of computations, because of their efficiency. A formal proof is a certificate of the correctness of the output of a SAT- Solver. Furthermore, from formal proofs information can be extracted about some underlying problem. For example, Interpolants [29] can be extracted from proofs, as done in [26].

Typically problems tackled by automated systems are huge. Therefore the produced proofs are huge. For example, [28] reports about a 13 GB proof of one case of the Erdős Discrepancy Conjecture. With such proof sizes, computer system reach their boundaries and that is why it is necessary to compress proofs. Our work presents two methods for proof compression.

The first method removes redundancies in the congruence part of SMT- proofs. Congruence reasoning deduces equations from a set of given equations, using the four axioms *reflexivity*, *symmetry*, *transitivity*, and *congruence*. We found that SMT- Solver often use an unnecessarily big set of input equations to deduce one particular equality. We want to find smaller sets of equations, that suffice to proof the same result and therefore replace subproofs with shorter ones. Furthermore, we will proof the NP - Completeness of the problem of finding the shortest explanation of one equation within a set of input equations.

The second method investigates the memory requirements of proofs. While processing a proof, not all parts of the proof have to be kept in memory at all times. Subproof can be loaded into memory when needed and can be removed from memory again when they are not. In which traversal order subproofs are visited is essential to the maximum memory consumption during proof processing. We want to construct traversal orders with low memory requirements using heuristics.

Kurzfassung

Diese Arbeit befasst sich mit der Komprimierung von formalen Beweisen. Formale Beweise sind von großer Bedeutung in der modernen Informatik. Sie können verwendet werden um deduktive Systeme miteinander zu kombinieren. Ein Beispiel sind SAT- Solver [9], welche ob ihrer Effektivität gerne für diverse Berechnungen verwendet werden. Ein formaler Beweis kann als Zertifikat für die Korrektheit des Ergebnisses eines SAT- Solvers dienen. Des Weiteren können aus ihnen Informationen, wie etwa Interpolants [29], extrahiert werden, welche zur Lösung eines Problems beitragen [26].

Formale Beweise sind typischerweise sehr groß, siehe etwa [28] für einen 13GB Beweis eines Falles der Erdős Discrepancy Conjecture. Bei solchen Beweisgrößen stoßen Computersysteme an ihre Grenzen und deswegen ist es erforderlich Beweise zu komprimieren. Unsere Arbeit präsentiert zwei Methoden zur Beweiskomprimierung.

Die erste Methode entfernt Redundanzen im Kongruenzteil von SMT-Beweisen. Kongruenzbeweise schließen von einer Menge an Gleichungen auf neue Gleichungen mit der Vorraussetzung der vier Axiome: *Reflexivität*, *Symmetrie*, *Transitivität* und *Kongruenz*. Beweise, die von SMT-Solvern erzeugt werden, schließen oft auf neue Gleichungen aus einer unnötig großen Menge. Wir wollen kleinere Mengen finden, die für den Beweis der selben Aussage ausreichen und somit redundante Beweise durch kürzere ersetzen. Außerdem werden wir die NP - Completeness des Problems der kürzesten Erklärung einer Gleichung beweisen.

Die zweite Methode untersucht die Speicherplatzanforderungen von Beweisen. Beim Bearbeiten von Beweisen muss nicht der gesamte Beweis zu jeder Zeit im Speicher gelagert werden. Teilbeweise werden erst in den Speicher geladen, wenn sie benötigt werden und werden wieder aus diesem entfernt, sobald sie nicht mehr benötigt werden. In welcher Ordnung die Teilbeweise geladen werden, ist essentiell für die maximale Speicherplatzanforderung. Wir wollen Ordnungen mit niedrigen Speicherplatzanforderungen mit Hilfe von Heuristiken konstruieren.

General Information

This document is intended as a template and guideline and should support the author in the course of doing the master's thesis. Assessment criteria comprise the quality of the theoretical and/or practical work as well as structure, content and wording of the written master's thesis. Careful attention should be given to the basics of scientific work (e.g., correct citation).

Organizational Issues

A master's thesis at the Faculty of Informatics has to be finished within six months. During this period regular meetings between the advisor(s) and the author have to take place. In addition, the following milestones have to be fulfilled:

1. Within one month after having fixed the topic of the thesis the master's thesis proposal has to be prepared and must be accepted by the advisor(s). The master's thesis proposal must follow the respective template of the dean of academic affairs. Thereafter the proposal has to be applied for at the deanery. The necessary forms may be found on the web site of the Faculty of Informatics. `http://www.informatik.tuwien.ac.at/dekanat/formulare.html`

2. Accompanied with the master's thesis proposal, the structure of the thesis in terms of a table of contents has to be provided.

3. Then, the first talk has to be given at the so-called "Seminar for Master Students". The slides have to be discussed with the advisor(s) one week in advance. Attendance of the "Seminar for Master Students" is compulsory and offers the opportunity to discuss arising problems among other master students.

4. At the latest five months after the beginning, a provisional final version of the thesis has to be handed over to the advisor(s).

5. As soon as the provisional final version exists, a first poster draft has to be made. The making of a poster is a compulsory part of the "Seminar for Master Students" for all master studies at the Faculty of Informatics. Drafts and design guidelines can be found at `http://www.informatik.tuwien.ac.at/studium/richtlinien`.

6. After having consulted the advisor(s) the second talk has to be held at the "Seminar for Master Students".

7. At the latest six months after the beginning, the corrected version of the master's thesis and the poster have to be handed over to the advisor(s).

8. After completion the master's thesis has to be presented at the "epilog". For detailed information on the epilog see:
   `http://www.informatik.tuwien.ac.at/studium/epilog`

Structure of the Master's Thesis

If the curriculum regulates the language of the master's thesis to be English (like for "Business Informatics"), the thesis has to be written in English. Otherwise, the master's thesis may be written in English or in German. The structure of the thesis is predetermined. The table of contents is followed by the introduction and the main part, which can vary according to the content. The master's thesis ends with the bibliography (compulsory) and the appendix (optional).

- Cover page
- Acknowledgements
- Abstract of the thesis in English and German
- Table of contents
- Introduction
  - motivation
  - problem statement (which problem should be solved?)
  - aim of the work
  - methodological approach
  - structure of the work
- State of the art / analysis of existing approaches
  - literature studies
  - analysis
  - comparison and summary of existing approaches
- Methodology
  - used concepts
  - methods and/or models
  - languages
  - design methods
  - data models
  - analysis methods
  - formalisms
- Suggested solution/implementation
- Critical reflection
  - comparison with related work
  - discussion of open issues
- Summary and future work
- Appendix: source code, data models, . . .
- Bibliography

Congruence Resolution

The proofs considered in this work are resolution proofs [**?**] extended by the axioms of congruence. In this chapter we define the calculus

Let $\mathcal{C}$ be a finite set of constant symbols and let f be a binary function symbol. The set of terms $\mathcal{T}$ is defined recursively.

$$\begin{aligned}
\mathcal{T}_0 &:= \mathcal{C} \\
\mathcal{T}_i &:= \mathcal{T}_{i-1} \cup \{f(t_1, t_2) \mid t_1, t_2 \in \mathcal{T}_{i-1}\} \\
\mathcal{T} &:= \bigcup_{n \in \mathbb{N}} \mathcal{T}_n
\end{aligned}$$

Let $\mathcal{Q}_{\mathcal{T}} = \{t_1 = t_2 \mid t_1, t_2 \in \mathcal{T}$ be the set of equations for a set of terms $\mathcal{T}$ Let V be a finite set of propositional variables. The set of equality atoms $\mathcal{E}$ is defined as $V \cup Q_{\mathcal{T}}$ An equality literal $\ell_{\mathcal{T}}$ is an equality atom e or a negated equality atom $\neg e$. We will abbreviate write $\neg(t_1 = t_2)$ by $t_1 \neq t_2$. An equality clause is a set of equality literals. As usual a clause is interpreted as the disjunction of its literals and a set of clauses is interpreted as the conjunction its clauses.

The axioms of congruence $EqAxioms$ for some set of terms $\mathcal{T}$ is defined as $R \cup S \cup T \cup C$ where

$$\begin{aligned}
R &= \{t = t \mid t \in \mathcal{T}\} \\
S &= \{t_1 \neq t_2, t_2 = t_1 \mid t_1, t_2 \in \mathcal{T}\} \\
T &= \{t_1 \neq t_2, t_2 \neq t_3, t_1 = t_3 \mid t_1, t_2, t_3 \in \mathcal{T}\} \\
C &= \{t_1 \neq t_3, t_2 \neq t_4, f(t_1, t_2) = f(t_3, t_4) \mid t_1, t_2, t_3, t_4 \in \mathcal{T}\}
\end{aligned}$$

Note that every congruence axiom has one positive equality literal. From now on we will omit the set of terms $\mathcal{T}$ if it is clear from context.

Next we will define the resolution calculus extended by congruence axioms. Let ℓ be an equality literal and C_1, C_2 be equality clauses such that $\ell \in C_1$ and $\neg\ell \in C_2$. The clause $C_1 \setminus \{\ell\} \cup C_2 \setminus \{\neg\ell\}$ is the resolvent of C_1 and C_2 with pivot ℓ.

Let $F = \{C_1, \ldots, C_n\}$ be a set of clauses. The notion of a congruence derivation for F is defined inductively. The sequence $\langle C_1, \ldots, C_n\rangle$ is a congruence derivation for F. If $\langle C_1, \ldots, C_m\rangle$ is a congruence derivation for F then $\langle C_1, \ldots, C_{m+1}\rangle$ is a congruence derivation for F if $C_{m+1} \in EqAxioms$ or C_{m+1} is a resolvent of C_i and C_j with $1 \leq i, j \leq m$. A congruence derivation containing the empty clause is a congruence refutation.

Let $D = \langle C_1, \ldots, C_m\rangle$ be a congruence derivation. The longest subsequence $\langle C_{i_1}, \ldots, C_{i_k}\rangle$ of D, such that $\{C_{i_1}, \ldots, C_{i_k}\} \subseteq EqAxioms$ is called the equality reasoning part of D.

Pebbling

Introduction

Proofs generated by SAT-solvers can be huge. Checking their correctness can not only take a long time but also consume a lot of memory. In an ongoing project for controller synthesis based on the extraction of interpolants from SMT-proofs [26], for example, post-processing a proof takes hours and may reach the limit of memory available today in a single node of a computer cluster (256GB). This issue is even more relevant in application scenarios in which the proof consumer, who is interested in independently checking the correctness of the proof, might have less available memory than the proof producer. This is in part because, while the proof checker reads a usual proof file and checks the proof it contains, every proof node (containing a clause) that is loaded into memory has to be kept there until the end of the whole proof checking process, since the proof checker does not know whether a proof node will still need to be used and re-reading the proof file to reload and recheck proof nodes would be too time-consuming.

To address this issue, recently proposed proof formats such as DRUP [24] and BDRUP [25] allow enriching a proof file with instructions that inform a proof checker when a proof node can be released from memory. Other proof formats, such as the TraceCheck format [7] could also be enriched analogously. Such node deletion instructions can be added by a proof-generating SAT-solver during proof search in the periodic clean-up of its database of derived learned clauses; for every clause the SAT-solver deletes during this phase, this deletion can be recorded in the proof file.

This paper explores the possibility of post-processing a proof in order to increase the amount of deletion instructions in the proof file. The more deletion instructions, the less memory the proof checker will need. Therefore, this *deletion-during-proof-postprocessing* approach ought to be seen not as a replacement but rather as an independent complement to the *deletion-during-proof-search* already performed by state-of-the-art proof-generating SAT-solvers.

The new methods proposed here exploit an analogy between proof checking and playing *Pebbling Games* [22,27]. The particular version of pebbling game relevant for proof checking is defined precisely in Section and the analogy to proof checking is explained in detail in Section . The proposed pebbling algorithms are greedy (Section) and based on heuristics (Section 7). As discussed in Sections and , approaches based on exhaustive enumeration or on encoding as a SAT problem would not fare well in practice.

The proof space compression algorithms described here are not restricted to proofs generated by SAT-solvers. They are general DAG pebbling algorithms, that could be applied to proofs represented in any calculus where proofs are directed acyclic graphs (including the special case of tree-like proofs). It is, nevertheless, in SAT and SMT that proofs tend to be largest and in most need of space compression. The underlying propositional resolution calculus (described in Section) satisfies the DAG requirement. The experiments (Section 7) evaluate the proposed algorithms on thousands of SAT- and SMT-proofs.

Propositional Resolution Calculus

A *literal* is a propositional variable or the negation of a propositional variable. The *complement* of a literal ℓ is denoted $\overline{\ell}$ (i.e. for any propositional variable p, $\overline{p} = \neg p$ and $\overline{\neg p} = p$). The set of all literals is denoted by $\mathcal{L}$. A *clause* is a set of literals. $\perp$ denotes the *empty clause*.

Definition 0.0.1 (Proof)**.** A *proof* φ is a tuple $\langle V, E, v, \Gamma \rangle$, such that $\langle V, E \rangle$ is a labeled directed acyclic graph whose edges are labeled with literals $\ell \in \mathcal{L}$, i.e. $E \subseteq V \times \mathcal{L} \times V$, $v \in V$, Γ is a clause and one of the following holds:

1. $V = \{v\}, E = \emptyset$
2. There are proofs $\varphi_L = \langle V_L, E_L, v_L, \Gamma_L \rangle$ and $\varphi_R = \langle V_R, E_R, v_R, \Gamma_R \rangle$ such that there exists a literal ℓ such that $\overline{\ell} \in \Gamma_L$, $\ell \in \Gamma_R$, and let $v \notin (V_L \cup V_R)$ then

$$
\begin{aligned}
V &= (V_L \cup V_R) \cup \{v\} \\
E &= E_L \cup E_R \cup \left\{ v_L \xrightarrow{\overline{\ell}} v, v_R \xrightarrow{\ell} v \right\} \\
\Gamma &= \left(\Gamma_L \setminus \{\overline{\ell}\}\right) \cup (\Gamma_R \setminus \{\ell\})
\end{aligned}
$$

v is called the *root* of φ and Γ its *conclusion*. φ_L and φ_R are *premises* of φ and φ is a *child* of φ_L and φ_R. Γ is called the *resolvent* of Γ_L and Γ_R with *pivot* ℓ. A proof ψ is a subproof of a proof φ, if there is a path from φ to ψ in the transitive closure of the premise relation. A subproof ψ of φ which has no premises is an *axiom* of φ. V_φ and A_φ denote, respectively, the set of nodes and axioms of φ. P_v^φ denotes the premises and C_v^φ the children of the subproof with root v in a proof φ. When a proof is represented graphically, the root is drawn at the bottom and the axioms at the top. The *length* of a proof φ is the number of nodes in V_φ and is denoted by $l(\varphi)$. □

Note that in case 2 of Definition 0.0.1 V_L and V_R are not required to be disjunct. Therefore the underlying structure of a proof is a directed acyclic graph and not simply a tree. Modern SAT- and SMT-solvers, using techniques of conflict driven clause learning, produce proofs with a general DAG structure [9,10]. The reuse of proof nodes plays a central role in proof compression [20]. Also note that a DAG corresponding to a proof has exactly one sink, which is called its root node, by definition.
For example a node of a proof $\langle V, E, v, \Gamma \rangle$ will be meant to be some $s \in V$. From hereon, if free from ambiguity, proofs and their underlying DAGs will not be distinguished.

Pebbling Game

Pebbling games are played on directed acyclic graphs and pebbles are placed on nodes following the rules of the game. The goal is to put a pebble on some target node. Pebbling games were introduced in the 1970's to model programming language expressiveness [35, 42] and compiler construction [38]. More recently, pebbling games have been used to investigate various questions in parallel complexity [11] and proof complexity [6,18,33]. They are used to obtain bounds for space and time requirements and trade-offs between the two measures [5,41]. Space requirements are modeled with the number of pebbles used. Time requirements are reflected by the number of rounds played. From hereon *to pebble* means to mark a node with a pebble and *to unpebble* means to remove the mark off a node.

Definition 0.0.2 (Bounded Pebbling Game)**.** The *Bounded Pebbling Game* is played by one player on a DAG $G = (V, E)$ with one distinguished node $s \in V$. The goal of the game is to pebble s, respecting the following rules:

1. A node v is pebbleable *iff* all predecessors of v in G are pebbled and v is currently not pebbled.

2. Pebbled nodes can be unpebbled at any time.

3. Once a node has been unpebbled, it may not be pebbled in a later round.

The game is played in rounds. Every round the player chooses a node $v \in V$, such that v is pebbled or pebbleable. The move of the player in this round is $p(v)$, if v is pebbleable and $u(v)$ if v is pebbled, where $p(.)$ and $u(.)$ correspond to pebbling and unpebbling a node respectively. □

Not that due to rule 1 the move in each round is uniquely defined by the chosen node v. The distinction of the two kinds of moves is just made for presentation purposes. Also note that as a consequence of rule 1, pebbles can be put on nodes without predecessors at any time. Playing the game on a proof φ means to play the game on the underlying DAG with the distinguished node being the root of φ.

In this work we investigate space requirements when time requirements are fixed. Fixing time is a design choice, see Section , and it corresponds to rule 3. Including this rules sets a bound $O(|V|)$ for the number of rounds.

Definition 0.0.3 (Strategy)**.** A *pebbling strategy* σ for the Bounded Pebbling Game, played on a DAG $G = (V, E)$ and distinguished node s, is a sequence of moves $(\sigma_1, \ldots, \sigma_n)$ of the player such that $\sigma_n = p(s)$.

Rules 2 and 3 The following definition allows to measure how many pebbles are required to play the Bounded Pebbling Game on a given graph.

Definition 0.0.4 (Pebbling number)**.** The *pebbling number of a pebbling strategy* $(\sigma_1, \ldots, \sigma_n)$ is $max_{i\in\{1\ldots n\}}|\{v \in V \mid v \text{ is pebbled in round } i\}|$. The *pebbling number of a DAG G and node s* is the minimum pebbling number of all pebbling strategies for G and s.

Note that Definitions 0.0.2 and 0.0.3 leave the player freedom when to do unpebbling moves. With the aim of finding strategies with low pebbling numbers, for every unpebbling move there is a canonical round make them, as will be shown in Section .

The Bounded Pebbling Game from definition 0.0.2 differs from the Black Pebbling Game discussed in [23, 36] in two aspects. Firstly, the Black Pebbling Game does not include rule 3. Excluding this rule allows for pebbling strategies with lower pebbling numbers ([38] has an example on page 1), at the expense of an exponential upper bound on the number of rounds [41]. Secondly, when pebbling a node in the Black Pebbling Game, one of its predecessors' pebbles can be used instead of a fresh pebble (i.e. a pebble can be moved). The trade-off between moving pebbles and using fresh ones is discussed in [41]. Deciding whether the pebbling number of a graph G and node s is smaller than k is PSPACE-complete in the absence of rule 3 [22] and NP-complete when rule 3 is included [38].

Pebbling and Proof Processing

The problem of processing a proof with minimal memory consumption is analogous to the problem of finding a pebbling strategy with minimal pebbling number. Proof processing could be checking its correctness, manipulating it or extracting information from it. The following definition makes the notion of proof processing formal.

Definition 0.0.5 (Proof Processing)**.** Let φ be a proof with nodes V and T be an arbitrary set. A function $f : V \times T \times T \to T$ is a *processing function* if there is a function $g_f : V \to T$ such that for every $v \in V$ with $P_v^\varphi = \emptyset$ (i.e. v represents an axiom), $g_f(v) = f(v, t_1, t_2)$ for all $\{t_1, t_2\} \subseteq T$. Let $\mathcal{F}$ be the set of processing functions. The *apply function* $\mathrm{ap} : V \times \mathcal{F} \to T$ is defined recursively as follows.

$$\mathrm{ap}(v, f) = \begin{cases} f(v, \mathrm{ap}(pr_1, f), \mathrm{ap}(pr_2, f)) & \text{if } v \text{ has premises } pr_1 \text{ and } pr_2 \\ g_f(v) & \text{otherwise} \end{cases}$$

Processing a node v with some processing function f means computing the value $\mathrm{ap}(v, f)$. *Processing a proof* means to process its root node. □

Example 0.0.1. Checking the correctness of a proof (i.e. checking for the absence of faulty resolution steps) can be checked in terms of the following processing function with $T = \{\top, \bot\}$ and $\wedge$ being the usual boolean and-operation.

$$f(v, w_1, w_2) = \begin{cases} \top & \text{if } v \text{ has no premises} \\ w_1 \wedge w_2 & \text{if the conclusion of } v \text{ is a resolvent} \\ & \quad \text{of the conclusions of its premises} \\ \bot & \text{otherwise} \end{cases}$$

Processing a proof with this processing function yields $\top$ *iff* the proof is a correct resolution proof. □

In Section it was pointed out that strategies with minimal pebbling numbers may require to play exponentially many rounds when rule 3 is not used. Every round of the game corresponds to an I/O operation and, if the action of the player is to pebble a node, the processing of the node. The goal of proof compression is to make proof processing less expensive, therefore requiring exponentially many I/O operations and processing steps is not a viable option. That is the reason why we chose the Bounded Pebbling Game for our purpose. In the Bounded Pebbling Game the number of rounds is linear in the number of nodes.

In order to process a node, the results of processing its premises are used and therefore have to be stored in memory. The requirement of having premises in memory corresponds to rule 1 of the Bounded Pebbling Game. Processing a node and I/O operations are typically more expensive than extra memory consumption, therefore in our setting every node can be processed only once, which corresponds to rule 3. A node that has been processed can be removed from memory, which corresponds to rule 2. Note that removing a node and its results too early in combination with 3 makes it impossible to process the whole proof. The optimal moment to remove a node from memory is uniquely determined by the order nodes are processed, see Theorem 0.0.1.

Definition 0.0.5 does not specify in what order to process nodes. The order in which nodes are processed is essential for the memory consumption, just like the order of pebbling nodes in the pebbling game is essential for the pebbling number. The following definition allows us to relate pebbling strategies with orderings of nodes.

Definition 0.0.6 (Topological Order)**.** A topological order of a proof φ is a total order relation $\prec$ on V_φ, such that for all $v \in V_\varphi$, for all $p \in P_v^\varphi : p \prec v$. A sequence of moves $(\sigma_1, \dots, \sigma_n)$ in the pebbling game *respects* a topological order $\prec$ if $j < i$ *iff* $\sigma_j \prec \sigma_i$. □

A topological order $\prec$ of a proof φ can be represented as a sequence $(v_1, \dots, v_n)$ of proof nodes, by defining $\prec := \{(v_i, v_j) \mid 1 \leq i < j \leq n\}$. The requirement that topological orders to order premises lower than their children corresponds to rule 1 of the Bounded Pebbling Game. The antisymmetry together with the fact that $V = \{v_1, \dots, v_n\}$ correspond to rule 3. Theorem 0.0.1 shows that the moments for unpebbling moves are predefined by the pebbling moves, when the goal is to find strategies with small pebbling numbers. Therefore there is a bijection between topological orders and pebbling strategies.

Definition 0.0.7 (Canonical Topological Pebbling Strategy)**.** The *canonical topological pebbling strategy* σ for a proof φ, its root node s and a topological order $\prec$ represented as a sequence $(v_1, \dots, v_n)$ is defined recursively:

$$\sigma_1 = p(v_1)$$
$$\sigma_i = \begin{cases} u(v) & \text{for all } c \in C_v^\varphi \text{ exists } k < i \text{ such that } \sigma_k = p(u) \\ p(v) & \text{otherwise, where } v = min_\prec(w \mid \text{ for all } l < i : \sigma_l \neq p(w)) \end{cases}$$

□

The following theorem shows that unpebbling moves can be omitted from strategies for the Bounded Pebbling Game, when the goal is to produce strategies with low pebbling numbers.

Theorem 0.0.1. *The canonical pebbling strategy has the minimum pebbling number among all pebbling strategies that respect the topological order $\prec$.*

Proof. Definition 0.0.7 prioritizes unpebbling over pebbling moves. Therefore the canonical topological pebbling strategy makes unpebbling moves as soon as possible. Consider the moment for unpebbling an arbitrary node v in the canonical pebbling strategy. Unpebbling it later could only possibly increase the pebble number. To reduce the pebble number, v would have to be unpebbled earlier than some preceding pebbling move. But, by definition of canonical pebbling strategy, the immediately preceding pebbling move pebbles the last child of v w.r.t. $\prec$. Therefore, unpebbling v earlier would make it impossible for its last child to be pebbled later without violating the rules of the game. □ □

As a consequence of Theorem 0.0.1 finding pebbling strategies with low pebbling numbers can be reduced to constructing topological orders. The memory required to process a proof using some topological order can be measured by the pebbling number of the canonical pebbling strategy corresponding to the order.

Definition 0.0.8 (Space)**.** The *space* $s(\varphi, \prec)$ of a proof φ and a topological order $\prec$ is the pebbling number of the canonical topological pebbling strategy of φ, its root and $\prec$. □

The problem of compressing the space of a proof φ and a topological order $\prec$ is the problem of finding another topological order $\prec'$ such that $s(\varphi, \prec') < s(\varphi, \prec)$. The following theorem shows that the number of possible topological orders is very large; hence, enumeration is not a feasible option when trying to find a good topological order.

Theorem 0.0.2. *There is a sequence of proofs $(\varphi_1, \ldots, \varphi_m, \ldots)$ such that $l(\varphi_m) \in O(m)$ and $|T(\varphi_m)| \in \Omega(m!)$, where $T(\varphi_m)$ is the set of possible topological orders for φ_m.*

Proof. Let φ_m be a perfect binary tree with m axioms. Clearly, $l(\varphi_m) = 2m - 1$. Let $(v_1, \ldots, v_n)$ be a topological order for φ_m. Let $A_\varphi = \{v_{k_1}, \ldots, v_{k_m}\}$, then $(v_{k_1}, \ldots, v_{k_m}, v_{l_1}, \ldots, v_{l_{n-m}})$, where $(l_1, \ldots, l_{n-m}) = (1, \ldots, n) \setminus (k_1, \ldots, k_m)$, is a topological order as well. Likewise, $(v_{\pi(k_1)}, \ldots, v_{\pi(k_m)}, v_{l_1}, \ldots, v_{l_{n-m}})$ is a topological order, for every permutation π of $\{k_1, \ldots, k_m\}$. There are $m!$ such permutations, so the overall number of topological orders is at least factorial in m (and also in n). □

Pebbling as a Satisfiability Problem

To find the pebble number of a proof, the question whether the proof can be pebbled using no more than k pebbles can be encoded as a propositional satisfiability problem. In this section let φ be a proof with nodes $v_1, \ldots, v_n$ and let v_n be its root node. Due to rule 3 of the Bounded Pebbling Game, the number of moves that pebble nodes is exactly n and due to theorem 0.0.1 determining the order of these moves is enough to define a strategy. For every $x \in \{1, \ldots, k\}$, every $j \in \{1, \ldots, n\}$ and every $t \in \{0, \ldots, n\}$ there is a propositional variable $p_{x,j,t}$. The variable $p_{x,j,t}$ being mapped to $\top$ by a valuation is interpreted as the fact that in the t'th round of the game node v_j is marked with pebble x. Round 0 is interpreted as the initial setting of the game before any move has been done.

Definition 0.0.9 (Pebbling SAT encoding)**.** The propositional formula obtained by conjuncting the following four constraints expresses the existence of a pebbling strategy for φ with pebbling number smaller or equal k.

1. The root is pebbled in the last round

$$\Psi_1 = \bigvee_{x=1}^{k} p_{x,n,n}$$

2. No node is pebbled initially

$$\Psi_2 = \bigwedge_{x=1}^{k} \bigwedge_{j=1}^{n} (\neg p_{x,j,0})$$

3. A pebble can only be on one node in one round

$$\Psi_3 = \bigwedge_{x=1}^{k} \bigwedge_{j=1}^{n} \bigwedge_{t=1}^{n} \left(p_{x,j,t} \rightarrow \bigwedge_{i=1, i \neq j}^{n} \neg p_{x,i,t} \right)$$

4. For pebbling a node, its premises have to be pebbled the round before and only one node is being pebbled each round.

$$\Psi_4 = \bigwedge_{x=1}^{k} \bigwedge_{j=1}^{n} \bigwedge_{t=1}^{n} \Bigg(\left(\neg p_{x,j,t} \wedge p_{x,j,(t+1)} \right) \rightarrow \left(\bigwedge_{i \in P_j^{\varphi}} \bigvee_{y=1, y \neq x}^{k} p_{y,i,t} \right) \wedge \left(\bigwedge_{i=1}^{n} \bigwedge_{y=1, y \neq x}^{k} \neg \left(\neg p_{y,i,t} \wedge p_{y,i,(t+1)} \right) \right) \Bigg)$$

The sets A_φ and P_j^φ are to be understood as sets of indices of the respective nodes.

This encoding is polynomial, both in n and k. However constraint 4 accounts to $O(n^3 * k^2)$ clauses. Even small resolution proofs have more than 1000 nodes and pebble numbers bigger than 100, which adds up to 10^{13} clauses for constraint 4 alone. Therefore, although theoretically possible to play the pebbling game via SAT-solving, this is practically infeasible for compressing proof space. The following theorem proves the correctness of the encoding.

Theorem 0.0.3 (Correctness of pebbling SAT encoding)**.** $\Psi = \Psi_1 \wedge \Psi_2 \wedge \Psi_3 \wedge \Psi_4$ *is satisfiable* iff *there exists a pebbling strategy using no more than k pebbles*

Proof. Suppose Ψ *is satisfiable* and let $\mathcal{I}$ be a satisfying variable assignment in form of the set of true variables. We will use $P(x, j, t)$ as an abbreviation for $p_{x,j,(t-1)} \notin \mathcal{I}$ and $p_{x,j,t} \in \mathcal{I}$. Since $\mathcal{I}$ satisfies Ψ_3, in $P(x, j, t)$ x is uniquely defined by j and t and we can write $P(j, t)$

instead. We will prove the following assertion. For every $t \in \{1,\dots,n\}$ there exists exactly one $j \in \{1,\dots,n\}$ such that $P(j,t)$.

Ψ_1 states that the root v_n has to be pebbled in the last round and Ψ_2 states that no node is pebbled initially. So for n there has to be a $t \in \{1,\dots,n\}$ such that $P(n,t)$. $\mathcal{I}$ satisfies Ψ_4, therefore for every predecessor of v_j of v_n there exists $x \in \{1,\dots,k\}$ such that $p_{x,j,(t-1)}$. Using the same argument for v_j as for v_n there has to be a $t' \in \{1,\dots,(t-1)\}$ such that $P(j,t')$. Every node of the proof is a recursive ancestor of the root, therefore for every $j \in \{1,\dots,n\}$ there exists at least one $t \in \{1,\dots,n\}$ such that $P(n,t)$. For every $t \in \{1,\dots,n\}$, Ψ_4 ensures that if $P(n,t)$ then there is no $i \in \{1,\dots,n\}, i \neq j$ such that $P(i,t)$, which proves the assertion. The assertion implies the existence of a bijection $\tau : \{1,\dots,n\} \to \{v_1,\dots,v_n\}$ such that $\tau(n) = v_n$ and $\tau(t) = j$ *iff* $P(j,t)$. Therefore $\sigma := \{\tau(1),\dots,\tau(n)\}$ is well defined. σ is a pebbling strategy, because $\tau(n) = v_n$, rule 1 is obeyed because of Ψ_4, rule 2 is obeyed, because unpebbling moves are given implicitly (see Theorem 0.0.1) and rule 3 is obeyed because τ is a bijection. Ψ_3 being satisfied ensures that σ uses no more than k pebbles.

Suppose there is a pebbling strategy σ using no more than k pebbles. Let the function $\mathrm{free} : \{1,\dots,n\} \to 2^{\{1,\dots,k\}} \setminus \emptyset$ be defined recursively as follows and $\mathrm{peb}(t) = \min(\mathrm{free}(t))$.

$$\mathrm{free}(t) = \begin{cases} \{1,\dots,k\} & : t = 1 \\ \mathrm{free}(t-1) \setminus \{\mathrm{peb}(t-1)\} \quad \cup & : otherwise \\ \Big\{ \mathrm{peb}(s) \mid \sigma_s \in P^{\varphi}_{\sigma_{t-1}}, s \in \{1,\dots,t-2\} \text{ and for all } v \in C^{\varphi}_{\sigma_s} & \\ \text{there exists } r \in \{1,\dots,t-1\} : \sigma_r = v \Big\} & \end{cases}$$

Intuitively, $\mathrm{free}(.)$ keeps track of the unused pebbles in each round. If a pebble is placed on a node, it is not free anymore. Pebbles are made free again by unpebbling moves, which correspond to the second set in the recursive definition of $\mathrm{free}(.)$. Since σ uses no more than k pebbles, $\mathrm{free}(.)$ is well defined.

Let $\mathcal{I}$ be a set of variables of Ψ defined as follows. $p_{x,j,t} \in \mathcal{I}$ *iff* $t > 0$ and there exists $s \in \{1,\dots,t\}$ such that $\mathrm{peb}(s) = x$, $\sigma_s = v_j$ and for all $r \in \{s+1,\dots,t\} : x \notin \mathrm{free}(r)$.

$\mathcal{I}$ is a satisfying assignment for Ψ. Ψ_1 is satisfied, because $\sigma_n = v_n$, therefore trivially $p_{\mathrm{peb}(n),n,n} \in \mathcal{I}$. Clearly Ψ_2 is satisfied by $\mathcal{I}$ as no variables with $t = 0$ are included in $\mathcal{I}$. To see that Ψ_3 is satisfied, suppose there exist x, t, i, j such that $i \neq j$ and $\{p_{x,j,t}, p_{x,i,t}\} \subseteq \mathcal{I}$. Then by definition of $\mathcal{I}$ there exist unique t_1 and t_2 such that $\mathrm{peb}(t_1) = x, \sigma_{t_1} = v_j$ and $\mathrm{peb}(t_2) = x, \sigma_{t_2} = v_i$. From $i \neq j$ follows $v_i \neq v_j$, therefore $t_1 \neq t_2$ w.l.o.g. suppose $t_1 > t_2$. From $\mathrm{peb}(t_2) = x$, $p_{x,i,t} \in \mathcal{I}$ and $t \geq t_1 > t_2$ follows $x \notin \mathrm{free}(t_1)$, which is a contradiction to $\mathrm{peb}(t_1) = x$. Let $P(x,j,t)$ be defined as above. Then from $P(x,j,t)$ follows $\mathrm{peb}(t) = x$ and $\sigma_t = v_j$. Rule 1 of the Bounded Pebbling Game ensures that there if $P(x,j,t)$ is true, then there exists a $y\{1,\dots,k\} \setminus \{x\}$ such that $p_{y,i,t-1} \in \mathcal{I}$. Suppose $P(x,j,t)$ and $P(y,i,t)$ both hold for some t, $x \neq y$ and $i \neq j$, then $y = \mathrm{peb}(t) = x$ and $v_j = \sigma_t = v_i$ are both contradictions. Therefore also Ψ_4 is satisfied by $\mathcal{I}$. □ □

Greedy Pebbling Algorithms

Theorem 0.0.2 and the remarks in the end of section indicate that obtaining an optimal topological order either by enumerating topological orders or by encoding the problem as a satisfiability problem is impractical. This section presents two greedy algorithms that aim at finding good though not necessarily optimal topological orders. They are both parameterized by some heuristic described in Section 7, but differ in the traversal direction in which the algorithms operate on proofs.

Top-Down Pebbling

Top-Down Pebbling (Algorithm 0.1) constructs a topological order of a proof φ by traversing it from its axioms to its root node. This approach closely corresponds to how a human would play the Bounded Pebbling Game. A human would look at the nodes that are available for pebbling in the current round of the game, choose one of them to pebble and remove pebbles if possible. Similarly the algorithm keeps track of pebblable nodes in a set N, initialized as A_φ. When a node v is pebbled, it is removed from N and added to the sequence representing the topological order. The children of v that become pebbleable are added to N. When N becomes empty, all nodes have been pebbled once and a topological order has been found.

Algorithm 0.1: `Top-Down Pebbling`

Input: proof φ
Output: sequence of nodes S representing a topological order $\prec$ of φ

$S = ()$; `// the empty sequence`
$N = A_\varphi$; `// initialize pebbleable nodes with Axioms`
while N *is not empty* **do**
  choose $v \in N$ heuristically;
  $S = S ::: (v)$; `// ::: is the concatenation of sequences`
  $N = N \setminus \{v\}$;
  for each $c \in C_v^\varphi$ **do** `// check whether` c `is now pebbleable`
    if $\forall p \in P_c^\varphi : p \in S$ **then**
      $N = N \cup \{c\}$;
return S;

Unfortunately Top-Down Pebbling often ends up finding a sub-optimal pebbling strategy regardless of the heuristic used. The following example shows such a situation.

Example 0.0.2. Consider the graph shown in Figure 1 and suppose that top-down pebbling has already pebbled the initial sequence of nodes $(1, 2, 3)$. For a greedy heuristic that only has information about pebbled nodes, their premises and children, all nodes marked with $4?$ are considered equally worthy to pebble next. Suppose the node marked with 4 in the middle graph is chosen to be pebbled next. Subsequently, pebbling 5 opens up the possibility to remove a pebble after the next move, which is to pebble 6. After that only the middle subgraph has to be pebbled. No matter in which order this is done, the strategy will use six pebbles at some point.

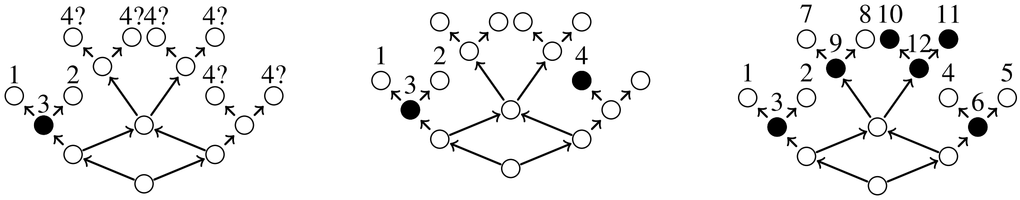

Figure 1: Top-Down Pebbling

One example sequence and the point where six pebbles are used are shown in the rightmost picture in Figure 1. However the pebbling number of this proof is 5.

Bottom-Up Pebbling

Bottom-Up Pebbling (Algorithm 0.2) constructs a topological order of a proof φ while traversing it from its root node r to its axioms. The algorithm constructs the order by visiting nodes and their premises recursively. For every node v the order in which the premises of v are visited is decided heuristically. After visiting the premises, n is added to the current sequence of nodes. Since axioms do not have any premises, there is no recursive call for axioms and these nodes are simply added to the sequence. The recursion is started with the call $BUpebble(\varphi, r, \emptyset, ())$. Since all proof nodes are ancestors of the root, the recursive calls will eventually visit all nodes once and a topological total order will be found. Bottom-Up Pebbling corresponds to the apply function $\mathrm{ap}(.)$ defined in Section with the addition of a visit order of the premises. Also previously visited nodes are not visited again.

Algorithm 0.2: BUpebble

Input: proof φ
Input: node v
Input: set of visited nodes V
Input: initial sequence of nodes S
Output: sequence of nodes

$V_1 = V \cup \{v\}$;
$N = P_v^\varphi \setminus V$; `// Only unprocessed premises are visited`
$S_1 = S$;
while N *is not empty* **do**
  choose $p \in N$ heuristically; $N = N \setminus p$;
  $S_1 = S_1 ::: BUpebble(\varphi, p, V, S)$; `// ::: is the concatenation of sequences`
return $S_1 ::: (v)$;

Example 0.0.3. Figure 2 shows part of an execution of Bottom-Up Pebbling on the same proof as presented in Figure 1. Nodes chosen by the heuristic, to be processed before the respective

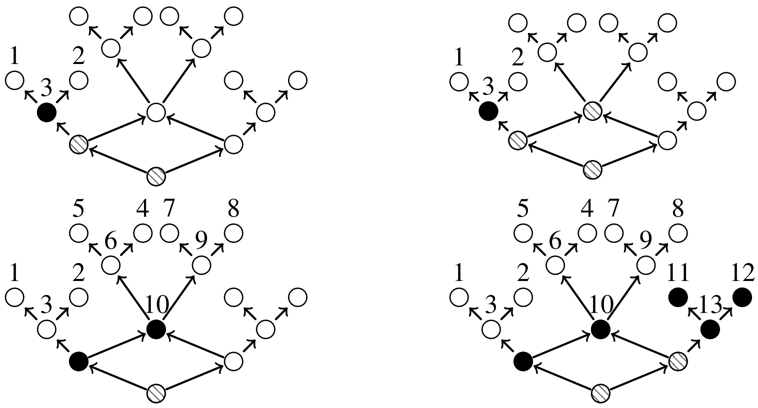

Figure 2: Bottom-Up Pebbling

other premise, are marked dashed. Suppose that similarly to the Top-Down Pebbling scenario, nodes have been chosen in such a way that the initial pebbling sequence is $(1, 2, 3)$. However, the choice of where to go next is predefined by the dashed nodes. Consider the dashed child of node 3. Since 3 has been completely processed, the other premise of its dashed child is visited next. The result is that the middle subgraph is pebbled while only one external node is pebbled, while it were two in the Top-Down scenario. At no point more than five pebbles will be used for pebbling the root node, which is shown in the bottom right picture of the figure. This is independently of the heuristic choices.

Remarks about Top-Down and Bottom-Up Pebbling

Every topological order of a given proof can be constructed using Top-down or Bottom-up Pebbling. A heuristic that orders nodes according to the desired topological order achieves this goal. Of course such a heuristic is not very useful in practice, as we do not know the desired topological order beforehand. Both algorithms traverse the proof only once and have linear run-time in the proof length (assuming that the heuristic choice requires constant time). Therefore both algorithms are theoretically equally good in constructing topological orders.

The experiments presented in Section 7 show that in practice, Bottom-Up Pebbling performs much better. Example 0.0.2 shows two principles that result in pebbling strategies with small pebbling numbers and are likely to be violated by the Top-Down Pebbling algorithm.

Firstly, pebbling strategies should make local choices. By local choices we mean that it should pebble nodes that are close w.r.t. undirected edges in the graph to other pebbled nodes. Such local choices allow to unpebble other nodes earlier and therefore keep the pebbling number low. Bottom-Up Pebbling makes local choices by construction, because premises are queued up and the second premise is visited as soon as possible. Top-Down Pebbling does not have knowledge about the recursive structure of children nodes, therefore it is hard to make local choices. The algorithm simply does not know which pebbleable nodes are close to other pebbled ones.

Secondly, pebbling strategies should pebble subproofs with a high pebbling number early. Pebbling such subproofs late will result in other pebbles staying on nodes for a high number of rounds. This likely results in increasing the overall pebbling number, as this adds extra pebbles to the already high pebbling number of the subproof. The principle is more subtle than the first one, because pebbling one subproof can influence the number of pebbles used for another subproof in situations where nodes are share between subproofs. The principle is demonstrated in the following example.

Example 0.0.4. Figure 3 shows a simple proof φ with two subproofs φ_0 (left branch) and φ_1 (right branch). As shown in the leftmost diagram, assume $s(\varphi_0, \prec_0) = 4$ and $s(\varphi_1, \prec_1) = 5$, where $\prec_0$ and $\prec_1$ represent some topological order of the respective subproofs with the corresponding pebbling numbers. After pebbling one of the subproofs, the pebble on its root node has to be kept there until the root of the other subproof is also pebbled. Only then the root node can be pebbled. Therefore, $s(\varphi, \prec) = s(\varphi_j, \prec_j) + 1$ where $\prec$ is obtained by first pebbling according to $\prec_j$, then by $\prec_{1-j}$ followed by pebbling the root. Choosing to pebble the less spacious subproof φ_0 first results in $s(\varphi, \prec) = 6$, while pebbling the more spacious one first gives $s(\varphi, \prec) = 5$.

Note that this example shows a simplified situation. The two subproofs do not share nodes. Pebbling one of them does not influence the pebbling number of the other.

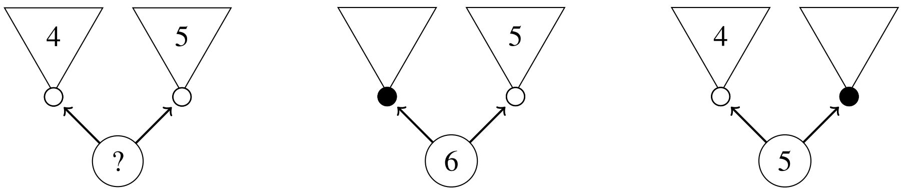

Figure 3: Spacious subproof first

Heuristics

Heuristics are used in both pebbling algorithms to choose one node out of a set N. For `Top-Down` Pebbling, N is the set of pebbleable nodes, and for `Bottom-Up` Pebbling, N is the set of unprocessed premises of a node.

Definition 0.0.10 (Heuristic and Full Heuristics)**.** Let φ be a proof with nodes V. A *heuristic* h for φ is a totally ordered set S_h together with a *node evaluation* function $e_h : V \to S_h$. A *full heuristic* for φ is finite a sequence $(e_{h_1}, \ldots, e_{h_n})$ of heuristics such that the node evaluation e_{h_n} is injective. The *choice* of the full heuristic for a set $N \subseteq V$ is some $v \in N$ such that $v = argmax_{v \in N} e_{h_1}(v)$ if v is unique and the choice of the full heuristic $(e_{h_2}, \ldots, e_{h_n})$ for $\{v \in N \mid v = argmax_{v \in N} e_{h_1}(v)\}$. This process will eventually terminate, because of the limitation to e_{h_n}.

Note that to satisfy the requirement for e_{h_n}, some trivial node evaluation like mapping nodes to their address in memory can be used. In the next chapters we present heuristics, which are cheap to compute and are justified by relating them to the semantics of the Bounded Pebbling Game. We will not elaborate on effects of reordering the heuristics within full heuristics.

Number of Children Heuristic ("Ch")

The `Number of Children` heuristic uses the number of children of a node v as evaluation function, i.e. $e_h(v) = |C_v^\varphi|$ and $S_h = \mathbb{N}$. The intuitive motivation for this heuristic is that nodes with many children will require many pebbles, and subproofs containing nodes with many children will tend to be more spacious. Example 0.0.4 shows the idea behind pebbling spacious subproofs early.

Last Child Heuristic ("Lc")

As discussed in Section in the proof of Theorem 0.0.1, the best moment to unpebble a node v is as soon as its last child w.r.t. a topological order $\prec$ is pebbled. This insight is used for the `Last Child` heuristic that chooses nodes that are last children of other nodes. Pebbling a node that allows another one to be unpebbled is always a good move. The current number of used pebbles (after pebbling the node and unpebbling one of its premises) does not increase. It might even decrease, if more than one premise can be unpebbled. For determining the number of premises of which a node is the last child, the proof has to be traversed once, using some topological order $\prec$. Before the traversal, $e_h(v) = 0$ for every node v. During the traversal $e_h(v)$ is incremented by 1, if v is the last child of the currently processed node w.r.t. $\prec$. For this heuristic $S_h = \mathbb{N}$. To some extent, this heuristic is paradoxical: v may be the last child of a node v' according to $\prec$, but pebbling it early may result in another topological order $\prec^*$ according to which v is not the last child of v'. Nevertheless, sometimes the proof structure ensures that some nodes are the last child of another node irrespective of the topological order. An example is shown in Figure 4, where the dashed line denotes a recursive predecessor relationship and the bottommost node is the last child of the top right node in every topological order.

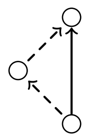

Figure 4: Bottommost node as necessary last child of right topmost node

Node Distance Heuristic ("$Dist(r)$")

In Example 0.0.2 and Section 7 it has been noted that `Top-Down` Pebbling may perform badly if nodes that are far apart are selected by the heuristic. The Node `Distance` heuristic prefers to pebble nodes that are close to pebbled nodes. It does this by calculating spheres with a radius up to the parameter r around nodes. A sphere $K_r^G(v)$ with radius r around the node v in the graph $G = (V, E)$ is the set $\{p \in V \mid v \text{ can be reached from } p \text{ visiting at most } r \text{ edges}\}$, where edges are considered undirected. The heuristic uses the following functions based on the spheres:

$$
\begin{aligned}
d(v) &:= \begin{cases} -min(D) \text{ such that } D = \{r \mid K_r^G(v) \text{ contains a pebbled node}\} \neq \emptyset \\ \infty \text{ otherwise} \end{cases} \\
s(v) &:= |K^G_{-d(v)}(v)| \\
l(v) &:= max_{\prec} K^G_{-d(v)}(v) \\
e_h(v) &:= (d(v), s(v), l(v))
\end{aligned}
$$

where $\prec$ denotes the order of previously pebbled nodes. So $S_h = \mathbb{Z} \times \mathbb{N} \times P$ together with the lexicographic order using, respectively, the natural smaller relation $<$ on $\mathbb{Z}$ and $\mathbb{N}$ and $\prec$ on N. The spheres $K_r(v)$ can grow exponentially in r. Therefore the maximum radius has to be kept small.

Decay Heuristics ("$Dc(h_u, \gamma, d, com)$")

`Decay` heuristics denote a family of meta heuristics. The idea is to not only use the evaluation of a single node, but also to include the evaluations of its premises. Such a heuristic has four parameters: an underlying heuristic h_u defined by an evaluation function e_u together with a well ordered set S_u, a decay factor $\gamma \in \mathbb{R}^+ \cup \{0\}$, a recursion depth $d \in \mathbb{N}$ and a combining function $com : S_u^n \to S_u$ for $n \in \mathbb{N}$. The resulting heuristic node evaluation function e_h is defined with the help of the recursive function rec:

$$
\begin{aligned}
rec(v, 0) &:= e_u(v) \\
rec(v, k) &:= e_u(v) + com(rec(p_1, k-1), \ldots, rec(p_n, k-1)) * \gamma \\
&\quad \text{where } P_v^{\varphi} = \{p_1, \ldots, p_n\} \\
e_h(v) &:= rec(v, d)
\end{aligned}
$$

Experiments

All the pebbling algorithms and heuristics described in the previous sections have been implemented in the hybrid functional and object-oriented programming language Scala (`www.scala-lang.org`) as part of the Skeptik library for proof compression (`github.com/Paradoxika/Skeptik`) [34].

To evaluate the algorithms and heuristics, experiments were executed[1] on four disjoint sets of proof benchmarks (Table 1). TraceCheck_1 and TraceCheck_2 contain proofs produced by the SAT-solver `PicoSAT` [8] on unsatisfiable benchmarks from the SATLIB (`www.satlib.org/benchm.html`) library. The proofs[2] are in the TraceCheck proof format, which is one of the three formats accepted at the *Certified Unsat* track of the SAT-Competition. veriT_1 and veriT_2 contain proofs produced by the SMT-solver VeriT (`www.verit-solver.org`) on

[1]The Vienna Scientific Cluster VSC-2 (`http://vsc.ac.at/`) was used.

[2]SAT proofs: `www.logic.at/people/bruno/Experiments/2014/Pebbling/tc-proofs.zip`

Name	**Number of proofs**	**Maximum length**	**Average length**
TraceCheck_1	2239	90756	5423
TraceCheck_2	215	1768249	268863
veriT_1	4187	2241042	103162
veriT_2	914	120075	5391

Table 1: Proof benchmark sets

Algorithm Heuristic	**Relative Performance** (%)	**Speed** (nodes/ms)
Bottom-Up		
Children	17.52	**88.6**
LastChild	**26.31**	84.5
Distance(1)	9.46	21.2
Distance(3)	-0.40	0.5
Top-Down		
Children	-27.47	0.3
LastChild	-31.98	1.9
Distance(1)	-70.14	0.6
Distance(3)	**-74.33**	**0.1**

Table 2: Experimental results

unsatisfiable problems from the SMT-Lib (`www.smtlib.org`). These proofs[3] are in a proof format that resembles SMT-Lib's problem format and they were translated into pure resolution proofs by considering every non-resolution inference as an axiom.

Table 2 summarizes the results of the experiments. The two presented algorithms are tested in combination with the four presented heuristics. The Children and LastChild heuristics were tested on all four benchmark sets. The Distance and Decay heuristics were tested on the sets TraceCheck_2 and veriT_2. The relative performance is calculated according to Formula 1, where f is an algorithm with a heuristic, P is the set of proofs the heuristic was tested on and G are all combinations of algorithms and heuristics that were tested on P. The time used to construct orders is measured in processed nodes per millisecond. Both columns show the best and worst result in boldface.

[3]SMT proofs: `www.logic.at/people/bruno/Experiments/2014/Pebbling/smt-proofs.zip`

Decay γ	**Depth** d	**Combination** com	**Performance Improvement** (%)	**Speed** (nodes/ms)
0.5	1	mean	0.50	47.7
0.5	1	maximum	0.40	47.0
0.5	7	mean	0.85	14.0
0.5	7	maximum	0.76	15.3
3	1	mean	0.48	64.0
3	1	maximum	0.43	**64.4**
3	7	mean	0.21	15.3
3	7	maximum	**0.94**	15.3

Table 3: Improvement of LastChild using Decay Heuristic

$$\text{relative\_performance}(f, P, G) = \frac{1}{|P|} * \sum_{\varphi \in P} \left(1 - \frac{s(\varphi, f(\varphi))}{avg_{g \in G} s(\varphi, g(\varphi))} \right) \tag{1}$$

Table 2 shows that the Bottom-Up algorithm constructs topological orders with much smaller space measures than the Top-Down algorithm. This fact is visualized in Figure 5, where each point represents a proof φ. The x and y coordinates are the smallest space measure among all heuristics obtained for φ using, respectively, the Top-Down and Bottom-Up algorithm. The results for Top-Down range far beyond 15000, but to display the discrepancy between the two algorithms the plot scales from 0 to 15000 on both axis. The biggest best space measure for Top-Down is 131 451, whereas this number is 11 520 for the Bottom-Up algorithm. The LastChild heuristic produces the best results and the Children heuristic also performs well. The Distance heuristic produces the worst results, which could be due to the fact that the radius is too small for big proofs with thousands of proof nodes.

Table 3 summarizes results of the Decay Heuristic with the best results highlighted in bold-face. Decay Heuristics were tested with the Bottom-Up algorithm, using LastChild as underlying heuristic. For the parameters decay factor, recursion depth and combining function two values and all their combinations have been tested. The performance improvement is calculated using Formula 1 with G being the singleton set of the Bottom-Up algorithm with the LastChild heuristic. The results show, that Decay Heuristics can improve the result, but not by a landslide. The improvement comes at the cost of slower speed, especially when the recursion depth is big.

Some additional heuristics, not described in this work, designed specifically for Top-Down Pebbling were tested on small benchmark sets. These heuristics aimed at doing local pebbling without having to calculate full spheres. For example pebbling nodes that allow other nodes to be unpebbled in the next move can be preferred. Unfortunately, none of the additional heuristics showed promising results.

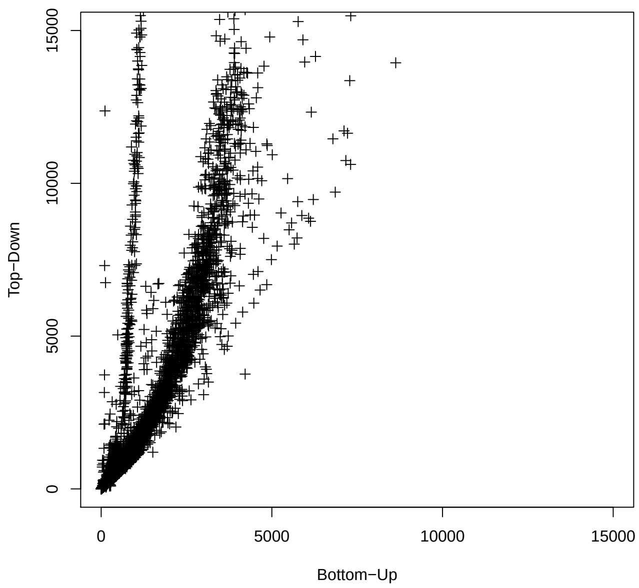

Figure 5: Space measures of best Bottom-Up and Top-Down result

The Bottom-Up algorithm does not only produce better results, it is also much faster, as can be seen in the last column of Table 2. The reason probably is the number of comparisons that the algorithms make. For Bottom-Up the set N of possible choices consists of the premises of a single node only, and usually $|N| \in O(1)$ (e.g. for a binary resolution proof, $N \leq 2$ always). For Top-Down the set N is the set of currently pebbleable nodes, which can be large (e.g. for a perfect binary tree with $2n - 1$ nodes, initially $|N| = n$). Possibly for some heuristics, Top-Down algorithms could be made more efficient by using, instead of a set, an ordered sequence of pebbleable nodes together with their memorized heuristic evaluations.

Unsurprisingly the radius used for the Distance Heuristic has a severe impact on the speed, which decreases rapidly as the maximum radius increases. With radius 5, only a few small proofs were processed in a reasonable amount of time.

On average the smallest space measure of a proof is 44.1 times smaller than its length. This shows the impact that the usage of deletion information together with well constructed topological orders can have. When these techniques are used, on average 44.1 times less memory is required for storing nodes in memory while proof processing.

Conclusion

Several algorithms for compressing proofs with respect to space have been conceived. The experimental evaluation clearly shows that the so-called Bottom-Up algorithms are faster and compress more than the more natural, straightforward and simple Top-Down algorithms. Both kinds of algorithms are parameterized by a heuristic function for selecting nodes. The best performances are achieved with the simplest heuristics (i.e. Last Child and Number of Children). More sophisticated heuristics provided little extra compression but cost a high price in execution time. Future work could investigate heuristics that take advantage of the particular shape of proofs generated by analysis of conflict graphs.

Acknowledgments: We would like to thank Armin Biere for clarifying why resolution chains are not left-associative in the TraceCheck proof format.

Congruence

Preliminaries

In this section, we summarize basic notions that we will use throughout the following sections.

Definition 0.0.11 (Terms)**.** Let $\mathcal{F}$ be a finite set of function symbols and $arity : \mathcal{F} \to \mathbb{N}$. A tuple $\Sigma = \langle \mathcal{F}, arity \rangle$ is called a *signature*. A function symbol with arity zero is called a *constant*, one with arity one is called a *unary* function symbol and one with arity 2 is called *binary*. For a given signature Σ, the set of *terms* $\mathcal{T}^\Sigma$ is defined inductively.

$$\begin{aligned}
\mathcal{T}_0^\Sigma &= \{a \in \mathcal{F} \mid arity(a) = 0\} \\
\mathcal{T}_{i+1}^\Sigma &= \{g(t_1, \ldots, t_n) \mid arity(g) = n \text{ and } t_1, \ldots, t_n \in \mathcal{T}_i\} \\
\mathcal{T}^\Sigma &= \bigcup_{i \in \mathbb{N}} \mathcal{T}_i
\end{aligned}$$

We will ommit the index Σ, if it is clear from context.
Todo: introduce all used notions

Resolution extended with equality

Let $\mathcal{T}$ be a set of terms. A relation $R \subset \mathcal{T} \times \mathcal{T}$ is a congruence relation, if has the following four properties:

- reflexive: for all $t \in \mathcal{T} : (t, t) \in R$
- symmetric: $(s, t) \in R$ implies $(t, s) \in R$
- transitive: $(r, s) \in R$ and $(s, t) \in R$ implies $(r, t) \in R$
- congruence: f is a n-ary function symbol and for all $i = 1, \ldots, n (t_i, s_i) \in R$ implies $f(t_1, \ldots, t_n), f(s_1, \ldots, s_n) \in R$

Every congruence relation partitions its underlying termset $\mathcal{T}$ into congruence classes, s.t. two terms (s, t) belong to the same class if and only if $(s, t) \in R$. The relations $\mathcal{T} \times \mathcal{T}$ and $\emptyset$ are trivial congruence relations.

Let E be a set of equations with terms in some set of terms $\mathcal{T}$. The set $E* \supseteq E$ is called the congruence closure of E, if $E*$ is a congruence relation on $\mathcal{T}$ and for every congruence relation C, such that $C \supset E$ follows $C \supseteq E*$. It is easily seen that congruence relations are closed under intersection. Therefore $E*$ always exists.

We write $E \models s \approx t$ if $(s,t) \in E*$.

We now extend the resolution calculus, presented in Section **??**, with the axioms of equality.

Todo: describe this calculus; prove relative correctness?

NP-completeness of shortest path decision problem

In this section we will assume that every propositional logic formula is given in conjunctive normal form and tuples of terms (u,v) are understood as equations $u = v$.

Definition 0.0.12 (Short explanation decision problem)**.** Let E be a set of input equations $s_1 = t_1, \ldots, s_n = t_n$ of terms $\mathcal{T}$. For $k \in \mathbb{N}$ and a target equation $s = t$, the short path decision problem is the question whether there exists an $E' \subseteq E$ such that $E' \models s = t$ and $|E'| \leq k$.

Note: $E \models s = t$ means s and t are in the same congruence class of the congruence closure of E.

Definition 0.0.13 (Congruence translation)**.** Let Φ be a propositional logic formula with clauses $C_1, \ldots, C_n$ using variables $x_1, \ldots, x_m$. The congruence translation E_Φ of Φ is defined as $Assignment \cup Pos \cup Neg \cup Connect$, where

$$\begin{aligned}
Assignment &= \{(\hat{x}_j, \top_j), (\hat{x}_j, \bot_j) \mid 1 \leq j \leq m\} \\
Pos &= \{(\hat{c}_i, t_i(\hat{x}_j)) \mid x_j \text{ appears positively in } C_i\} \\
Neg &= \{(\hat{c}_i, f_i(\hat{x}_j)) \mid x_j \text{ appears negatively in } C_i\} \\
Connect &= \{(t_i(\top_j), \hat{c}_{i+1}), (f_i(\bot_j), \hat{c}_{i+1}) \mid 1 \leq i \leq n+m, 1 \leq j \leq m\}
\end{aligned}$$

Additionally, for every $i = 1, \ldots, n$ and $j = 1, \ldots, m$ we define the following sets

$$\begin{aligned}
T_{ij} &= \{(\hat{c}_i, t_i(\hat{x}_j)), (\hat{x}_j, \top_j), (t_i(\top_j), \hat{c}_{i+1})\} \\
F_{ij} &= \{(\hat{c}_i, f_i(\hat{x}_j)), (\hat{x}_j, \bot_j), (f_i(\bot_j), \hat{c}_{i+1})\}
\end{aligned}$$

Example 0.0.5. Let $\Phi := (x_1 \vee x_2 \vee \neg x_3) \wedge (\neg x_2 \vee x_3) \wedge (\neg x_1 \vee \neg x_2)$. Figure 6 shows the graphical representation of the equations in Pos, Neg and $Connect$ for the congruence translation E_Φ of Φ. Figure 7 shows the subset $Assignment$ of E_Φ.

Let $\mathcal{I} := \{x_1, x_3\}$. It is easy to see that $\mathcal{I} \models \Phi$. Figure 8 shows a graphical representation of $\mathcal{I}$. Note that the satisfiability of Φ does not depend on x_2. Therefore replacing $Assignment$ with $Assignment' := Assignment \setminus \{(\hat{x}_2, \top_2)\} \cup \{(\hat{1}_2, \bot_1)\}$ in E_Φ leads to another explanation of $(\hat{c}_1, \hat{c}_4)$ of equal size. In Lemma 0.0.5 we exclude such ambiguous sets by introducing additional topological clauses.

Lemma 0.0.4 (Characterization of explanations)**.** *Let Φ be a propositional logic formula with n clauses and m variables. For every subset E of E_Φ, $E \models \hat{c}_1 = \hat{c}_{n+1}$ if and only if for every $i = 1, \ldots, n$ there is a $j = 1, \ldots, m$ such that $T_{ij} \subseteq E$ or $F_{ij} \subseteq E$.*

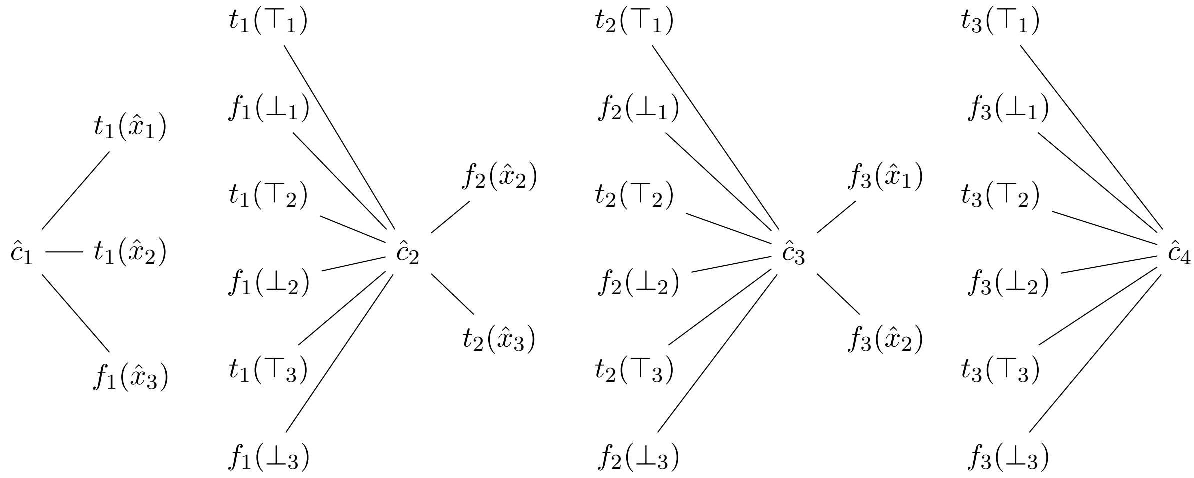

Figure 6: Pos, Neg and Connect for E_Φ

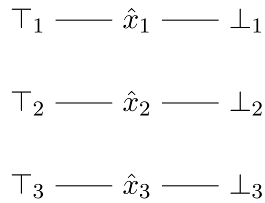

Figure 7: Assignment for E_Φ

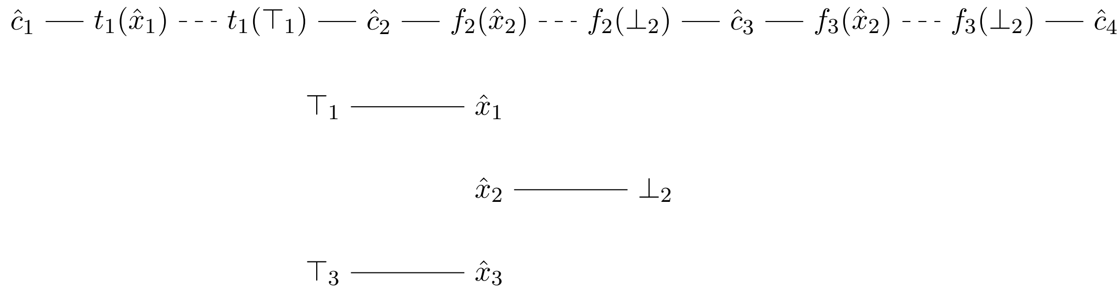

Figure 8: Explanation of $(\hat{c}_1, \hat{c}_4)$

Proof. Suppose that for every $i = 1, \ldots, n$ there is a $j = 1, \ldots, m$ such that $T_{ij} \subseteq E$ or $F_{ij} \subseteq E$. Clearly $T_{ij} \models \hat{c}_i = t_i(\hat{x}_j)$ and $T_{ij} \models t_i(\top_j) = \hat{c}_{i+1}$. Since $(\hat{x}_j, \top_j) \in E$ the fact $E \models t_i(\hat{x}_j) = t_i(\top_j)$ follows by an application of the deduction axiom. Using the transitivity axiom it follows that $T_{ij} \models \hat{c}_i = \hat{c}_{i+1}$. Similarly it can be shown that $F_{ij} \models \hat{c}_i = \hat{c}_{i+1}$. Therefore it follows from the assumption that $E \models \hat{c}_i = \hat{c}_{i+1}$ for every $i = 1, \ldots, n$. Using the transitivity axiom it follows that $E \models \hat{c}_1 = \hat{c}_{n+1}$.

We will show the other direction of the equivalence by induction on n.

Induction Base $n = 1$: Suppose that $E \models \hat{c}_1 = \hat{c}_2$. Since $\hat{c}_1$ is a constant, the deduction axiom can not be applied to any equation with $\hat{c}_1$ on one side. Therefore in order to satisfy $E \models \hat{c}_1 = t$ with $t \neq \hat{c}_1$ there has to be an equation $(\hat{c}_1, t) \in E$ for some term t. Since $E \subseteq E_\Phi$, the only possible such equations are of the form $(\hat{c}_1, t_1(\hat{x}_j))$ and $(\hat{c}_1, f_1(\hat{x}_j))$ for some j. The only equations in E involving terms with the function symbols t_1 and f_1 are of the form $(\hat{c}_1, t_1(\hat{x}_j)), (t_1(\top_j), \hat{c}_2)$ and $(\hat{c}_1, f_1(\hat{x}_j)), (f_1(\bot_j), \hat{c}_2)$. Therefore in order to satisfy $E \models \hat{c}_1 = t$ such that t is neither the constant $\hat{c}_1$ nor some term $t_1(\hat{x}_j), f_1(\hat{x}_j)$, it is necessary that $E \models t_1(\hat{x}_j) = t_1(\top_j)$ and $(\hat{c}_1, t_1(\hat{x}_j)) \in E$ or $E \models f_1(\hat{x}_j) = f_1(\bot_j)$ and $(f_1(\bot_j), \hat{c}_2) \in E$ for some j. The conditions can only be satisfied with equations of E_Φ if $\{(\hat{c}_1, t_1(\hat{x}_j)), (\hat{x}_j, \top_j)\} \subseteq E$ or $\{(\hat{c}_1, f_1(\hat{x}_j), (\hat{x}_j, \bot_j)\} \subseteq E$ respectively. From a similar argumentation about the equations involving c_2 and $t_1(\top_j)$ or $f_1(\bot_j)$ it follows that either $T_{1j} \subseteq E$ or $F_{1j} \subseteq E$ for some j.

Induction Hypothesis: For every subset E of E_Φ, $E \models \hat{c}_1 = \hat{c}_n$ if and only if for every $i = 1, \ldots, n-1$ there is a $j = 1, \ldots, m$ such that $T_{ij} \subseteq E$ or $F_{ij} \subseteq E$.

Induction Step: Suppose that $E \models \hat{c}_1 = \hat{c}_{n+1}$.
Similarly to the argumentation in the induction base, the only equations in E_Φ involving $\hat{c}_{n+1}$ are of the form $(t_n(\top_j), \hat{c}_{n+1})$ and $(f_n(\bot_j), \hat{c}_{n+1})$. The only possibility to enrich the congruence class of $\hat{c}_{n+1}$ with terms other than $\hat{c}_{n+1}$ and those of the form $t_n(\top_j)$ and $f_n(\bot_j)$, is that for some j, $(\hat{x}_j, \top_j) \in E$ or $(\hat{x}_j, \bot_j) \in E$ and subsequently also $(\hat{c}_n, t_n(\hat{x}_j)) \in E$ or $(\hat{c}_n, f_n(\hat{x}_j)) \in E$. Thus $T_{nj} \subseteq E$ or $F_{nj} \subseteq E$ and as a consequence $E \models \hat{c}_n = \hat{c}_{n+1}$. Using transitivity $E \models \hat{c}_1 = \hat{c}_{n+1}$ and $E \models \hat{c}_n = \hat{c}_{n+1}$ imply $E \models \hat{c}_1 = \hat{c}_n$ and from the induction hypothesis it follows that $T_{ij} \subseteq E$ or $F_{ij} \subseteq E$ for every $i = 1, \ldots, n-1$.

□

Lemma 0.0.5 (NP- hardness)**.** *The short path decision problem is NP- hard.*

Proof. We will reduce SAT to the short path decision problem. Let Φ be a propositional formula in conjunctive normal form with variables $x_1, \ldots, x_m$ and clauses $C_1, \ldots, C_n$. Let $C_{n+1}, \ldots, C_{n+m}$ be tautological clauses $\{x_1, \neg x_1\}, \ldots, \{x_m, \neg x_m\}$. Clearly Φ is satisfiable if and only if $\Phi' = \{c_1, \ldots, c_{n+m}\}$ is satisfiable. We will show that Φ' is satisfiable if and only if there exists $E \subseteq E_{\Phi'}$ such that $E \models \hat{c}_1 = \hat{c}_{n+m+1}$ and $|E| \leq 2n + 3m$.

Suppose Φ' is satisfiable and let $\mathcal{I}$ be a satisfying assignment.
For every clause C_i there is a literal $\ell_i \in C_i$ such that $\mathcal{I} \models \ell_i$. For every $i = 1, \ldots, n+m$ we define E_i to be T_{ij} if $\ell_i = x_j$ or F_{ij} if $\ell_i = \neg x_j$. From $\ell_i \in C_i$ it follows $E_i \subseteq E_{\Phi'}$. Let $E = \bigcup_i^n E_i$ then from Lemma 0.0.4 follows $E \models c_1 = c_{n+m+1}$. What remains to show is that $|E| \leq 2n + 3m$. Since the sets E_2, E_3 and E_4 in the definition of $E_{\Phi'}$ are pairwise disjoint, for $i \neq j$ $E_i \cap E_j \subseteq \{(\hat{x}_j, \top_j), (\hat{x}_j, \bot_j) \mid j = 1, \ldots, m\}$. Therefore E involves exactly $2(n+m)$ equations of E_2, E_3 and E_4. By construction of the sets E_i and the clauses $C_{n+1}, \ldots, C_{n+m}$ there is no $j = 1, \ldots, m$ such that $(\hat{x}_j, \top_j) \in E$ and $(\hat{x}_j, \bot_j) \in E$. Therefore E involves m equations of set E_1 in the definition of $E_{\Phi'}$. Overall we have $|E| = 2n + 3m$.

Suppose there exists $E \subseteq E_{\Phi'}$, $E \models \hat{c}_1 = \hat{c}_{n+m+1}$ and $|E| \leq 2n + 3m$.
We will show that $\mathcal{I} = \{\hat{x}_j \mid (\hat{x}_j, \top_j) \in E\}$ is a satisfying assignment for Φ'. Let C_i be an arbitrary clause of Φ'. From $E \models \hat{c}_1 = \hat{c}_{n+m+1}$ and Lemma 0.0.4 follows $T_{ij} \subseteq E$ or $F_{ij} \subseteq E$ for some $j = 1, \ldots, m$.
Assume $T_{ij} \subseteq E$ for some $j = 1, \ldots, m$. $E \subseteq E_{\Phi'}$ implies that x_j appears positively in C_i. By definition of $\mathcal{I} \models x_j$. Therefore $\mathcal{I} \models C_i$.
If $T_{ij} \nsubseteq E$ for all $j = 1, \ldots, m$, then $F_{ij} \subseteq E \subseteq E_{\Phi'}$, which implies that x_j appears negatively in C_i, $x_j \notin \mathcal{I}$. Therefore $\mathcal{I} \models C_i$.
Since i was arbitrary $\mathcal{I} \models \Phi'$.

□

Lemma 0.0.6 (In NP)**.** *The short path decision problem is in NP.*

Proof. Explanations are subsets of the input equations, therefore they are clearly polynomial in size. The congruence of two terms, i.e. verifying that a subset is actually an explanation, can be decided in $O(nlog(n))$ using for example the congruence closure algorithm presented in Section .

□

Lemma 0.0.5 and 0.0.6 establish the main result of this section.

Theorem 0.0.7 (NP - completeness)**.** *The short path decision problem is NP- complete.*

Algorithms

In this section, we present a congruence closure algorithm that is able to produce explanations. The algorithm is a mix of the approaches of the algorithms presented in [19] and [31, 32]. The basic structure of the algorithm is inherited from [19], which itself inherits its structure from the algorithm of Nelson and Oppen [30]. The technique to store and deduce equations of non constant terms is inspired from [31, 32]. Additionally the proof forest structure described below was proposed by [31, 32].

Preliminaries

Our congruence closure algorithm operates on curried terms. Curried terms use a single binary function symbol to represent general terms. More formally let $\mathcal{F}$ be a finite set of functions with a designated binary function symbol $f \in \mathcal{F}$ and let every other function symbol in $\mathcal{F}$ be a constant. A term w.r.t. a signature of this form is called a *curried term.*

It is possible to uniquely translate a general set of terms $\mathcal{T}^{\Sigma}$ with signature $\Sigma = \langle \mathcal{F}, arity \rangle$ into a set of curried terms $\mathcal{T'}^{\Sigma'}$. Σ' is obtained from Σ by setting $arity$ to zero for every function symbol in $\mathcal{F}$ and introducing the designated binary function symbol f to $\mathcal{F}$. The translation of a term $t \in \mathcal{T}^{\Sigma}$ is given in terms of the function $curry$.

$$curry(t) = \begin{cases} t & \text{if } t \text{ is a constant} \\ f(\ldots(f(f(g, curry(t_1)), curry(t_2)))\ldots, curry(t_n)) & \text{if } t = g(t_1, \ldots, t_n) \end{cases}$$

The idea of currying was introduced by M. Schönfinkel [37] in 1924 and independently by Haskell B. Curry [15] in 1958, who also lends his name to the concept. Currying is not restricted to terms. The general indea is to translate functions of type $A \times B \to C$ into functions of type $A \to B \to C$. There is a close relation between currying and lambda calculus [13]. Lambda calculus uses a single binary function λ. Its arguments can either be elements of some set or again lambda terms. For an introduction to lambda calculus, including currying in terms of lambda calculus and its relation to functional programming, see [3].

The benefit of working with curried terms is an easier and cleaner congruence closure algorithm that runs in optimal time $O(n \log(n))$.

Recently so called abstract congruence closure algorithms have been proposed and shown to be more efficient than traditional approaches [1]. The idea of abstract congruence closure is to introduce new constants for non constant terms. Doing so, all of equations the algorithm has to take into account are of the form $c = d$ and $c = f(a, b)$, where a, b, c, d are constants. This replaces tedious preprocessing steps, for example transformation to a graph of outdegree 2 [17],that are necessary for other algorithms to achieve the optimal running time.

Our method is does not employ the idea of abstract congruence closure. We found that using currying is enough to obtain an algorithm with optimal running time and no tedious preprocessing steps. The reason why we did not go for abstract congruence closure is, that we do not want to have the overhead of introducing and eliminating fresh constants. In the context of proof compression, our congruence closure algorithm will be applied to relatively small instances very often. We could introduce the extra constants for the whole proof before processing, but would still have to remove them from explanations every time we produce a new subproof. It would be interesting to investigate, whether our intuition in that regard is right, or if it pays off to deal with extra constants.

Comming back to the explanation producing congruence closure algorithms that inspired ours, [31, 32] describes an abstract one using currying. [19] uses a traditional algorithm without currying and extra constants. Our algorithm is a middle ground between them.

Congruence structure

We call the underlying data structure of our congruence closure algorithm a *congruence structure*. A congruence structure for set of terms $\mathcal{T}$ is a collection of the following data structures.

- Representative $r : \mathcal{T} \to \mathcal{T}$
- Congruence class $[.] : \mathcal{T} \to 2^{\mathcal{T}}$
- Left neighbors $lN : \mathcal{T} \to 2^{\mathcal{T}}$
- Right neighbors $rN : \mathcal{T} \to 2^{\mathcal{T}}$
- Lookup table $l : \mathcal{T} \times \mathcal{T} \to \mathcal{T}$
- Congruence graph g
- Queue $\mathcal{Q}$ of type $\mathcal{T} \times \mathcal{T}$

- Current explanations $\mathcal{M} : \mathcal{T} \times \mathcal{T} \rightarrow \mathcal{E}$

The representative is one particular term of a class of congruent terms. It is used to identify whether two terms are already in the same congruence class and the data structures used for detecting equalities derived from the congruence axiom are kept updated only for representatives. The congruence class structure represents a set of pairwise congruent terms. It is used to keep track which representatives have to be updated when merging the classes of two terms. The structures left neighbor and right neighbor for every term keep track of other terms that appear as the second argument in a compound term of the form $f(a, b)$. The lookup table is used to keep track of all compound terms in the congruence structure and to merge compound terms, which arguments are congruent. The congruence graph stores the derived equalities in a structured way, that allows to create explanations for a given pair of terms. Edges are added to the graph in a lazy way, meaning that they are buffered and only actually entered into the graph when demanded. The queue $\mathcal{Q}$ keeps track of the order in which edges should be added to the graph. The function $\mathcal{M}$ stores the explanation for a buffered edge. The main idea about buffering is to overwrite explanations, when an edge was added due to the congruence axiom, while it is entered as an input equation later on. We call the unique congruence structure for $\mathcal{T} = \emptyset$ the *empty congruence structure*. It is not by coincidence that many of the used data structures are described as functions. In fact our congruence closure algorithm can and is implemented in a functional way and the data structures can be implemented immutable.

Congruence closure algorithms

In this section we present the pseudocode of our congruence closure algorithm, state and prove its properties. Most importantly we show that it the method is sound and complete and has optimal running time $O(n \log(n))$. Computing the congruence closure of some set of equations E is done by adding all of them to an ever growing congruence structure, which initially is empty. Most algorithm pseudocodes do not include a return statement. In fact every algorithm implicitly returns a (modified) congruence structure or simply modifies a global variable, which is the current congruence structure. Since this has to be done in some order, we will often assume that E is given as a sequence of equations rather than a set. Adding an equation to a congruence structure is done with the addEquation method. The method adds boths sides of the equation to the current set of terms $\mathcal{T}$ using the addNode method and afterwards merges the classes of the two terms. The `addNode` method enlarges the set of terms and searches for equalities that are due to the congruence axiom. The updates of $\mathcal{T}$ are not outlined explicitly, but are understood to happen implicitly. The method `merge` initializes and guides the merging of terms. The actual merging is done by the method union by modifying the data structures.

Invariant 0.0.8 (Class)**.** *For every $s \in \mathcal{T}$ and every $t \in [r(s)]$, $r(t) = r(s)$.*

For all invariants I might need to be more specific as to when exactly they should old.

Proof. Clearly the invariant is true when intializing $[s]$ in line 2 of addNode.

The only other point in the code that changes $[s]$ is line 36 of union. Suppose the class of u is enlarged by the class of v in union and suppose the invariant holds before the union for

Algorithm 0.3: addEquation

Input: equation $s = t$ or null

1 addNode(s)
2 addNode(t)
3 merge($s, t, s = t$)

Algorithm 0.4: addNode

Input: term v

1 **if** r *is not defined for* v **then**
2     $r(v) \leftarrow v$
3     $[v] \leftarrow \{v\}$
4     $lN(v) \leftarrow \emptyset$
5     $rN(v) \leftarrow \emptyset$
6   **if** v *is of the form* $f(a, b)$ **then**
7   addNode(a)
8   addNode(b)
9   **if** l *is defined for* $(r(a), r(b))$ *and* $l(r(a), r(b)) \neq f(a, b)$ **then**
10   merge($l(r(a), r(b)), f(a, b), \emptyset$)
11   **else**
12             $l(r(a), r(b)) \leftarrow f(a, b)$
13         $lN(r(b)) \leftarrow lN(r(b)) \cup \{a\}$
14         $rN(r(a)) \leftarrow rN(r(a)) \cup \{b\}$

those terms. Before the update of $[r(u)]$ the representative of every term in $[r(v)]$ is set to $r(u)$. Therefore the invariant remains valid after the update.

□

Invariant 0.0.9 (Lookup)**.** *The lookup structure l is defined for a pair of terms (s, t) if and only if there is a term $f(a, b) \in \mathcal{T}$ such that $r(a) = r(s)$ and $r(b) = r(t)$.*

Proof. Suppose l is defined for some pair of terms (s, t). The value of $l(s, t)$ was either set in lines 32 or 18 of union or in line 40 of addNode. In the latter case, l is set to $f(a, b)$ for the tuple $(r(a), r(b))$ and therefore the invariant holds at this point. For changes to $r(a)$ or $r(b)$ in union the one implication of the invariant remains valid in case l is defined for the new representatives, or l is set for an additional pair of terms in lines 32 or 18. In case l is set to $(new\_left, r(u))$ or $(r(u), new\_right)$ in union, there is an l-entry l_v for which the invariant held before the union. The changes in representatives of x are reflected by new_left and new_right, while the representative of v is changed to $r(u)$. The new entry for l therefore respects the implication of the invariant.

To show the other implication, let $f(a, b) \in \mathcal{T}$. The term $f(a, b)$ is entered via the addEquation and subsequently via the addNode method. For compound terms lines and assert that l is

Algorithm 0.5: merge

Input: term s
Input: term t
Input: extended equation eq
if $r(s) \neq r(t)$ **then**
  $c \leftarrow \{s = t\}$
  $eq \leftarrow s = t$
  while $c \neq \emptyset$ **do**
    Let (u, v) be some element in c
    $c \leftarrow c \setminus \{(u, v)\} \cup union(u, v)$
    lazy_insert(u, v, eq)
    $eq \leftarrow null$

Algorithm 0.6: lazy_insert

Input: term s
Input: term t
Input: extended equation eq
if $\mathcal{M}$ *is set for* (s, t) *or* (t, s) **then**
  if *eq is not null* **then**
    $\mathcal{M}(s, t) \leftarrow s = t$
else
  $\mathcal{Q} \leftarrow \mathcal{Q}.enqueue(s, t)$
  $\mathcal{M}(s, t) \leftarrow eq$

Algorithm 0.7: lazy_update

while $\mathcal{Q}$ *is not empty* **do**
  $(u, v) \leftarrow \mathcal{Q}.dequeue$
  $eq \leftarrow \mathcal{M}(u, v)$
  $g.insert(u, v, eq)$

defined for $(r(a), r(b))$. All changes to $r(a)$ or $r(b)$ must happen in union and they are reflected by matching updates to the l structure.

□

Invariant 0.0.10 (Neighbours)**.** *For every $s \in \mathcal{T}$, every $t_r \in rN(r(s))$ and $t_l \in lN(r(s))$, l is defined for $(r(s), r(t_r))$ and $(r(t_l), r(s))$.*

Proof. We show the result for the structure rN. The result about lN can be obtained similarly. Since rN is initialized with the empty set in line 4 of addNode, the invariant clearly holds initially. To show that the invariant always holds, it has to be shown that all modifications of r and rN do not change the invariant. The structure l is not modified after initialization. The structure r is modified in line 35 of union. The structure rN is modified in line 13 of addNode and line 38 of union.

Line 13 of addNode adds b to $rN(r(a))$ and the four lines before that addition show that l is defined for $(r(a), r(b))$.

Union modifies rN in such a way that it adds all right neighbors of some representative $r(v)$ to $rN(r(u))$. Lines 19 to 33 make sure that l is defined for all these right neighbors.

□

A consequence of this invariant is and the fact that the statement is true after inserting, that for every term $t \in \mathcal{T}$ of the form $f(a, b)$, l is defined for $(r(a), r(b))$.

Proposition 0.0.11 (Sound- & Completeness)**.** *Let $\mathcal{C}$ be the congruence structure obtained by adding equations $E = \langle(u_1, v_1), \ldots, (u_n, v_n)\rangle$ to the empty congruence structure. For every $s, t \in \mathcal{T}$: $E \models s \approx t$ if and only if $r(s) = r(t)$.*

Proof. **Completeness**

We show that from $E \models s \approx t$ follows $r(s) = r(t)$ by induction on n.

Base case $n = 1$: $E \models s \approx t$ implies either $s = t$ or $\{u_1, v_1\} = \{s, t\}$. In the first case $r(s) = r(t)$ is trivial. In the second case, the claim follows from the fact that, when (u_1, v_1) is entered, union is called with arguments s and t. After this operation $r(s) = r(t)$.

Induction hypothesis: For every sequence of equations E_n with n elements and every $s, t \in \mathcal{T}_{E_n}$: $E_n \models s \approx t$ then $r(s) = r(t)$.

Induction step: Let $E = \langle(u_1, v_1), \ldots, (u_{n+1}, v_{n+1})\rangle$ and $E_n = \langle(u_1, v_1), \ldots, (u_n, v_n)\rangle$. There are two cases: $E_n \models s \approx t$ and $E_n \not\models s \approx t$. In the former case, the claim follows from the induction hypothesis, the invariant class and the fact that union always changes representatives for all elements of a class. We still have to show the claim in the latter case. We write $E \models_n u \approx v$ as an abbreviation for $E_n \not\models u \approx v$ and $E \models u \approx v$. We show the claim by induction on the structure of the terms s and t.

Base case: s or t is a constant and therefore the transitivity reasoning was used to derive $E \models_n s \approx t$. In other words, there are l terms $t_1, \ldots, t_l$ such that $s = t_1$, $t = t_l$ and for all $i = 1, \ldots, l-1 : E \models_n t_i \approx t_{i+1}$. We prove by yet another induction on l that $r(t_1) = r(t_l)$. Base case $l = 2$. It has to be the case (up to swapping u_{n+1} with v_{n+1}), that $E_n \models s \approx u_{n+1}$ and $E_n \models t \approx v_{n+1}$, and the outmost induction hypothesis implies $r(s) = r(u_{n+1})$ and $r(t) = r(v_{n+1})$. Therefore it follows from Invariant Class, that after the call to union for (u_{n+1}, v_{n+1}) it is the case that $r(t_1) = r(t_2)$. Suppose that the claim holds for some $l \in \mathbb{N}$. In the induction step, going from l to $l+1$, the claim follows from a simple application of the transitivity axiom, since $t_1, \ldots, t_l$ and $t_2, \ldots, t_{l+1}$ are both sequences of length l.

For the induction step of the term-structure induction, suppose that $s = f(a, b)$ and $t = f(c, d)$. There are two cases such that $E \models_n s \approx t$ can be derived. Using a transitivity chain, the claim can be shown just like in the base case. Using the congruence axiom, it has to be the case that $E \models_n a \approx c$ and $E \models_n b \approx d$ (in fact one of those can also be the case without the n index). The terms a, b, c, d are of lower structure than s and t. Therefore it follows from the induction hypothesis that $r(a) = r(c)$ and $r(b) = r(d)$. The Invariants Neighbour and Lookup imply that either $r(s) = r(t)$ or (s, t) is added to d in line 14 or line line 28 of union. Subsequently union is called for s and t, after which $r(s) = r(t)$ holds.

The triple induction is a little bit weird and can surely still be improved. Also I am not 100% sure whether I am sometimes too loose with argumentation. On the other hand I have the feeling that it gets really technical sometimes.

Soundness

For $s = t$ the claim follows trivially. Therefore we show soundness in case $s \neq t$. We show that from $r(s) = r(t)$ follows $E \models s \approx t$ by induction on the number k of calls to union induced by adding all equations of E to the empty congruence structure for all s and t that are arguments of some call to union. The original claim then follows from invariant Class, since only union modifies the r structure and the fact that two terms are in the same class if and only if union was called for some elements in the respective classes.

Base case $k = 1$: $r(s) = r(t)$ implies $\{u_1, v_1\} = \{s, t\}$ and $E \models s \approx t$ is trivial.

Induction hypothesis: For every $l < k$, if a set of equations F induces l calls to union, then from $r(s) = r(t)$ follows $F \models s \approx t$ for all terms s, t that are arguments of some call to union.

Induction step: Suppose $E = \langle (u_1, v_1), \ldots, (u_n, v_n) \rangle$ induces k calls to union with arguments $(h_1, g_1), \ldots, (h_k, g_k)$. The subsequence $E_n = \langle (u_1, v_1), \ldots, (u_{n-1}, v_{n-1}) \rangle$ induced the first l calls to union for some $n - 1 \leq l < k$. In other words, adding (u_n, v_n) to the congruence structure induces the calls to union with arguments $(h_{k-l}, g_{k-l}), \ldots, (h_k, g_k)$. The first call to union with arguments (h_{k-l}, g_{k-l}) is either an original input equation, or a deduced equality from line 9 of `addNode`. In both cases $E \models h_{k-l} \approx g_{k-l}$, which is trivial in the former case and an application of the induction hypothesis in the latter case. Union induces additional union calls in such a way that the arguments of the additional call are on parent terms of the respective original arguments. Therefore, using induction on the structure of terms, the original induction hypothesis, Invariants Lookup and Neighbour and lines 5 to 33 of `union`, it can be shown that for all pairs (h_m, g_m) and all $m = k - l + 1, \ldots, k$ it is the case that $E \models h_m \approx g_m$.

□

Proposition 0.0.12 (Runtime)**.** *Let E be a set of equations that uses n terms. Computing the congruence closure with our congruence closure algorithm takes worst-case time $O(n \log(n))$.*

Proof. There are three loops in the method `union`, which are nested within the loop of `merge`. These loops are clearly the dominating factor for runtime.

Lines reverse1 and reverse2 of `union` make sure that everytime the representative of a term is changed, the size of its congruence class is doubled. The maximum size of a congruence class is n. Therefore the representative of a single term is changed maximally $\log(n)$ times overall and line 35 of `union` is not executed more than $n \log(n)$ times.

I am not yet sure how to prove, and to be honest even whether it is true, that line the right neighbor, left neighbor loops of union are not executed n$^2 times overall$.

□

Algorithm 0.8: union

```
Input: term s
Input: term t
Output: a set of deduced equations
if [r(s)] ≥ [r(t)] then
    reverse1 (u, v) ← (s, t)
else
    reverse2 (u, v) ← (t, s)
d ← ∅
for every x ∈ lN(r(v)) do
    l_v ← l(r(x), r(v))
    if r(x) = r(v) then
        new_left ← r(u)
    else
        new_left ← r(x)
    if l is defined for (new_left, r(u)) then
        l_u ← l(new_left, r(u))
        if r(l_u) ≠ r(l_v) then
            d ← d ∪ {(l_u, l_v)}
        else
            lN(r(v)) ← lN(r(v)) \ {x}
    else
        l(new_left, r(u)) ← l_v
for every x ∈ rN(r(v)) do
    l_v ← l(r(v), r(x))
    if r(x) = r(v) then
        new_right ← r(u)
    else
        new_right ← r(x)
    if l is defined for (r(u), new_right) then
        l_u ← l(r(u), new_right)
        if r(l_u) ≠ r(l_v) then
            d ← d ∪ {(l_u, l_v)}
        else
            rN(r(v)) ← rN(r(v)) \ {x}
    else
        l(r(u), new_right) ← l_v
[r(u)] ← [r(u)] ∪ [r(v)]
for every x ∈ [r(v)] do
    r(x) ← r(u)
[r(u)] ← [r(u)] ∪ [r(v)]
lN(r(u)) ← lN(r(u)) ∪ lN(r(v))
rN(r(u)) ← rN(r(u)) ∪ rN(r(v))
return d
```

Congruence graph

The main goal of this work is to replace redundant explanations with shorter ones. For this purpose the input equations and deduced equalities have to be stored in a data structure that supports the production of explanations. We support two different such data structures. Both structures store equations in labeled graphs, which we call congruence graphs. A node in such a graph represents a term and an edge between two nodes denotes that the represented terms are congruent w.r.t. the set of input equations. A path in a congruence graph is a sequence of undirected, unweighted, labeled edges in the underlying graph. The set of labels for both types of graphs is the set of extended equations $\mathcal{E}$. Depending on the type of congruence graph used, it is not guaranteed that when `lazy_insert` is called with arguments s and t, there is an edge between s and t. However it is guaranteed that they are connected in the graph afterwards, i.e. there is a path between the nodes representing the terms.

Invariant 0.0.13 (Paths)**.** *For terms s, t such that $s \neq t$ and a congruence structure with representative function r holds $r(s) = r(t)$ if and only if there is a path in the congruence graph of the structure between s and t*

Proof. We show the claim by an induction on $|[r(s)]|$. The proof relies on the invariant Class, which shows the consistency between classes and representatives.

In the induction base $[r(s)] = \{s\}$, i.e. $r(s) = r(t)$ is false for every term $t \neq s$. We have to show that there is no edge (s, t) for $t \neq s$ in the congruence graph. Edges are only added to the congruence graph via the `lazy_insert` method which is only called in `merge`. Clearly `merge` does not call `union` for s and some term $t \neq s$, since otherwise $t \in [r(s)]$. Therefore `merge` also does not add an edge for s and some term $t \neq s$ to the congruence graph.

Let the induction hypothesis be, that for every term s such that $|[r(s)]| \leq n$, for every term $t \neq s$ it is the case that $r(s) = r(t)$ if and only if there is a path between s and t in the congruence graph.

Suppose $[r(s)]$ is an arbitrary class with cardinality $n + 1$. Then there are two terms $u, v \in [r(s)]$ such that `union` was called for u and v. Before the union $|[r(u)]|$ and $|[r(v)]|$ both were strictly smaller than $n+1$. In case they both belong to the same class before the union, the claim follows trivially by the induction hypothesis, since existing paths are not removed by adding new edges to the graph. Suppose $s \in [r(u)]$ and $t \in [r(v)]$, then by induction hypothesis there are paths p_1 between s and u and p_2 between t and v. Right after the union of u and v, an edge is inserted between them, so p_1 concatenated with (u, v) and p_2 is a path between s and t. In case one of the terms did not belong to one of the classes before the union, it does not belong to the merged class after the union. Also there was no path between the two terms before and since the only addition paths are between elements of $[r(u)]$ and $[r(v)]$, there is no path between the terms after the union.

□

Invariant 0.0.14 (Deduced Edges)**.** *For every edge in a congruence structure between vertices u, v with label $null$, there are $a, b, c, d \in \mathcal{T}$ such that $u = f(a, b)$, $v = f(c, d)$ and there are paths in the underlying graphs between a and c aswell as b and d.*

Proof. Edges with label $null$ are added, when `merge` is called from `addNode`, or `union` induces an additional merge. In both cases there are subterms with respective equal representatives. The claim follows by using the invariant Paths.

□

The method `explain` returns a path between its two arguments, if one exists. Depending on the actual type of graph used, this path can be unique or not. The method `inputEqs` for a path in the congruence graph returns the input equations that were used to derive the equality between the first and the last node of the path. For an input equation, this is simply the equation itself. For a deduced equality, this is the set of input equations that were used for deduction. Combining these two methods, the statement `inputEqs(explain(s,t,g),g)` returns an explanation for $E \models s \approx t$ if there is one.

Algorithm 0.9: inputEqs

Input: path p in g
Input: congruence graph g
Output: set of input equations used in p

Let p be $(u_1, l_1, v_1), \ldots, (u_n, l_n, v_n)$
$eqs \leftarrow \emptyset$ **for** $i \leftarrow 1 ton$ **do**
    if $l_i = null$ **then**
        $f(a, b) \leftarrow u_i$
        $f(c, d) \leftarrow v_i$
        $p1 \leftarrow$ explain(a, c, g)
        $p2 \leftarrow$ explain(b, d, g)
        $eqs \leftarrow eqs \cup inputEqs(p1, g) \cup inputEqs(p2, g)$
    else
        $eqs \leftarrow eqs \cup \{l_i\}$
return eqs

In the following, we describe the two types of congruence graphs we support. They differ in the type of graph they use and how explanations are produced.

Equation Graph

A equation graph stores input and deduced equalities in a labeled weighted undirected graph (V, E) with $V \subseteq \mathcal{T}$, $E \subseteq V \times \mathcal{E} \times V \times \mathbb{N}$. The weight for an edge is the number of input equalities used to derive the equality between its two nodes. This number is one for input equalities and the size of the explanation for deduced equalities. Edges are added to the graph, regardless whether the nodes are already connected in the graph. Therefore there is a choice which path the explain method returns. To produce short explanations, the shortest path w.r.t. the edge weights is returned.

Finding the shortest path between two nodes in a weighted graph is not trivial. The single source shortest path problem (SSSP) is a classical graph problem in computer science. The task is to find the shortest path in a graph between one designated node, the source, and all other nodes in the graph. To the best knowledge of the authors, there is no algorithm to find the shortest path between two nodes which has better asymptotic runtime than one to solve SSSP. There is a whole variety of algorithms that solve SSSP. Classical algorithms for SSSP are those of Dijkstra [16] and Bellman-Ford [4, 21]. The algorithms work on different kinds of graphs. Our setting is an undirected graph with positive integer weights. We chose to use Dijkstra's algorithm, even though the algorithm does not have optimal asymptotic runtime. It's worst-case runtime is $O(n \log(n))$ [14], if the priority queue is implemented as a Fibonacci Heap, which is the case in our implementation. [40] reports of an linear time algorithm for the undirected single source shortest path with positive integer weights problem. However, the algorithm has a big overhead and needs several precomputations. [12] is an extensive study of several shortest path algorithms which shows that Dijkstra's algorithm performs well in practice.

Dijkstra's algorithm finds shortest paths to an increasing set of nodes, until every node has been discovered. It does so by keeping track of the the shortest paths and the distances, being the combined weights of edges on the path, of nodes to the source. Initially, the only discovered node is the source itself and the distance to every other node is infinite. The algorithm discovers new nodes by selecting the lowest weight outgoing edge of all nodes that have been discovered so far and updates shortest paths and distances while doing so. It is a greedy algorithm in the sense that it always locally chooses lowest weight edges and never discards previously made decisions.

The algorithm has been slightly modified to take into account decisions that are edges for deduced equalities. These edges represent explanations, which are a sets of input equations. Previously included input equations do not increase the size of the global explanation when including them again. Therefore the modified Dijkstra algorithm adds an edge with weight 0 for every input equation in the explanation of a deduced equality edge. This is done to reduce the size of explanations. Since previous decisions are not discarded, it is not guaranteed that the modified algorithm returns the shortest path in the final graph, including the extra edges. Example 0.0.6 demonstrates that the modified shortest path algorithm does not always produce the shortest explanation, but can produce shorter explanations than the unmodified version in some situations. The shortest path algorithm's inability to return shortest explanations is not surprising, since it runs in $O(n \log(n))$ and in Section **??** it was shown that finding the shortest explanation is NP-complete.

Example 0.0.6. Consider the congruence graph shown in Figure 9, where solid edges are input equation and the dashed edge marks an application of the congruence axiom. The equality of $f(c_1, e)$ and $f(c_4, e)$ was deduced using the equations $(c_1, c_2), (c_2, c_3), (c_3, c_4)$, which is the shortest path in the graph between c_1 and c_4, obtained from a previous call to the shortest path algorithm.

Suppose we want to compute an explanation for $a \approx b$. Clearly the input equalities $(a, f(c_1, e))$, $(f(c_4, e), c_1)$ and the explanation for $f(c_1, e) \approx f(c_4, e)$ have to be included in the explanation. Additionally $c_1 \approx b$ has to be explained. For this equality the set $(c_1, d_1), (d_1, d_2), (d_2, b)$ is the shortest explanation in the original graph. This sub explanation adds three new equations to the explanation for $a \approx b$. Therefore when the shortest path algorithm iterates over the edge $(f(c_1, e), f(c_4, e))$, it can add add zero weight edges $(c_1, c_2), (c_2, c_3), (c_3, c_4)$ to the graph. By doing so the shortest explanation for $c_1 \approx b$ becomes $(c_1, c_2), (c_2, c_3), (c_3, c_4), (c_4, b)$, which only adds one extra equation to the global explanation.

This method is successful in finding the shortest explanation in this example if the search begins in the node a. Should the search begin in the node b, the edges including d_1, d_2 are added to the shortest path before the edge $(f(c_1, e), f(c_4, e))$ is touched. Therefore the undesired long explanation would be returned.

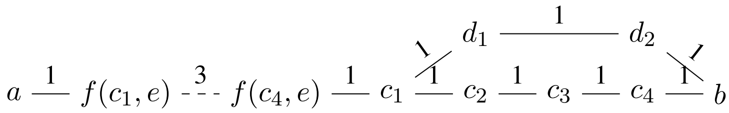

Figure 9: Short explanation example

Algorithm 0.10: insert (equation graph)

Input: term s
Input: term t
Input: equation $eq \in \mathcal{E}$
if $eq! = null$ **then**
  add edge $(s, (eq, \emptyset), t, 1)$ to g
else
  $f(a, b) \leftarrow s$
  $f(c, d) \leftarrow t$
  $p1 \leftarrow$ shortest path between a and c in g
  $p2 \leftarrow$ shortest path between b and d in g
  $w \leftarrow \#(p1.inputEqs \cup p2.inputEqs)$
  add edge $(s, (null), t, w)$

Algorithm 0.11: explain

Input: term s
Input: term t
Input: equation graph g
Output: Path in g
return shortest path between s and t in g

Proof Forest

A proof forest is a collection of proof trees. A proof tree is a labeled tree with nodes in $\mathcal{T}$ and edge labels in $\mathcal{E}$. For every congruence class in the current status of a congruence structure, there is one proof tree. Inserting an edge between nodes s and t of different proof trees is done by making one the child of the other. To maintain a tree structure, all edges between the new child and the root of its tree are reversed. To limit the number of edge revering steps, the smaller tree is always attached to the bigger one. This results in $O(n\log(n))$ edge reversing steps, where n is the number terms in the input equation set. This bound can be shown using the same argument as in the proof of Proposition 0.0.12. As stated earlier, we understand a path as a sequence of undirected edges. In case of a proof tree, a path between s and t of the same tree is the combined sequence of edges between the nodes and their nearest common ancestors. The structure, up to small changes, was proposed by [31, 32]. Its benefit is the quick access of explanations and good overall runtime. Its downside is its inflexibility when it comes to producing alternative explanations. In fact the explanation returned is always the first one to occur during edge insertion. The authors improve the structure for the special case of flattened terms, for which no term has nesting depth greater than one.

Do I cite [31, 32] too often? Should I only cite their newer paper?

Algorithm 0.12: insert (proof forest)

Input: term s
Input: term t
Input: equation $eq \in \mathcal{E}$
if *s is not in g* **then**
  add tree with single node s
if *t is not in g* **then**
  add tree with single node t
$sSize \leftarrow$ size of tree of s
$tSize \leftarrow$ size of tree of t
if $sSize \leq tSize$ **then**
  $(u, v) \leftarrow (s, t)$
else
  $(u, v) \leftarrow (t, s)$
reverse all edges on the path between u and its root node
insert edge (v, eq, u)

Example 0.0.7. Add equations from example 0.0.6 with d - terms first into structure and show why equation graph produces longer proof.

Algorithm 0.13: explain

Input: term s
Input: term t
Input: proof forest g

if *s and t are in the same proof tree P* **then**
  Let nca be the nearest common ancestor of s and t in P
  $p1 \leftarrow$ path from s to nca
  $p2 \leftarrow$ path from nca to s
  return $p1 :: p2$
else
  return the empty path

Proof Production

In this section we describe how to produce resolution proofs from paths in a congruence graph. The method to carry out this operation is `prodcureProof`. The basic idea is to traverse the path, creating a transitivity chain of equalities between adjacent nodes, while keeping track of the deduced equalities in the chain. From invariant Deduced Edges follows that for the deduced equalities there have to be paths between the respective arguments of the compound terms. These paths are transformed into proof recursively and resolved with a suiting instance of the congruence axiom. Afterwards the subproof is resolved with the original transitivity chain. Since terms can never be equal to their subterms, the procedure will eventually terminate. The result of this procedure is a resolution proof with a root, such that the equations of the negative literals are an explanation of the target equality. In other words, let $s \approx t$ be the equality to be explained and suppose `produceProof` returns a proof with root ρ. Then for ρ it is the case that $F := \{(u, v) \mid u \neq v \text{ is a literal in } \rho\} \models s \approx t$ and F is a subset of the input equations.

Example 0.0.8. Consider again the congruence graph shown in Figure 9 and suppose we want a proof for $a \approx b$. Suppose we found the path $p_1 := \langle a, f(c_1, e), f(c_4, e), c_1, c_2, c_3, c_4, b\rangle$ as an explanation and that the explanation for $f(c_1, e) \approx f(c_4, e)$ is the path $\langle c_1, c_2, c_3, c_4\rangle$. We transform p_1 and p_2 into instances of the transitivity axiom C_1 and C_2 respectively. The clause C_2 is resolved with the instance of the congruence axiom C_3, which is then resolved with the instance of the reflexive axiom C_4 resulting in clause C_5. Finally, C_1 is resolved with C_5 to obtain the final clause C_6. The proof is shown in Figure 10.

Figure 10: Example proof

Algorithm 0.14: produceProof

```
Input: term s
Input: term t
Output: Resolution proof for s = t or null
p ← explain(s, t, g)
d ← ∅
e ← ∅
while p is not empty do
    (u, l, v) ← first edge of p
    p ← p \ (u, l, v)
    e ← e ∪ {u ≠ v}
    if l = null then
        f(a, b) ← u
        f(c, d) ← v
        p1 ← produceProof(a, c, g)
        p2 ← produceProof(b, d, g)
        con ← {a ≠ c, b ≠ d, f(a, b) = f(c, d)}
        res ← resolve con with non null roots of p1 and p2
        d ← d ∪ res
if #e > 1 then
    proof ← e ∪ {s = t}
    while d is not empty do
        int ← some element in d
        d ← d \ {int}
        proof ← resolve proof with int
    return proof
else if d = {ded} then
    return ded
else
    if e = {(u, l, u)} then
        return {u = u}
    else
        return null
```

Congruence Compressor

In Section **??** processing of a proof was defined. The most important application of proof processing for this work is proof compression. We want to make use of the short explanations found by the congruence closure algorithm described above. To this end we replace subproofs with conclusions that contain unnecessary long explanations with new proofs that have shorter conclusions. Shorter conclusions lead to less resolution steps further down the proof and possibly big chunks of the proof can simply be discarded. There is however a tradeoff in overall proof length when introducing new subproofs. The subproof corresponding to a short explanation can be longer in proof length, i.e. involve more resolution nodes, than one with a longer explanation. Example 0.0.9 displays this issue. Additionally it can be the case that by introducing a new subproof, we only partially remove the old subproof. Some nodes of the old subproof might still be used in other parts of the proof. Therefore the replacement of a subproof by another, smaller one does not necessarily lead to a smaller proof. Nevertheless, the meta heuristic favoring smaller conclusions should still dominate such effects, especially on large proofs. The results in Section **??** confirm this intuition.

Example 0.0.9. Consider the set of equations $E = \{(f(f(a,b), f(a,a)), a), (a,b), (b, f(f(b,a), f(b,b)))\}$ and the target equality $f(f(a,b), f(a,a)) \approx f(f(b,a), f(b,b))$. For presentation purposes, throughout this example we will abbreviate the term $f(f(a,b), f(a,a))$ with t_a and $f(f(b,a), f(b,b))$ with t_b. Using equations in E, one can prove the the target equality in two ways. Either one uses the instance of the transitivity axiom $\{t_a \neq a, a \neq b, b \neq t_b, t_a = t_b\}$ or a repeated applications of instances of the congruence axiom, e.g. $\{a \neq b, f(a,a) = f(b,b)\}$. The corresponding explanations are E and $\{(a,b)\}$.

The two resulting proofs are shown in Figure 11. The proof with the longer explanation E is only one proof node, whereas the proof with the singleton explanation has proof length 5.

Figure 11: Short explanation, long proof

The Congruence Compressor compresses processes a proof replacing subproofs as described above. It is defined upon the processing function specified in pseudocode in Algorithm 16. The idea of the processing function is simple. Axioms are not changed by the function. For all other nodes the `fixNode` is called method, to maintain a correct proof. Then in line 4 it is decided whether the explanation finding congruence closure algorithm should be used to find a replacement for the current node. One trivial criteria could be true for every node. Testing every node will result in a slow algorithm, but the best possible compression. Some nodes do not need to be checked, since they contain optimal explanations by definition or there is no hope of finding an explanation at all. The following definition classifies nodes to define a more sophisticated decision criteria.

Definition 0.0.14 (Types of nodes)**.** An axiom is a *theory lemma* if it is an instance of one of the congruence axioms. Otherwise it is called *input derived*. The classification of internal nodes is defined recursively. An internal node is input derived, if one of its premises is input derived. Otherwise it is a theory lemma. We call a node a *low theory lemma* if it is a theory lemma and has a child that is input derived.

We suspect that most redundancies in proofs are to be found in low theory lemmas, since they reflect the explanations found by the proof producing solver. Therefore an alternative criteria is to only find replacements for low theory lemmas. The question whether a node is a low theory lemma is not trivial to answer while traversing the proof in a top to bottom fashion.

Therefore a preliminary traversal is necessary to determine the classification of nodes. Experiments have shown that using this criteria speeds up the algorithm a lot, while losing only very little compression.

Add all equations of the antecedent to an empty congruence structure and check whether these equations induce a proof for one of the equations in the succedent that has a shorter conclusion than the original subproof. If there is such a proof, we replace the old subproof by the new one. Futher criteria for deciding whether to replace or not could be size of the subproof or a global metric that tries to predict the global compression achieved by replacement.

Algorithm 0.15: compress

Global: Set of input equations E
Input: resolution node n
Input: pr : tuple of resolution nodes (p_1, p_2) or null
Output: resolution node

if $pr = null$ **then**
    return n
else
    $m \leftarrow fixNode(n, (p_1, p_2))$
    if m *fulfills criteria* **then**
        $lE \leftarrow \{(a, b) \mid (a \neq b) \in m\}$
        $rE \leftarrow \{(a, b) \mid (a = b) \in m\}$
        $con \leftarrow$ empty congruence structure
        for (a, b) *in* lE **do**
            $con \leftarrow con.addEquality(a, b)$
        for (a, b) *in* rE **do**
            $con \leftarrow con.addNode(a).addNode(b)$
            $proof \leftarrow con.prodProof(s, t)$
            if $proof \neq null$ *and* $|proof.conclusion| < |m.conclusion|$ **then**
                $m \leftarrow proof$
    return m

The compressor (Algorithm 16) uses the method `fixNode` to maintain a correct proof. The method modifies nodes with premises that have earlier been replaced by the compressor. Nodes with unchanged premises are not changed. Let n be a proof node that was derived using pivot ℓ in the original proof and which updated premises are pr_1 and pr_2 . Depending on the presence of ℓ in pr_1 and pr_2, n is either replaced by the resolvent of pr_1 and pr_2 or by one of the updated premises. It assumed that the values pr_1, pr_2 and ℓ are stored together with the node and can be accessed in constant time. In case both updated premises do not contain the original pivot element, replacing the node by either one of them maintains a correct proof. Since we are interested in short proofs, we return the one with the shorter clause. This method of maintaining a correct proof was proposed in [2] in the context of similar proof compression algorithms.

Algorithm 0.16: fixNode

Input: resolution node n
Input: pr : tuple of resolution nodes (p_1, p_2) or null
Output: resolution node

if $pr = null$ *or* $(n.premise_1 = p_1$ *and* $n.premise_2 = p_2)$ **then**
  return n
else
  if $n.pivot \in p_1$ *and* $n.pivot \in p_2$ **then**
    return $resolve(p_1, p_2)$
  else if $n.pivot \in p_1$ **then**
    return p_2
  else if $n.pivot \in p_2$ **then**
    return p_1
  else
    return node with smaller clause

Future Work

[1] compares the running times of several congruence closure algorithms. It would be interesting to do a similar comparison including the congruence closure algorithm presented in Section 7. A comparison to the classic congruence closure algorithms of Nelson and Oppen [30], Downey, Sethi and Tarjan [17] and Shostak [39] and their abstract counterparts, as described in [1], would show whether our method can compete in terms of computation speed. Comparing our method with the explanation producing algorithms presented in [19] and [31,32] could be done not only in terms of speed, but also in terms of explanation size.

In Section 7 it was shown that the problem of finding the shortest explanation is NP-complete. Therefore methods and heuristics to find short explanations could be investigated. The idea of using shortest path algorithms for explanation finding is a step in that direction. In **??** we describe a modification of Dijkstra's algorithm [**?**] to make it sensitive to previously used equations. Further modifications, possibly using heuristics, could lead to a short explanation algorithm. Furthermore translating the problem into a SAT instance could result in an algorithm to derive shortest explanations in acceptable time.

The congruence closure algorithm could be implemented into a SMT solver. Such solvers usually have high requirements regarding computation time. It would be interesting to see, whether the method presented in this work can match these requirements.

[32] extends the congruence closure algorithm to integer offsets; this could be incorporated

Conclusions

Conclusions

Literature Search

Information on online libraries and literature search, e.g., interesting magazines, journals, conferences, and organizations may be found at `http://www.big.tuwien.ac.at/teaching/info.html`.

BibTeX

BibTeX should be used for referencing.

Blabla article1 [**?**], laber laber article2 [**?**].

Bibliography

[1] Leo Bachmair and Ashish Tiwari. Abstract congruence closure and specializations. In *CADE*, pages 64–78, 2000.

[2] Omer Bar-Ilan, Oded Fuhrmann, Shlomo Hoory, Ohad Shacham, and Ofer Strichman. Linear-time reductions of resolution proofs. In *Haifa Verification Conference*, pages 114–128, 2008.

[3] Henk Barendregt. The impact of the lambda calculus in logic and computer science. *Bulletin of Symbolic Logic*, 3(02):181–215, 1997.

[4] Richard Bellman. On a routing problem. Technical report, DTIC Document, 1956.

[5] Eli Ben-Sasson. Size space tradeoffs for resolution. In *STOC*, pages 457–464, 2002.

[6] Eli Ben-Sasson and Jakob Nordström. Short proofs may be spacious: An optimal separation of space and length in resolution. *Electronic Colloquium on Computational Complexity (ECCC)*, 16:2, 2009.

[7] Armin Biere. Tracecheck resolution proof format, 2006.

[8] Armin Biere. Picosat essentials. *JSAT*, 4(2-4):75–97, 2008.

[9] Armin Biere, Marijn Heule, Hans van Maaren, and Toby Walsh, editors. *Handbook of Satisfiability*, volume 185 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2009.

[10] Thomas Bouton, Diego Caminha B. de Oliveira, David Déharbe, and Pascal Fontaine. verit: An open, trustable and efficient smt-solver. In *CADE*, pages 151–156, 2009.

[11] Siu Man Chan. Pebble games and complexity. 2013.

[12] Boris V. Cherkassky, Andrew V. Goldberg, and Tomasz Radzik. Shortest paths algorithms: Theory and experimental evaluation. *Math. Program.*, 73:129–174, 1996.

[13] Alonzo Church. An unsolvable problem of elementary number theory. *American journal of mathematics*, pages 345–363, 1936.

[14] Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms*. The MIT Press and McGraw-Hill Book Company, 1989.

[15] Haskell B. Curry. *Combinatory Logic*. Amsterdam, North-Holland Pub. Co., 1958.

[16] Edsger W Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.

[17] Peter J. Downey, Ravi Sethi, and Robert Endre Tarjan. Variations on the common subexpression problem. *J. ACM*, 27(4):758–771, 1980.

[18] Juan Luis Esteban and Jacobo Torán. Space bounds for resolution. *Inf. Comput.*, 171(1):84–97, 2001.

[19] Pascal Fontaine. *Techniques for verification of concurrent systems with invariants*. PhD thesis, PhD thesis, Institut Montefiore, Université de Liege, Belgium, 2004.

[20] Pascal Fontaine, Stephan Merz, and Bruno Woltzenlogel Paleo. Compression of propositional resolution proofs via partial regularization. In *CADE*, pages 237–251, 2011.

[21] Lester Randolph Ford. Network flow theory. 1956.

[22] John R. Gilbert, Thomas Lengauer, and Robert Endre Tarjan. The pebbling problem is complete in polynomial space. *SIAM J. Comput.*, 9(3):513–524, 1980.

[23] Philipp Hertel and Toniann Pitassi. Black-white pebbling is pspace-complete. *Electronic Colloquium on Computational Complexity (ECCC)*, 14(044), 2007.

[24] Marijn Heule. Drup proof format, 2007.

[25] Marijn Heule. Bdrup proof format, 2013.

[26] Georg Hofferek, Ashutosh Gupta, Bettina Könighofer, Jie-Hong Roland Jiang, and Roderick Bloem. Synthesizing multiple boolean functions using interpolation on a single proof. *CoRR*, abs/1308.4767, 2013.

[27] Takumi Kasai, Akeo Adachi, and Shigeki Iwata. Classes of pebble games and complete problems. *SIAM J. Comput.*, 8(4):574–586, 1979.

[28] Boris Konev and Alexei Lisitsa. A sat attack on the erdos discrepancy conjecture. *CoRR*, abs/1402.2184, 2014.

[29] Kenneth L. McMillan. Applications of craig interpolants in model checking. In *TACAS*, pages 1–12, 2005.

[30] Greg Nelson and Derek C. Oppen. Fast decision procedures based on congruence closure. *J. ACM*, 27(2):356–364, 1980.

[31] Robert Nieuwenhuis and Albert Oliveras. Proof-producing congruence closure. In *RTA*, pages 453–468, 2005.

[32] Robert Nieuwenhuis and Albert Oliveras. Fast congruence closure and extensions. *Inf. Comput.*, 205(4):557–580, 2007.

[33] Jakob Nordström. Narrow proofs may be spacious: Separating space and width in resolution. *SIAM J. Comput.*, 39(1):59–121, 2009.

[34] Bruno Woltzenlogel Paleo, Joseph Boudou, and Andreas Fellner. Skeptik system description.

[35] Nicholas Pippenger. Comparative schematology and pebbling with auxiliary pushdowns (preliminary version). In *STOC*, pages 351–356, 1980.

[36] Nicholas Pippenger. *Advances in pebbling*. Springer, 1982.

[37] Moses Schönfinkel. Über die bausteine der mathematischen logik. *Mathematische Annalen*, 92(3):305–316, 1924.

[38] Ravi Sethi. Complete register allocation problems. *SIAM J. Comput.*, 4(3):226–248, 1975.

[39] Robert E. Shostak. An algorithm for reasoning about equality. *Commun. ACM*, 21(7):583–585, 1978.

[40] Mikkel Thorup. Undirected single-source shortest paths with positive integer weights in linear time. *J. ACM*, 46(3):362–394, 1999.

[41] Peter van Emde Boas and Jan van Leeuwen. Move rules and trade-offs in the pebble game. In *Theoretical Computer Science*, pages 101–112, 1979.

[42] S. A. Walker and H. Raymond Strong. Characterizations of flowchartable recursions. *J. Comput. Syst. Sci.*, 7(4):404–447, 1973.