



Final year internship report




Verification in Isabelle/HOL of Hopcroft's algorithm for minimizing DFAs including runtime analysis

Vincent Trélat

29th August 2023



Contents

1	Preamble	2
2	Technical University of Munich	3
3	Introduction	3
3.1	The Refinement Framework	5
3.2	Setting up the environment 	5
3.3	Hopcroft's algorithm	6
3.4	Original formalization 	7
3.5	Modern formalization	7
4	Overview of the project	9
4.1	Objectives of the project	9
4.2	Gantt chart	11
4.3	Main difficulties	12
5	Proof of correctness	12
6	Time complexity analysis	13
6.1	Introduction	13
6.2	Paper proof	14
6.3	Towards a formal proof 	21
6.3.1	Abstract level	21
7	Further work	32
8	Conclusion	33

1 Preamble

This report is the result of a 6-month internship at the Chair of Logic and Verification of the Technical University of Munich (TUM) under the supervision of [Tobias Nipkow](#) and [Peter Lammich](#).

The main goal of this internship was to verify Hopcroft’s algorithm for minimizing DFAs [[Hop71](#)] in Isabelle/HOL and to analyze its worst-case time complexity using the Refinement Framework [[Lam16](#)]. Some considerable work had already been carried out by Peter Lammich and Thomas Türk on the proof of correctness ten years ago in an older version of Isabelle [[LT12](#)]. The goal was to port this proof to the latest version of Isabelle and to extend it with a time complexity analysis using the NREST framework, namely the extension of the Refinement Framework to time [[HL19](#)].

In order to bring the reader closer to the reality of the project, some sections of this report dive a little deeper into the details of the interesting parts of my research. They are marked with the following gear symbol ⚙️ and can be skipped without loss of understanding of the main ideas.

2 Technical University of Munich

The Technical University of Munich (TUM) stands as a beacon of academic excellence, renowned globally for its commitment to cutting-edge research, innovation, and education. Nestled in the heart of Germany, TUM’s legacy spans over 150 years, consistently ranking among the top universities worldwide¹. At TUM, scholars and students alike converge to explore and shape the forefront of various fields, fostering a dynamic environment that thrives on intellectual curiosity and collaboration.

Within the expansive landscape of TUM’s academic endeavors lies the Chair of Logic and Verification – a distinguished hub of expertise dedicated to the study and application of formal methods – which operates at the intersection of computer science, mathematics, logic and engineering, harnessing the power of logical reasoning to ensure the precision and reliability of complex systems.

Among the illustrious figures leading the charge at the Chair of Logic and Verification is Professor Tobias Nipkow. A luminary in the realm of formal methods, he has made indelible contributions to the development and propagation of logical verification techniques.

Central to the Chair’s accomplishments is the Isabelle proof assistant – an innovation born from the ingenuity of Professor Tobias Nipkow and his dedicated team. Isabelle stands as a testament to the power of collaborative intellect, serving as a versatile tool for rigorous formal reasoning. This proof assistant facilitates the creation, manipulation, and verification of mathematical proofs, underpinning the very essence of formal methods’ application. Its impact reverberates not only through academic corridors but also through the industries and systems that rely on robust verification.

In summary, the Technical University of Munich’s Chair of Logic and Verification represents an ecosystem of brilliance, where researchers, developers, and applied experts converge to elevate the tenets of formal methods. Professor Tobias Nipkow’s visionary leadership and the Isabelle proof assistant embody the spirit of innovation that defines this dynamic academic haven, propelling the boundaries of formal reasoning and verification.

3 Introduction

Since the whole project relies on many deeply theoretical notions such as abstraction, logic or semantics, this section only gives a brief introduction to formal methods in general and their purpose.

¹TUM is ranked first in Europe in Computer Science and Mathematics.

Formal methods refer to a field of theoretical computer science whose main purpose is to provide logical links between mathematics and programming languages, so that one can state logical properties about formally described algorithms or mathematical procedures and relate them to actual implementations. The main application of formal methods is to prove correctness of programs with respect to a specification, i.e. a set of rules which must hold throughout the execution of the program. Such a program is then ensured to behave in a way² that is intended by its specification.

Such work is usually carried out using a proving assistant such as [Isabelle](#), i.e. a program with an implemented logic allowing to reason with respect to that logic. Isabelle has the advantage to rely on a really light kernel, which is the only part that can only be trusted. Everything else – i.e. libraries, called theories – stems from this kernel. It also has nice interface with \LaTeX syntax adding little overhead to usual mathematical notations and is very flexible in terms of syntax and semantics.

A common way to proceed when we want to prove correctness of an algorithm and generate correct executable code is to follow a top-down strategy via successive refinements. We start from the highest level – i.e. with the highest level of abstraction – and we refine towards the implementation. As an example, if an algorithm manipulates sets at the abstract level, we may use lists or trees in its implementation. When performing a refinement, we have to give links – and prove them – between levels, e.g. that the behavior of the refined one does not contradict the abstract one. In Peter Lammich’s Refinement Framework [\[Lam16\]](#), this is done via binary relations and abstraction/concretization functions³. The Refinement Framework is briefly introduced in the next section.

If formal methods allow for correctness and termination theorems, it is less trivial to analyze complexity of a program, e.g. in terms of time or memory consumption. Since complexity of a program is not a logical property, but a property of the execution of the program, a model of computation must be defined, i.e. a way to execute the program, and then prove that the execution of the program in this model is equivalent to the execution of the program in the real world. Maximilian Haslbeck and Peter Lammich have extended

²In the context of industrial applications, we often differentiate between two sorts of assertions, namely safety properties and liveness properties, so that we can explicitly check that the system works (it is “alive”) and that it is safe (it works properly).

³A concretization function maps an abstract value to a set of concrete values and conversely for an abstraction function, so that invariants can be carried and preserved through the refinements. For example, we may abstract the set \mathbb{Z} of integers to $\{Neg, Zero, Pos\}$ where *Pos* (resp. *Neg*) represents positive (resp. negative) integers and *Zero* represents the singular set $\{0\}$.

the Refinement Framework to include worst-case time complexity analysis in Isabelle [HL19] and have also implemented a verified framework to analyze algorithms down to LLVM code [HL22]. The principle is to model costs of operations with resources and currencies and embed this new dimension in the program, where operations consume resources.

Since the project also relies on automata theory, it is recommended to have some basic knowledge about automata theory and formal languages.

3.1 The Refinement Framework

The Refinement Framework [Lam16] is a methodology for designing and verifying computer systems, particularly in critical applications like aviation and healthcare. It was mostly developed by Peter Lammich and Maximilian P. L. Haslbeck, often in collaboration with tools like Isabelle.

At its core, the Refinement Framework provides a systematic way to start with a high-level, abstract description of a system and gradually refine it into a concrete implementation. It works the same way one would build a house. First, a basic blueprint with the overall design and room layouts is designed: it is the abstract specification. Then, this plan is gradually refined by adding more detail at each step. For instance, the materials, dimensions, and exact placement of walls, plumbing, and electrical systems might be specified. Each refinement step brings the model closer to the actual construction until a fully detailed plan is obtained and ready for builders to follow.

Software development, is completely analogous. We start with a high-level description of what a software system should do (the abstract specification), and it is refined through several stages, adding more and more detail until we have a concrete and reliable software implementation. This process helps ensure that the final software system matches what one originally intended, and it provides a way to verify that it works correctly.

The key benefit of the Refinement Framework is that it helps catch errors and problems early in the development process, reducing costly mistakes and improving the quality of the final product. It's a systematic way to bridge the gap between abstract concepts and real-world software, making sure they align properly.

3.2 Setting up the environment

In order to be able to comfortably browse the theories and inspect the proofs, it is recommended to use Isabelle⁴/jEdit. Yet, the project is not part of

⁴The project was carried out using Isabelle2022.

the [Archive of Formal Proofs](#) (AFP) and can only be downloaded from the [GitHub repository](#).

Then, building an image of the main framework is strongly recommended, as it will greatly speed up the compilation of the theories, especially for the time complexity analysis. This may take a few minutes, but only needs to be done once:

```
cd Hopcroft_Verif/Hopcroft_Time
isabelle build -d . -l Isabelle_LLVM_Time Hopcroft_Time.thy
```

Isabelle may then be launched with jEdit by running the following command. This will open the main theory file `Hopcroft_Time.thy` in jEdit and should take less than a minute:

```
isabelle jedit -d . -l Isabelle_LLVM_Time Hopcroft_Time.thy
```

3.3 Hopcroft’s algorithm

John E. Hopcroft’s algorithm for minimizing deterministic finite automata (DFA) was first presented in his original 1971 paper [Hop71] as a formal algorithm. It has been a major breakthrough in the field of automata theory, since it allowed for minimization in linearithmic⁵ time in the worst case⁶. Thanks to this, minimization has become almost costless and thus this algorithm has been widely used in the field of formal verification, e.g. in model checking, since it allows to reduce the size of the automaton to be verified.

The formal idea behind the algorithm is to partition the set of states of the DFA into equivalence classes, i.e. sets of states that are equivalent⁷ with respect to the language they recognize. The algorithm starts with two partitions, namely the set of final states and the set of non-final states. Then, it iteratively refines the partitions by splitting them into smaller ones until no more refinement is possible. The algorithm is guaranteed to terminate since the number of partitions is finite and strictly decreases at each iteration. The equivalence class of the set of states of the DFA is then the set of partitions, so that any two different sets of states from that partition recognize different languages.

⁵ $O(n \log n)$

⁶Other minimization algorithms were presented prior to Hopcroft, e.g. by Moore (quadratic) or Brzozowski (exponential).

⁷Two states are equivalent if they accept the same language.

3.4 Original formalization

Algorithm 1: Hopcroft's original formal algorithm

Data: a finite DFA $\mathcal{A} = (\mathcal{Q}, \Sigma, \delta, q_0, \mathcal{F})$
Result: the equivalence class of \mathcal{Q} under state equivalence

- 1 Construct $\delta^{-1}(q, a) := \{t \in \mathcal{Q} \mid \delta(t, a) = q\}$ for all $q \in \mathcal{Q}$ and $a \in \Sigma$;
- 2 Construct $P_1 := \mathcal{F}$, $P_2 := \mathcal{Q} \setminus \mathcal{F}$ and $s_{a,i} := \{q \in P_i \mid \delta^{-1}(q, a) \neq \emptyset\}$
for all $i \in \{1, 2\}$ and $a \in \Sigma$;
- 3 Let $k := 3$;
- 4 For all $a \in \Sigma$, construct $L_a := \arg \min_{1 \leq i < k} |s_{a,i}|$;
- 5 **while** $\exists a \in \Sigma, L_a \neq \emptyset$ **do**
- 6 Pick $a \in \Sigma$ such that $L_a \neq \emptyset$ and $i \in L_a$;
- 7 $L_a := L_a \setminus \{i\}$;
- 8 **forall** $j < k, \exists q \in P_j, \delta(q, a) \in s_{a,i}$ **do**
- 9 $P'_j := \{t \in P_j \mid \delta(t, a) \in s_{a,i}\}$ and $P''_j := P_j \setminus P'_j$;
- 10 $P_j := P'_j$ and $P_k := P''_j$; construct $s_{a,j}$ and $s_{a,k}$ for all $a \in \Sigma$
accordingly ;
- 11 For all $a \in \Sigma$, $L_a := \begin{cases} L_a \cup \{j\} & \text{if } j \notin L_a \wedge |s_{a,j}| \leq |s_{a,k}| \\ L_a \cup \{k\} & \text{otherwise} \end{cases}$;
- 12 $k := k + 1$;
- 13 **end**
- 14 **end**

Algorithm 1 is a direct translation of the original algorithm from [Hop71] with only slight changes in the notations.

3.5 Modern formalization

The algorithm is now usually given in a more mathematical and formalised way⁸, which loosens up the choice for an implementation. The algorithm is given below in Algorithm 2 using sets and partitions.

⁸See for example [EB23].

Algorithm 2: Hopcroft's algorithm in a modern style

Data: a DFA $\mathcal{A} = (\mathcal{Q}, \Sigma, \delta, q_0, \mathcal{F})$
Result: the language partition \mathcal{P}

```

1 if  $\mathcal{F} = \emptyset \vee \mathcal{Q} \setminus \mathcal{F} = \emptyset$  then
2   return  $\mathcal{Q}$ 
3 else
4    $\mathcal{P} := \{\mathcal{F}, \mathcal{Q} \setminus \mathcal{F}\}$  ;
5    $\mathcal{W} := \{(a, \min\{\mathcal{F}, \mathcal{Q} \setminus \mathcal{F}\}), a \in \Sigma\}$  ;
6   while  $\mathcal{W} \neq \emptyset$  do
7     Pick  $(a, C)$  from  $\mathcal{W}$  and remove it;
8     forall  $B \in \mathcal{P}$  do
9       Split  $B$  with  $(a, C)$  into  $B_0$  and  $B_1$  ;
10       $\mathcal{P} := (\mathcal{P} \setminus \{B\}) \cup \{B_0, B_1\}$  ;
11      forall  $b \in \Sigma$  do
12        if  $(b, B) \in \mathcal{W}$  then
13           $\mathcal{W} := (\mathcal{W} \setminus \{(b, B)\}) \cup \{(b, B_0), (b, B_1)\}$  ;
14        else
15           $\mathcal{W} := \mathcal{W} \cup \{(b, \min\{B_0, B_1\})\}$  ;
16        end
17      end
18    end
19  end
20 end

```

We justify this transformation. First, instead of storing indices for the blocks, we directly deal with sets and explicitly work with a partition \mathcal{P} of \mathcal{Q} . Then, we define splitters.

Definition 1. Let $\mathcal{A} = (\mathcal{Q}, \Sigma, \delta, q_0, I, F)$ be a DFA. Let P be a partition of \mathcal{Q} . Let $B \in P$. For $a \in \Sigma$ and $C \in P$ we say that (a, C) splits B or is a splitter of B if

$$\exists q_1, q_2 \in B \quad \delta(q_1, a) \in C \wedge \delta(q_2, a) \notin C.$$

Remark 1. If (a, C) is a splitter of B , P can be updated to $P \setminus \{B\} \cup \{B', B''\}$, where

$$B' := \{q \in B \mid \delta(q, a) \in C\} \text{ and } B'' := B \setminus B'.$$

Definition 2. For a splitter $s \in \Sigma \times P$, we define its size $\|s\|$ as the size of its second component, i.e. if $s = (a, C)$, then $\|s\| := |C|$.

In Algorithm 1, picking an index $i \in L_a$ corresponds to picking a set $s_{a,i}$, i.e. – by definition – the subset of P_i of states having at least one predecessor. We cannot directly replace L_a and $s_{a,i}$ by a set of splitters (a, P_i) as is, however the next step in the algorithm is to go over all blocks in the partition that have a successor in $s_{a,i}$. If such a block B is found, it is split into two blocks, namely the set of states having a successor in $s_{a,i}$ and the others. Overall, we look for B such that P_i has a predecessor in B . We add the condition that B must also have at least one successor not in P_i to avoid splitting B into $\{B, \emptyset\}$. Since this is equivalent to splitting B with (a, P_i) , we can replace L_a by a set of splitters (a, P_i) . Since the symbol a is chosen such that $L_a \neq \emptyset$, we can define a workset \mathcal{W} of all splitters, with all symbols of Σ . Therefore, L_a is empty if and only if there is no a -splitter in \mathcal{W} . Because testing emptiness is equivalent to testing membership, and because of the existential quantifier in the definition of splitters, the $s_{a,i}$ are no longer needed.

We can explicitly write the trivial cases where either \mathcal{F} or $\mathcal{Q} \setminus \mathcal{F}$ is empty to improve performance for this specific case. This is not done in Algorithm 1 but this does not produce wrong results. If one of the two sets is empty, one of the $s_{a,1}$ or $s_{a,2}$ will be empty for all $a \in \Sigma$, and the *while* loop will be entered $|\Sigma|$ times but the inner *for* loop will never be entered. This results in a useless $O(|\Sigma|)$ computation. Another difference in that case is that it would return the partition $\min\{\mathcal{F}, \mathcal{Q} \setminus \mathcal{F}\}$ – where \min returns the set with the least number of elements – which we avoid in the modern version by just returning \mathcal{Q} .

4 Overview of the project

4.1 Objectives of the project

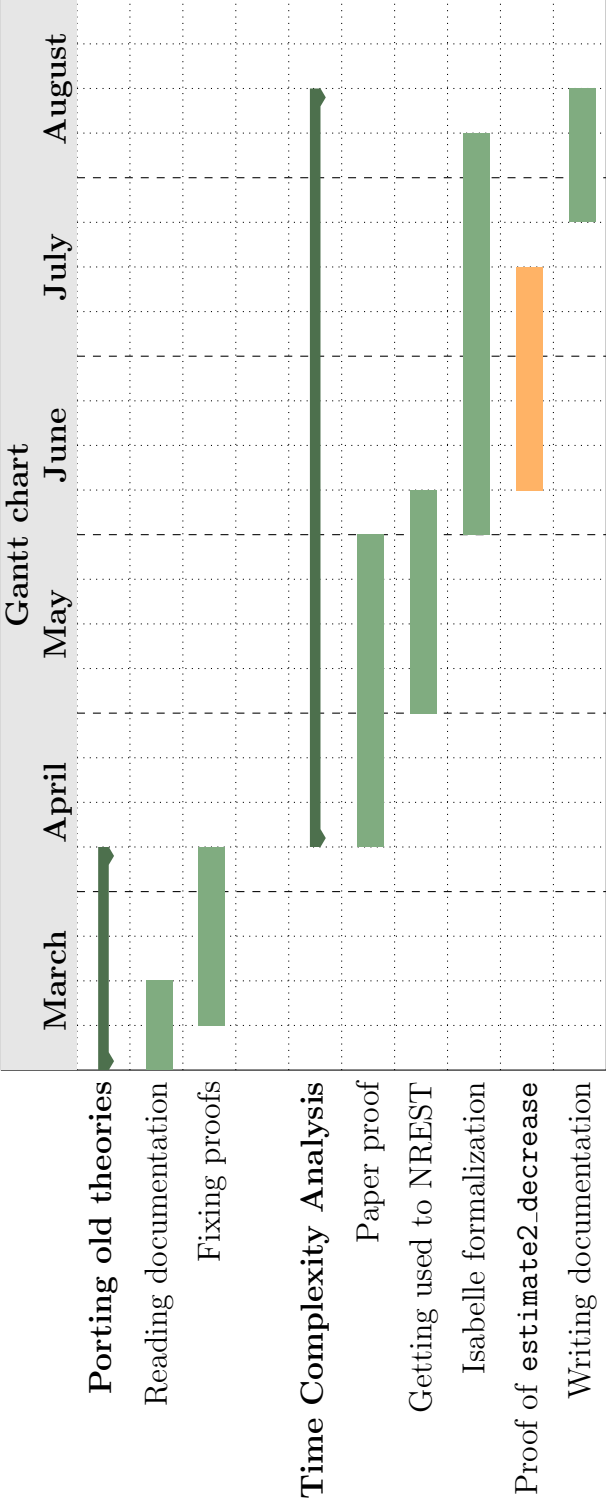
The project can be divided into two main parts:

- (i) porting the old formalization of the algorithm and its proof of correctness to the latest version of Isabelle,
- (ii) extending the proof with a time complexity analysis using the NREST Framework [HL19], i.e. the extension of the Refinement Framework to time.

The first part is rather straightforward since it only involves a good understanding of the original proof and the Refinement Framework. The main proofs are fixed by finding syntax changes and adapting them accordingly.

the second one is more challenging since it requires a good understanding of the algorithm and the Refinement Framework, as well as a good knowledge of the NREST Framework.

4.2 Gantt chart



4.3 Main difficulties

Like any project in formal methods, the main difficulty is to find a good abstraction level for the formalization and a good formalization. Fortunately, I could base mine on the existing one used for the proof of correctness in which the different levels of abstraction – i.e. the different stages of refinement – were already designed down to the implementation, so the choice for the datastructures was even already done.

Of course, paper proofs serve as foundations for the formal proof. I consider them as the “fun” part since paper is the first place where actual research starts. Draft proofs may also be considered as the “easy” part since a lot of details – of implementation, or formal details – are often omitted⁹ during this phase. Finding a good paper proof involves being cautious with potential inaccuracies that would make the formal proof harder at a later stage: it happens that a paper proof seems correct at first but is actually wrong or unprovable formally – at least modulo some changes.

The “real” challenges lie in formalizing the different levels of abstraction: they should be gradual enough to make proofs digestible. My formalization follows the formalization used for the proof of correctness with slight adjustments for the time complexity analysis, so all the trouble I had with the formalization was linked to using the NREST framework properly. It involves theoretical notions of semantics and concepts such as, resources, currencies, and the NREST monad, but also practical notions such as specific commands or tactics in Isabelle that come with the framework.

5 Proof of correctness

Since the proof of correctness is not the purpose of this project, I just sketch the main ideas that are also helpful for the time complexity analysis and stand as a good introduction to the formalization. John E. Hopcroft’s original paper [Hop71] contains a proof of correctness for his low-level algorithm, however the ideas behind the invariants are the same for the formalization. The proof typically involves two main components: partition refinement and equivalence checking.

The basic idea of the proof is to show that the algorithm correctly partitions the states of the input DFA into equivalence classes, where states within the same class are indistinguishable from each other with respect to the language recognized by the DFA. This partition represents the minimal DFA that recognizes the same language as the input DFA.

⁹Be it deliberate or not!

The algorithm starts with an initial partition that distinguishes accepting states from non-accepting states. Then, it iteratively refines this partition by considering pairs of states that are currently in different equivalence classes and determining if they can be distinguished by a symbol from the input alphabet.

During each iteration, the algorithm examines each symbol in the alphabet and checks if it distinguishes any pair of states in different equivalence classes. If a symbol does distinguish a pair, it splits the equivalence class containing those states into two new classes. This process continues until no more splits can be made, and the resulting partition represents the minimal DFA.

To prove the correctness of Hopcroft’s algorithm, one needs to show two key properties:

- the resulting partition is a valid equivalence relation, meaning it satisfies the properties of reflexivity, symmetry, and transitivity,
- the resulting partition is indeed the minimal partition that correctly distinguishes states based on their language equivalence.

The proof typically involves reasoning about the behavior of the algorithm during each iteration, considering how the partition is refined inductively and ensuring that it converges to the minimal partition.

In the Refinement Framework, the algorithm is first described at the abstract level, so its correctness can be proved independently of any implementation details. Then, the algorithm is progressively refined into a concrete implementation, and the correctness of each refinement until the implementation is proved by showing that it refines the abstract algorithm.

6 Time complexity analysis

6.1 Introduction

In order to give a glimpse of my contributions to this project, this section sketches the main ideas behind the time complexity analysis of Hopcroft’s algorithm, since it was the main goal of this internship. First, I give my raw notes on the paper proof. This acts as a pre-structured draft for the formalization in Isabelle. Then, I give a brief overview of the formalization in Isabelle.

The idea of the proof is the following: the algorithm can be reduced to a

loop, so first, we find an upper bound¹⁰ for the time taken for an iteration of this loop. Then, we have to find how many iterations are executed until termination of the algorithm. To do so, we design what is referred to as an “estimate” of the time spent in the loop until its termination. At each step, this estimate is refined with the actual time taken by the previous steps, so that it decreases at each step. Finding a good estimate is the meat of the proof.

6.2 Paper proof

In a first time, we focus on the original algorithm presented in Algorithm 1 in order to work on the arguments given in [Hop71]. The data structures used at that time were mostly linked lists, but let us rather give some requirements for the data structures instead of actual implementations. The goal is to show that the algorithm can be executed in $O(m \cdot n \log n)$ time, where m is the number of symbols in the alphabet and n is the number of states in the DFA.

The following requirements come directly from [Hop71] and are specific to the algorithm presented in Algorithm 1.

Requirement 1. Sets such as $\delta^{-1}(q, a)$ and L_a must be represented in a way that allows $O(1)$ time for addition and deletion in “front position”¹¹.

Requirement 2. Vectors must be maintained to indicate whether a state is in a given set.

Requirement 3. Sets such as P_i must be represented in a way that allows $O(1)$ time for addition and deletion at any given position.

Remark 2. From Req. 2 and Req. 3, we can show that for a state q in the representation of a set P_i or $s_{a,i}$, its “position” – in the implemented data structure – must be determined in $O(1)$ time.

Lemma 1. In Algorithm 1, lines 1 to 4 can be executed in $O(|\Sigma| \cdot |\mathcal{Q}|)$ time.

¹⁰This upper bound will depend on some parameters such as an input (a letter) and a set of states of the automaton (a block of the partition).

¹¹This does not make sense for sets and is specific for a data structure. What is meant here is a bit informal and designates a position directly accessible, e.g. the head for a stack, the value of the root for a tree, etc.

Proof. The non trivial part is the computation of the inverse transition function $\delta^{-1}(q, a)$, for all $q \in \mathcal{Q}$ and $a \in \Sigma$. This can be done in $O(|\Sigma| \cdot |\mathcal{Q}|)$ time by iterating over Σ and traversing the automaton (e.g. with a DFS) while keeping track of the predecessor at each step. ■

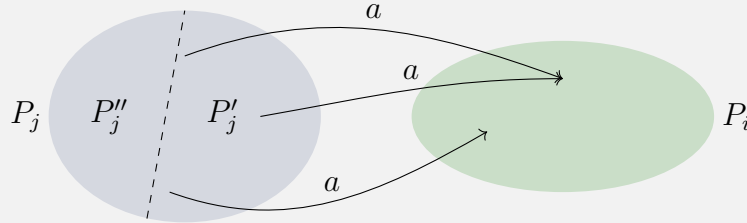
Lemma 2. An iteration of the *while* loop in both algorithms for a splitter $s = (a, C)$ takes a time proportional to the number of transitions terminating in C and the number of symbols in the alphabet, i.e. $\Theta(|\Sigma| \cdot |\delta^{-1}(C, a)|)$ time.

Proof. We start by showing it for Algorithm 1. We pick $a \in \Sigma$ such that $L_a \neq \emptyset$ and an $i \in L_a$. We need to examine all $j < k$ such that $\exists q \in P_j, \delta(q, a) \in s_{a,i}$ to construct the sets corresponding to splitting the block P_j w.r.t. a and P_i .

Let $j < k$. From the definition of $s_{a,i}$, we obtain the following:

$$\begin{aligned} \exists q \in P_j, \delta(q, a) \in s_{a,i} &\iff \exists q \in P_j, \delta(q, a) \in P_i \wedge \underbrace{\delta^{-1}(\delta(q, a), a) \neq \emptyset}_{\text{true}} \\ &\iff \exists q \in P_j, \delta(q, a) \in P_i \end{aligned}$$

Which corresponds to finding states in P_j having an outgoing a -transition to a state in P_i , as represented in the following scheme:



This set of states can be expressed via the inverse transition function:

$$\{q \in P_j \mid \delta(q, a) \in P_i\} = \left(\bigcup_{q' \in P_i} \delta^{-1}(q', a) \right) \cap P_j$$

Since δ^{-1} was already computed in the first step of the algorithm, we can determine using req. 3 whether a state of $\bigcup_{q \in P_i} \delta^{-1}(q, a)$ is also in P_j in $\Theta(1)$ time.

Thus, instead of examining P_j for all $j < k$, we rather go through the table of δ^{-1} and for each state q such that $\delta(q, a) \in P_i$, we know from req. 2 that we can determine the index $j < k$ (because there are k blocks) of the block P_j containing q in $\Theta(1)$ time. The sets P'_j and $P''_j = P_k$ resulting from the partition can be constructed on the fly without any additional time cost. The construction of the sets $s_{b,j}$ and $s_{b,k}$ as well as the update of L_b for all $b \in \Sigma$ can also be done on the fly but require $\Theta(1)$ time for each symbol $b \in \Sigma$ and thus add up to a total of $\Theta(|\Sigma|)$ time.

Overall, one iteration of the loop runs in time

$$\Theta \left(|\Sigma| \cdot \left| \bigcup_{q \in P_i} \delta^{-1}(q, a) \right| \right)$$

We now show the result for Algorithm 2. A splitter $s =: (a, C)$ is picked from \mathcal{W} . Then, we iterate over all blocks $B \in \mathcal{P}$ that may be split with s . Let $B \in \mathcal{P}$.

$$\begin{aligned} s \text{ splits } B &\iff \exists q_1, q_2 \in B \quad \delta(q_1, a) \in C \wedge \delta(q_2, a) \notin C \\ &\iff \{q_1 \in B, \delta(q_1, a) \in C\} \neq \emptyset \wedge \{q_2 \in B, \delta(q_2, a) \notin C\} \neq \emptyset \\ &\iff \left(\bigcup_{q \in C} \delta^{-1}(q, a) \right) \cap B \neq \emptyset \wedge \left(\bigcup_{q \in B} \delta(q, a) \right) \setminus C \neq \emptyset \end{aligned}$$

Likewise, instead of examining all blocks $B \in \mathcal{P}$, we go through δ^{-1} and we update the splitters for all symbols in Σ and one iteration of the loop runs in time

$$\Theta \left(|\Sigma| \cdot \left| \bigcup_{q \in C} \delta^{-1}(q, a) \right| \right)$$

■

For the sake of simplicity, we now denote $\bigcup_{q \in C} \delta^{-1}(q, a)$ by $\overset{\leftarrow a}{C}$.

We will now justify the logarithmic factor in the time complexity. We briefly explain why a logarithm stands out and prove the statement by induction. The idea is that for each symbol $a \in \Sigma$, a state $q \in \mathcal{Q}$ can be in at most one of the splitters in \mathcal{W} . When the loop iterates over this splitter, it will split the block and keep the smaller one, whose size will be at most the size of the splitter divided by two. This means that a splitter s can be processed at most $\log \|s\|$ times. We now properly state and prove the property

by induction.

We see splitters as updatable program variables. This means that $s_{a,i}$ as a set may differ along the loop, however we know there exists some $q \in \mathcal{Q}$ such that $s_{a,i}$ is the unique set containing q throughout the execution. This gives a way to characterize splitters throughout the whole execution.

Lemma 3. Any splitter in the workset $s \in \mathcal{W}$ is processed – i.e. picked at the beginning of the *while* loop – at most $\lfloor \log \|s\| \rfloor$ times.

Proof. We first show the following statement:

During an iteration, the chosen splitter s will either be removed from the set of splitters or its size will be reduced by at least $\frac{\|s\|}{2}$ after the iteration^a.

Let $\{P_1, \dots, P_\ell\}$ be the current partition and let $s = (a, C)$ be the picked splitter. Thus, s is no longer in the set of splitters \mathcal{W} . Since we go over all splittable blocks, C may also be split by s .

- If C is not split by s , then s was already removed from the splitters. Note that it can be added again later if it is split by another splitter.
- If C is split into C' and C'' , then $(a, \min\{C', C''\})$ is added to the splitters and its size is at most $\frac{\|s\|}{2}$.

Therefore, since a splitter cannot be empty, s can be processed at most m times where m is such that

$$1 \leq \frac{\|s\|}{2^m} < 2$$

which is equivalent to

$$m \leq \log \|s\| < m + 1 \quad \text{i.e.} \quad \lfloor \log \|s\| \rfloor = m$$

■

^aThat is a bit informal because, the original splitter is actually removed and another “similar” splitter (cf the paragraph above the statement of the lemma) of size at most $\frac{\|s\|}{2}$ will be added.

Lemma 4. Any block B resulting from a split and not added as a splitter is processed at most $\lfloor \log \frac{|B|}{2} \rfloor$ times.

Proof. Let $P \in \mathcal{P}$ be a block in the partition and $s =: (a, C)$ be a splitter such that s splits P into S and B so that (a, S) is added to the workset and B is not. In order for B to be processed, some (b, B) must be added to the workset.

Since the only way to create a fresh splitter is to split B , this means that there exists some later step in the loop such that some splitter (b, B') splits B into B_1 and B_2 and $(b, \min\{B_1, B_2\})$ is added to the workset.

From lemma 3, we know that this splitter can be processed at most $\lfloor \log |\min\{B_1, B_2\}| \rfloor$ times. Since $|\min\{B_1, B_2\}| \leq \frac{|B|}{2}$ and since there is no splitter containing B in the workset, we obtain that B can be processed at most $\lfloor \log \frac{|B|}{2} \rfloor$ times. ■

Lemma 5. Let us consider some step in the algorithm such that \mathcal{P} is the current partition. We give an estimation of the total time spent in the loop until termination as an upper bound:

$$T := \theta \cdot \left(\sum_{(a,C) \in \mathcal{W}} |\overset{\leftarrow a}{C}| \log |C| + \sum_{(a,B) \in (\Sigma \times \mathcal{P}) \setminus \mathcal{W}} |\overset{\leftarrow a}{B}| \log \frac{|B|}{2} \right)$$

where θ is the constant factor of proportionality that may be obtained from lemma 2.

Proof. We show the result by induction over the steps.

Base case: the current partition is $\{P_1, P_2\}$ and we may assume w.l.o.g. that P_1 is the smaller set, so that $\mathcal{W} = \{(a, P_1), a \in \Sigma\}$. We have to show that the total time spent in the loop is bounded by

$$T = \theta \cdot \sum_{a \in \Sigma} \left(|\overset{\leftarrow a}{P_1}| \log |P_1| + |\overset{\leftarrow a}{P_2}| \log \frac{|P_2|}{2} \right)$$

We know from lemma 2 that an iteration of the loop for (a, P_i) takes $\theta |\overset{\leftarrow a}{P_i}|$ time.

- From lemma 3, for all $a \in \Sigma$, (a, P_1) can be processed at most $\log \|(a, P_1)\| = \log |P_1|$ times^a, hence the total time for (a, P_1) is bounded by $\theta |\overset{\leftarrow a}{P_1}| \log |P_1|$ and thus the total time for any splitter

(a, P_1) is bounded by:

$$\theta |\overset{\leftarrow a}{P}_1| \log |P_1|$$

- From lemma 4, for all $a \in \Sigma$, (a, P_2) can be processed at most $\log \frac{\| (a, P_2) \|}{2} = \log \frac{|P_2|}{2}$ times. Thus, the total time for any non-splitter (a, P_2) is bounded by:

$$\theta |\overset{\leftarrow a}{P}_2| \log \frac{|P_2|}{2}$$

Inductive step: Let $\{P_1, \dots, P_\ell\} := \mathcal{P}$ be the current partition. The induction hypothesis states that the estimate for the total time spent in the loop until termination for a is bounded by

$$T := \theta \left(\sum_{(a, C) \in \mathcal{W}} |\overset{\leftarrow a}{C}| \log |C| + \sum_{(a, B) \in \Sigma \times \mathcal{P} \setminus \mathcal{W}} |\overset{\leftarrow a}{B}| \log \frac{|B|}{2} \right)$$

By going through one more step, some blocks of \mathcal{P} may be split and we define a new estimate \hat{T} over this new partition and we have to show that $\hat{T} \leq T$.

Suppose some $B \in \mathcal{P}$ is split into B' and B'' by any splitter. We have to consider the cases where (a, B) is a splitter or not for all $a \in \Sigma$. Let $a \in \Sigma$.

- $(a, B) \in \mathcal{W}$: \mathcal{W} is updated as follows:

$$\mathcal{W} := \mathcal{W} \setminus \{(a, B)\} \cup \{(a, B'), (a, B'')\}$$

Thus, instead of taking a time $\theta |\overset{\leftarrow a}{B}| \log |B|$ time for (a, B) , it now takes a time bounded by:

$$\theta |\overset{\leftarrow a}{B}'| \log |B'| + \theta |\overset{\leftarrow a}{B}''| \log |B''|$$

Finally, we show the following:

$$|\overset{\leftarrow a}{B}'| \log |B'| + |\overset{\leftarrow a}{B}''| \log |B''| \leq |\overset{\leftarrow a}{B}| \log |B|$$

From $B = B' \cup B''$ and because the automaton is deterministic, we get $|B| = |B'| + |B''|$ and $|\overset{\leftarrow a}{B}| = |\overset{\leftarrow a}{B'}| + |\overset{\leftarrow a}{B''}|$, hence:

$$\begin{aligned}
& |\overset{\leftarrow a}{B'}| \log |B'| + |\overset{\leftarrow a}{B''}| \log |B''| - |\overset{\leftarrow a}{B}| \log |B| \\
&= |\overset{\leftarrow a}{B'}| \log |B'| + (|\overset{\leftarrow a}{B}| - |\overset{\leftarrow a}{B'}|) \log (|B| - |B'|) - |\overset{\leftarrow a}{B}| \log |B| \\
&= |\overset{\leftarrow a}{B'}| \log \frac{|B'|}{|B| - |B'|} + |\overset{\leftarrow a}{B}| \log \frac{|B| - |B'|}{|B|} \\
&\leq |\overset{\leftarrow a}{B'}| \log \frac{|B|}{|B| - |B'|} + |\overset{\leftarrow a}{B}| \log \frac{|B| - |B'|}{|B|} \\
&= (|\overset{\leftarrow a}{B}| - |\overset{\leftarrow a}{B'}|) \underbrace{\log \frac{|B| - |B'|}{|B|}}_{\leq 1} \leq 0
\end{aligned}$$

- $(a, B) \notin \mathcal{W}$: $(a, \min\{B', B''\})$ is added to the set of splitters. We may assume w.l.o.g. that B' is the smaller set. Thus, the corresponding term in the sum $|\overset{\leftarrow a}{B}| \log \frac{|B|}{2}$ is updated as follows:

$$\underbrace{\theta |\overset{\leftarrow a}{B'}| \log |B'|}_{\text{lemma 3}} + \underbrace{\theta |\overset{\leftarrow a}{B''}| \log \left(\frac{|B''|}{2} \right)}_{\text{lemma 4}}$$

Since $B = B' \cup B''$, the sum above can be written as:

$$\theta |\overset{\leftarrow a}{B'}| \log |B'| + \theta (|\overset{\leftarrow a}{B}| - |\overset{\leftarrow a}{B'}|) \log \left(\frac{|B| - |B'|}{2} \right)$$

By using the fact that $|B'| \leq \frac{|B|}{2}$, we have:

$$\begin{aligned}
& |\overset{\leftarrow a}{B'}| \log |B'| + (|\overset{\leftarrow a}{B}| - |\overset{\leftarrow a}{B'}|) \log \left(\frac{|B| - |B'|}{2} \right) - |\overset{\leftarrow a}{B}| \log \frac{|B|}{2} \\
&\leq |\overset{\leftarrow a}{B'}| \log \frac{|B|}{2} + (|\overset{\leftarrow a}{B}| - |\overset{\leftarrow a}{B'}|) \log \left(\frac{|B| - |B'|}{2} \right) - |\overset{\leftarrow a}{B}| \log \frac{|B|}{2} \\
&\leq (|\overset{\leftarrow a}{B}| - |\overset{\leftarrow a}{B'}|) \log \left(\frac{|B| - |B'|}{|B|} \right) \leq 0
\end{aligned}$$

Overall, splitting any block of \mathcal{P} does not increase the total time spent over the loop. By iterating over all blocks, we obtain that the total time \hat{T} spent in the loop from this new iteration until termination is such that $\hat{T} \leq T$. ■

^aWe drop the floor function for simplicity. Note that since we are giving upper bounds, this is completely valid. However, the floor function will have some importance later in the formalization.

Theorem 1. Algorithm 1 and Algorithm 2 run in $O(|\Sigma| \cdot |\mathcal{Q}| \log |\mathcal{Q}|)$ time.

Proof. From lemma 5, we obtain in particular that for the initial partition the time spent in the loop until termination is bounded by:

$$T = \theta \cdot \sum_{a \in \Sigma} \left(|\overset{\leftarrow a}{P}_1| \log |P_1| + |\overset{\leftarrow a}{P}_2| \log \frac{|P_2|}{2} \right)$$

where θ is the constant of proportionality obtained in lemma 2 and $\{P_1, P_2\}$ is the initial partition.

Thus using concavity of the logarithm, the following holds:

$$T \leq \theta \sum_{a \in \Sigma} \underbrace{(|\overset{\leftarrow a}{P}_1| + |\overset{\leftarrow a}{P}_2|)}_{\leq |\mathcal{Q}|} \log \underbrace{(|P_1| + |P_2|)}_{\leq |\mathcal{Q}|} \leq \theta |\Sigma| |\mathcal{Q}| \log |\mathcal{Q}|$$
■

6.3 Towards a formal proof

Although the paper proof looks rather simple and already formalized, the objective of the project is to obtain a formal proof for the time complexity of the algorithm using the Refinement Framework with its extension to runtime analysis. We follow a top-down strategy, from the abstract level to the implementation.

6.3.1 Abstract level

The purpose of this section is to explain how the analysis at the abstract level is structured without adding too many details about Isabelle.

Setting up the framework: First, we define the algorithm on the abstract level using the NREST framework as shown below:

```

definition Hopcroft_abstractT where
Hopcroft_abstractT  $\mathcal{A} \equiv$ 
  if (check_states_empty_spec  $\mathcal{A}$ ) then mop_partition_empty else
  if (check_final_states_empty_spec  $\mathcal{A}$ ) then
    mop_partition_singleton ( $\mathcal{Q} \mathcal{A}$ ) else
  do {
    PL  $\leftarrow$  init_spec  $\mathcal{A}$ ;
    ( $P, \_$ )  $\leftarrow$  monadic_WHILEIET (Hopcroft_abstract_invar  $\mathcal{A}$ )
      ( $\lambda s. \text{estimate\_iteration } \mathcal{A} \ s$ ) check_b_spec
      (Hopcroft_abstract_f  $\mathcal{A}$ ) PL;
    RETURNT P
  }

```

It is parameterized by abstract costs for each operation, like checking emptiness, returning an empty partition or a singleton, initializing the state¹², etc. For the sake of simplicity and since names are rather self-explanatory, we only give the definition of `check_states_empty_spec` as an example and detail the *while* loop. Checking emptiness for the set of states costs one coin of the currency `check_states_empty` and can be defined as follows:

```

definition check_states_empty_spec  $\mathcal{A} \equiv$ 
  consume (RETURNT ( $\mathcal{Q} \mathcal{A} = \{\}$ )) (cost ''check_states_empty'' 1)

```

The purpose of using currencies is to be able to give an abstract description of the costs, and then to instantiate them with concrete values in the implementation which will be expressed in terms of concrete elementary costs like the number of comparisons or assignments. For example, if the set of states is represented as a tree t , there should be a method or an attribute `t.isempty` kept up to date during the execution of the algorithm, so that the cost of checking emptiness is equal to the cost of a function call plus the cost for reading memory at the location of the boolean `t.isempty`, that is constant time.

The *while* loop requires to be provided with an invariant, an estimation of the total amount of resource to be consumed – here time, and this estimation has to decrease over an iteration, the loop condition, the function to be executed in the loop body and the initial state. The loop body defined in `Hopcroft_abstract_f` is given below:

¹²States in this setting are tuples (P, L) where P is a partition of the set of states \mathcal{Q} of \mathcal{A} together with a workset L of splitters.

```

definition Hopcroft_abstract_f where
Hopcroft_abstract_f  $\mathcal{A}$  =
   $\lambda$ PL. do {
    ASSERT (Hopcroft_abstract_invar  $\mathcal{A}$  PL);
    (a,p)  $\leftarrow$  pick_splitter_spec PL;
    PL'  $\leftarrow$  update_split_spec  $\mathcal{A}$  PL a p;
    RETURN PL'
  }

```

The main operation, i.e. splitting blocks of the partition and updating the workset accordingly is specified in `update_split_spec`. As written in the informal paper proof of lemma 2, this operation is a bit tricky since both splitting and updating are performed on-the-fly at the same time, thus they are gathered together for the moment. The definition of `update_split_spec` is given below:

```

definition update_split_spec  $\mathcal{A} \equiv \lambda(P,L) \ a \ p.$ 
  SPEC ( $\lambda(P', L')$ .
    Hopcroft_update_splitters_pred  $\mathcal{A}$  p a P L L'  $\wedge$  P' =
      Hopcroft_split  $\mathcal{A}$  p a {} P
  ) ( $\lambda\_.$  cost ''update_split'' (enat ( $|\Sigma \mathcal{A}| * |\text{preds } \mathcal{A} \ a \ p|$ )))

```

where `preds \mathcal{A} a p` represents the set of all predecessors $\bigcup_{q \in p} \delta^{-1}(q, a) = \overset{\leftarrow a}{p}$ of all states of p labelled by a . The cost specified by this definition stems from the result of lemma 2, namely that the cost for splitting and updating the partition and the workset is $\Theta(|\Sigma| \cdot |\overset{\leftarrow a}{p}|)$.

The most important part of this definition is the estimate of the cost for the loop, i.e. the amount of resource consumed by the loop. From lemma 5, we have the following estimate:

$$T = \theta \cdot \left(\sum_{(a,C) \in \mathcal{L}} |\overset{\leftarrow a}{C}| \lceil \log_2 |C| \rceil + \sum_{(a,B) \in (\Sigma \times \mathcal{P}) \setminus \mathcal{L}} |\overset{\leftarrow a}{B}| \lceil \log_2 \frac{|B|}{2} \rceil \right) \quad (1)$$

Remark 3. In the following, we denote $x \mapsto \lfloor \log_2 x \rfloor$ by $\mapsto \log x$ for simplicity.

The constant θ comes from lemma 2 and represents constant costs inherent to the loop body, like the cost of checking emptiness, calling a function, etc. In Isabelle, this factor θ is defined following the definition of `Hopcroft_abstract_f` and `check_b_spec` as follows:

```

definition cost_1_iteration  $\equiv$ 
  cost ''call'' 1 + cost ''check_l_empty'' 1 + cost ''if'' 1 +
  cost ''pick_splitter'' 1 + cost ''update_split'' 1

```


Then, given a suitable definition `estimate` for the inner expression in T – i.e. the “raw” estimate – the estimate for the loop is defined as the product¹³ of this constant factor and the inner expression, as follows:

```
definition estimate_iteration where
  estimate_iteration A PL  $\equiv$  cost_1_iteration *m estimate A PL
```

Finding a suitable definition for the estimate: Finding a suitable definition for the inner expression in T is not straightforward. The first attempt was to define it as follows:

```
definition estimate1 A  $\equiv$   $\lambda(P,L).$ 
   $\sum\{(\text{card}(\text{preds } A \text{ (fst } s) \text{ (snd } s))) * \text{Discrete.log}(\text{card}(\text{snd } s)) \mid$ 
     $s. s \in L\} +$ 
   $\sum\{(\text{card}(\text{preds } A \text{ (fst } s) \text{ (snd } s))) * \text{Discrete.log}(\text{card}(\text{snd } s)/2) \mid$ 
     $s. s \in \Sigma A \times P - L\}$ 
```

Remark 4. The function `Discrete.log :: nat \Rightarrow nat` is the floor of the logarithm in base 2, i.e. it formalizes the function $x \mapsto \lfloor \log_2 x \rfloor$. Unlike the informal proof of lemma 5, the floor function cannot be dropped – as is – in Isabelle, mostly because the Refinement Framework works with natural numbers for the costs. This makes sense because the logarithm essentially captures the number of times blocks are split, which is a natural number.

The problem with this definition is that the function $s := (a, C) \mapsto \overset{\leftrightarrow a}{|C|} \log |C|$ has no reason to be injective either on L or its complement and thus the set comprehension expressions within the sums may actually contain duplicates that are merged together. One way to fix this is to sum over multisets. Multisets can also be defined via set comprehension in Isabelle, however having a structure allowing for induction is often more convenient. Thus, we prefer lists to multisets.

A very convenient way to link lists and multisets is to use permutations, which are defined in Isabelle using multisets: a list ℓ is a permutation of a list ℓ' if and only if ℓ and ℓ' have the same multiset, i.e. in Isabelle if and only if `mset ℓ = mset ℓ'` . We can lift this definition¹⁴ to permutations between lists and sets with the help of the function `mset_set` in Isabelle. Thus, we can define the estimate as follows:

¹³The operator `*m` lifts multiplication to cost expressions.

¹⁴A list ℓ permutes a set S , which we denote by $\ell <\sim\sim\sim> S$, iff `mset ℓ = mset_set S`.

```

definition estimate  $\mathcal{A} \equiv \lambda(P, L).$ 
  let xs = (SOME xs. xs <~~~> L) and
      ys = (SOME ys. ys <~~~>  $\Sigma \mathcal{A} \times P - L$ ) in
   $\sum_{s \leftarrow xs}. \text{card}(\text{preds } \mathcal{A} \text{ (fst } s) \text{ (snd } s)) * \text{Discrete.log}(\text{card}(\text{snd } s)) +$ 
   $\sum_{s \leftarrow ys}. \text{card}(\text{preds } \mathcal{A} \text{ (fst } s) \text{ (snd } s)) * \text{Discrete.log}(\text{card}(\text{snd } s)/2)$ 

```

Remark 5. This definition makes use of the non-deterministic choice operator `SOME` of Isabelle to choose a permutation of L and $\Sigma \times \mathcal{P} \setminus \mathcal{L}$ respectively. Since we are working with finite automata, such permutations always exist.

Simplifying the estimate: Unfortunately, such a definition makes the proofs very difficult. For a state transition $(\mathcal{P}, \mathcal{L}) \rightarrow (\mathcal{P}', \mathcal{L}')$, we would have to examine all cases for all blocks whether they are splitters in L or whether they are split by some other splitter or whether there are new splitters created from this splitter in L' . This would be very tedious to formalize, but here is a sketch of some ideas.

Splitters of the workset can be uniquely characterized thanks to “ a - q -splitters”. It is defined as follows for a state (P, L) :

Definition 3. Let $a \in \Sigma$ and $q \in \mathcal{Q}$. Let B be the unique block of the partition \mathcal{P} containing q . If $(a, B) \in L$, it is the a - q -splitter. Otherwise, there is no a - q -splitter.

This can be achieved in Isabelle using the option type. Although this is a good characterization of splitters, it is not easy to work with it. For example, if we want to show that the number of splitters decreases, we would have to show that the number of a - q -splitters decreases for any $a \in \Sigma$ and any $q \in \mathcal{Q}$. This is rather tricky to formalize.

In order to simplify the expression of the estimate (1), we can remove the halving factor in the second term, which allows both term to be gathered together under one sum. We obtain an upper bound for the previous expression and the difference between the two is “small”:

$$\begin{aligned}
T &\leq \theta \cdot \left(\sum_{(a, C) \in \mathcal{L}} |\overset{\leftarrow a}{C}| \log |C| + \sum_{(a, B) \in (\Sigma \times \mathcal{P}) \setminus \mathcal{L}} |\overset{\leftarrow a}{B}| \log |B| \right) \\
&= \theta \sum_{(a, C) \in \Sigma \times \mathcal{P}} |\overset{\leftarrow a}{C}| \log |C|
\end{aligned}$$

Thus, we work with the following simplified estimate:

$$\hat{T} := \theta \sum_{(a,C) \in \Sigma \times \mathcal{P}} |\overset{\leftarrow a}{C}| \log |C| \quad (2)$$

Remark 6. Recall that we dropped the floor function in the notations for simplicity. However, the floor function will have some importance later in the formalization. The full expression of the estimate is given by

$$\hat{T} = \theta \sum_{(a,C) \in \Sigma \times \mathcal{P}} |\overset{\leftarrow a}{C}| \lfloor \log_2 |C| \rfloor$$

The error ΔT is given by

$$\Delta T := |\hat{T} - T| = \theta \sum_{(a,B) \in (\Sigma \times \mathcal{P}) \setminus \mathcal{L}} |\overset{\leftarrow a}{B}| \log \frac{|B|}{2} \quad (3)$$

Thus, performing this simplification does not change the asymptotic behavior of the estimate. The definition in Isabelle is also simpler:

```
definition estimate  $\mathcal{A} \equiv \lambda(\mathbf{P}, \mathbf{L}).$ 
  let  $\mathbf{xs} = (\text{SOME } \mathbf{xs}. \mathbf{xs} <\sim\sim\sim> \Sigma \mathcal{A} \times \mathbf{P})$  in
   $\sum \mathbf{s} \leftarrow \mathbf{xs}. \text{card } (\text{preds } \mathcal{A} \text{ (fst } \mathbf{s}) \text{ (snd } \mathbf{s})) * \text{Discrete.log } (\text{card } (\text{snd } \mathbf{s}))$ 
```

Remark 7. The previous definition could be equivalently achieved by using Isabelle's summation over sets:

```
definition estimate  $\mathcal{A} \equiv \lambda(\mathbf{P}, \mathbf{L}).$ 
   $\sum \mathbf{s} \in \Sigma \mathcal{A} \times \mathbf{P}. \text{card } (\text{preds } \mathcal{A} \text{ (fst } \mathbf{s}) \text{ (snd } \mathbf{s})) * \text{Discrete.log } (\text{card } (\text{snd } \mathbf{s}))$ 
```

Unfortunately, I noticed it after the formalization was done. It may shorten some proofs and improve overall understanding, so it is worth switching to this definition.

Proving that the estimate decreases: We can now prove that the estimate decreases. We can consider two approaches:

- (1) either we follow the idea of the paper proof given in theorem 1 and a non-strict inequality is sufficient,
- (2) or we follow the rules given by the Refinement Framework (for a *while* loop given an estimate of the time spent in the loop) and we show a strict inequality.

We start with approach (1) since it is easier to prove. We first sketch out the idea of the proof, followed by some technical details about the proof in Isabelle. Given a state $(\mathcal{P}, \mathcal{L})$ satisfying some invariants and a state transition $(\mathcal{P}, \mathcal{L}) \rightarrow (\mathcal{P}', \mathcal{L}')$, we have to show that

$$\hat{T}(\mathcal{P}', \mathcal{L}') \leq \hat{T}(\mathcal{P}, \mathcal{L})$$

Forgetting the constant θ , we have to show that

$$\sum_{(a,C) \in \Sigma \times \mathcal{P}'} |\overset{\leftarrow a}{C}| \log |C| \leq \sum_{(a,C) \in \Sigma \times \mathcal{P}} |\overset{\leftarrow a}{C}| \log |C|$$

Let $\{P_1, \dots, P_m\} := \mathcal{P}$ be an enumeration of the m blocks of \mathcal{P} such that the first k blocks P_1, \dots, P_k are split and P_{k+1}, \dots, P_m are not. This means that for each $i \in \{1, \dots, k\}$, there exists P'_i and P''_i in \mathcal{P}' resulting from the split of P_i in \mathcal{P} . Thus, \mathcal{P}' can be written as follows:

$$\mathcal{P}' = \{P'_1, P''_1, \dots, P'_k, P''_k, P_{k+1}, \dots, P_m\}$$

Therefore, we can rewrite the previous inequality as follows:

$$\sum_{(a,C) \in \Sigma \times \{P'_1, P''_1, \dots, P'_k, P''_k, P_{k+1}, \dots, P_m\}} |\overset{\leftarrow a}{C}| \log |C| \leq \sum_{(a,C) \in \Sigma \times \{P_1, \dots, P_m\}} |\overset{\leftarrow a}{C}| \log |C|$$

By splitting the left-hand side sum into two sums over $\{P'_1, P''_1, \dots, P'_k, P''_k\}$ and $\{P_{k+1}, \dots, P_m\}$ and subtracting the terms over $\{P_{k+1}, \dots, P_m\}$ on both sides, we can reduce the problem to showing that

$$\sum_{(a,C) \in \Sigma \times \{P'_1, P''_1, \dots, P'_k, P''_k\}} |\overset{\leftarrow a}{C}| \log |C| \leq \sum_{(a,C) \in \Sigma \times \{P_1, \dots, P_k\}} |\overset{\leftarrow a}{C}| \log |C|$$

that is,

$$\sum_{i=1}^k \left(|\overset{\leftarrow a}{P'_i}| \log |P'_i| + |\overset{\leftarrow a}{P''_i}| \log |P''_i| \right) \leq \sum_{i=1}^k |\overset{\leftarrow a}{P_i}| \log |P_i|$$

Then, we consider the following facts¹⁵:

- A property induced by the concavity of the logarithm:

$$\forall a, b, x, y \in \mathbb{R}_+^* \quad (a + b) \log(x + y) > a \log x + b \log y$$

¹⁵The proofs of those facts are omitted here since they are simple and not very interesting.

- A property on the set of predecessors and the definition of a split:

$$(a, B) \text{ splits } C \text{ into } (C', C'') \implies \overset{\leftarrow a}{C} = \overset{\leftarrow a}{C'} \cup \overset{\leftarrow a}{C''} \wedge \overset{\leftarrow a}{C'} \cap \overset{\leftarrow a}{C''} = \emptyset$$

from which we deduce that

$$(a, B) \text{ splits } C \text{ into } (C', C'') \implies |\overset{\leftarrow a}{C}| = |\overset{\leftarrow a}{C'}| + |\overset{\leftarrow a}{C''}|$$

- A property on the definition of a split:

$$(a, B) \text{ splits } C \text{ into } (C', C'') \implies |C| = |C'| + |C''|$$

Given those properties, we show that

$$\forall i \in \{1, \dots, k\} \quad |\overset{\leftarrow a}{P'_i}| \log |P'_i| + |\overset{\leftarrow a}{P''_i}| \log |P''_i| \leq |\overset{\leftarrow a}{P_i}| \log |P_i|$$

which is sufficient to conclude the proof.

We now give some technical details about the proof in Isabelle. The goal is to show the following lemma:

```
lemma estimate_decrease: estimate  $\mathcal{A}$  (P', L') ≤ estimate  $\mathcal{A}$  (P, L)
```

The definition of the estimate using permutations allows us to do the disjunction between split and non-split blocks in the partition. First, we have to obtain a list \mathbf{xs} (resp. \mathbf{xs}') permuting $\Sigma \times \mathcal{P}$ (resp. $\Sigma \times \mathcal{P}'$) such that

```
estimate  $\mathcal{A}$  (P, L) =  $\sum \mathbf{s} \leftarrow \mathbf{xs}. \text{card}(\text{preds } \mathcal{A} (\text{fst } \mathbf{s}) (\text{snd } \mathbf{s})) * \log(\text{card}(\text{snd } \mathbf{s}))$ 
```

resp.

```
estimate  $\mathcal{A}$  (P', L') =  $\sum \mathbf{s} \leftarrow \mathbf{xs}'. \text{card}(\text{preds } \mathcal{A} (\text{fst } \mathbf{s}) (\text{snd } \mathbf{s})) * \log(\text{card}(\text{snd } \mathbf{s}))$ 
```

Then, we obtain two lists \mathbf{ys} and \mathbf{zs} (resp. \mathbf{ys}' and \mathbf{zs}') such that

```
 $\mathbf{xs} = \mathbf{ys} @ \mathbf{zs} \wedge$   
 $\forall \mathbf{e} \in \text{set } \mathbf{ys}. \text{snd } \mathbf{e} \in \mathcal{P} \cap \mathcal{P}' \wedge$   
 $\forall \mathbf{e} \in \text{set } \mathbf{zs}. \text{snd } \mathbf{e} \notin \mathcal{P} \cap \mathcal{P}'$ 
```

resp.

```
 $\mathbf{xs}' = \mathbf{ys}' @ \mathbf{zs}' \wedge$   
 $\forall \mathbf{e} \in \text{set } \mathbf{ys}'. \text{snd } \mathbf{e} \in \mathcal{P} \cap \mathcal{P}' \wedge$   
 $\forall \mathbf{e} \in \text{set } \mathbf{zs}'. \text{snd } \mathbf{e} \notin \mathcal{P} \cap \mathcal{P}'$ 
```

Just like above, by showing that $\mathbf{ys} = \mathbf{ys}'$ we reduce the current goal to showing that

$$\sum_{\mathbf{s} \leftarrow \mathbf{zs}} \text{card}(\text{preds } \mathcal{A} \text{ (fst } \mathbf{s}) \text{ (snd } \mathbf{s})) * \log(\text{card}(\text{snd } \mathbf{s})) \leq \sum_{\mathbf{s} \leftarrow \mathbf{zs}'} \text{card}(\text{preds } \mathcal{A} \text{ (fst } \mathbf{s}) \text{ (snd } \mathbf{s})) * \log(\text{card}(\text{snd } \mathbf{s}))$$

Then, we show two key properties which will be useful to conclude the proof. Since picking any $(A, B) \in \mathcal{P}' \times \mathcal{P}'$ does not imply that A and B result from the same split, we cannot directly obtain a single property with an equivalence capturing the whole idea. Instead, we have to prove two separate implications that are dual properties.

The idea is that given a block of the current partition, we can find the two blocks of the new partition resulting from the split of this block. Conversely, given a block of the new partition, we can find the other block of the new partition and the block from the current partition such that the two blocks from the new partition result from the split of the block from the current partition. The latter property is the one making a formulation with an equivalence quite tedious since we cannot directly pick such two blocks together from the new partition.

Those two properties are named `unique_split` and `unique_split_conv` and are defined in Isabelle as follows:

```
note unique_split=
B ∈ set zs ⇒ ∃!(B1, B2). B1 ∈ set zs' ∧ B2 ∈ set zs' ∧
  (snd B, snd B1, snd B2) ∈ HopcroftSplitted A C a {} P ∧
  fst B = fst B1 ∧ fst B = fst B2
```

and

```
note unique_split_conv=
B1 ∈ set zs' ⇒ ∃!(B, B2). B1 ∈ set zs ∧ B2 ∈ set zs' ∧
  split_pred A P (snd B) a C (snd B1) (snd B2) ∧
  fst B = fst B1 ∧ fst B = fst B2
```

Let us dissect those two properties and translate them into words in order to understand their meaning.

unique_split: Let $\sigma \in \Sigma$ be a label and let $B \in \mathcal{P} \setminus \mathcal{P}'$ be a block of the current partition \mathcal{P} that is about to be split, i.e. that is not in \mathcal{P}' – which is represented in Isabelle by (σ, B) being a member of the list \mathbf{zs} .

We can find two unique blocks B_1 and B_2 of the new partition \mathcal{P}' – actually, in $\mathcal{P}' \setminus \mathcal{P}$ which is represented in Isabelle by (α, B_1) and (α, B_2) being members of the list \mathbf{zs}' for some label $\alpha \in \Sigma$ – such that B_1 and

B_2 are the result of the split of B by some splitter¹⁶ – and B_1 and B_2 are both labelled by the same letter σ if they are splitters in the new workset, i.e. if $(\sigma, B) \in \mathcal{L}$ then $(\sigma, B_1) \in \mathcal{L}'$ and $(\sigma, B_2) \in \mathcal{L}'$.

This property allows us to uniquely characterize and obtain blocks resulting from a split.

unique_split_conv: Let $\sigma \in \Sigma$ be a label and let $B_1 \in \mathcal{P}' \setminus \mathcal{P}$ be a block of the new partition \mathcal{P}' that is not in \mathcal{P} – which is represented in Isabelle by (σ, B_1) being a member of the list \mathbf{zs}' .

We can find two unique blocks B of the current partition \mathcal{P} and B_2 of the new partition \mathcal{P}' – actually, in $\mathcal{P} \setminus \mathcal{P}'$ (resp. $\mathcal{P}' \setminus \mathcal{P}$) which is represented in Isabelle by (α, B) (resp. (α, B_2)) being a member of the list \mathbf{zs} (resp. \mathbf{zs}') for some label $\alpha \in \Sigma$ – such that B_1 and B_2 result from the split of B by some splitter and B, B_1 and B_2 are all labelled by the same letter σ if they are splitters in the current and new worksets, i.e. if $(\sigma, B_1) \in \mathcal{L}'$ then $(\sigma, B) \in \mathcal{L}$ and $(\sigma, B_2) \in \mathcal{L}'$.

Then, we define¹⁷ a function f which maps a block of the current partition to the two blocks of the new partition resulting from the split of this block. This function is defined as follows:

```
define f = λB. (THE (B1, B2).
  B1 ∈ set zs' ∧ B2 ∈ set zs' ∧
  (snd B, snd B1, snd B2) ∈ Hopcroft_splitted A C a {} P ∧
  fst B = fst B1 ∧ fst B = fst B2)
```

The operator **THE** in Isabelle allows us to select “the” element satisfying a predicate. Thanks to the property **unique_split** giving unicity for this predicate, the function f is well-defined.

Then, we prove that the image of \mathbf{zs} by f is exactly (a permutation of) \mathbf{zs}' . Formally, we prove that

```
⋃ { {fst (f B), snd (f B)} | B. B ∈ set zs } = set zs'
```

and

```
fold ((@) ∘ (λx. [fst (f x), snd (f x)])) zs [] <~~> zs'
```

We can then replace the expression on the blocks¹⁸ in the sum corresponding to the estimate for (P', L') over \mathbf{zs}' by their preimages in \mathbf{zs} by

¹⁶Any splitter, in particular it can be with the label σ or with the block B itself.

¹⁷This definition is local to this proof, thus f depends on all formerly obtained variables appearing in its definition.

¹⁸Actually, the tuples of letters and blocks.

f and we obtain the sum for the estimate for (P, L) over \mathbf{zs} . Formally, we prove that

$$\begin{aligned} \sum \mathbf{s} \leftarrow \mathbf{zs}'. \text{card}(\text{preds } \mathcal{A}(\text{fst } \mathbf{s})(\text{snd } \mathbf{s})) * \log(\text{card}(\text{snd } \mathbf{s})) = \\ \sum \mathbf{s} \leftarrow \mathbf{zs}. \text{card}(\text{preds } \mathcal{A}(\text{fst}(\text{fst}(\mathbf{f } \mathbf{s})))(\text{snd}(\text{fst}(\mathbf{f } \mathbf{s})))) * \\ \log(\text{card}(\text{snd}(\text{fst}(\mathbf{f } \mathbf{s})))) + \\ \text{card}(\text{preds } \mathcal{A}(\text{fst}(\text{snd}(\mathbf{f } \mathbf{s})))(\text{snd}(\text{snd}(\mathbf{f } \mathbf{s})))) * \\ \log(\text{card}(\text{snd}(\text{snd}(\mathbf{f } \mathbf{s})))) \end{aligned}$$

We then use the property `unique_split_conv` to show that the sum on the right-hand side is bounded by the estimate for (P, L) over \mathbf{zs} , which concludes the proof.

We now consider approach (2) which requires a strict inequality. This approach is still a work in progress and is not yet finished. We would need to find an estimate Γ such that for any state transition $(\mathcal{P}, \mathcal{L}) \rightarrow (\mathcal{P}', \mathcal{L}')$, we have

$$\Gamma(\mathcal{P}', \mathcal{L}') < \Gamma(\mathcal{P}, \mathcal{L})$$

My guess is that the estimate T defined in 1 would satisfy the strict inequality, however I am not sure that the estimate \hat{T} defined in 2 would satisfy it. The reason for this intuition is that T is defined on the workset \mathcal{L} whereas \hat{T} is defined on the partition \mathcal{P} . Indeed, we can prove that for any state transition $(\mathcal{P}, \mathcal{L}) \rightarrow (\mathcal{P}', \mathcal{L}')$ of the algorithm, we have $\mathcal{L} \neq \mathcal{L}'$ but it may happen that $\mathcal{P} = \mathcal{P}'$.

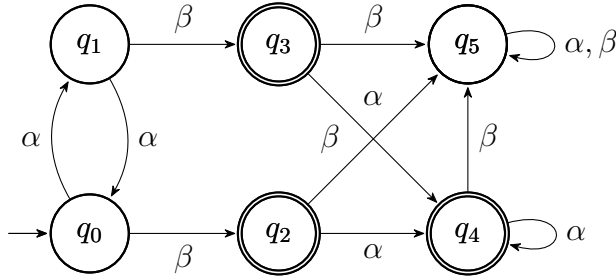


Figure 1: Note that q_3 has a transition to q_4 with an α and q_2 has a transition to q_5 with a β . There is no ambiguity since it is a DFA.

Remark 8. Consider the automaton given in Fig. 1. Note that the partition is stable after the second iteration of the algorithm. The workset however changes after each iteration.

Thus, it may happen that the estimate \hat{T} does not decrease after a state transition where the partition does not change. With the estimate T , we are

Splitter	Partition \mathcal{P}	Workset \mathcal{L}
–	$\{q_0, q_1, q_5\} \{q_2, q_3, q_4\}$	$(\alpha, \{q_0, q_1, q_5\}) (\beta, \{q_0, q_1, q_5\})$
$(\beta, \{q_0, q_1, q_5\})$	$\{q_0, q_1\} \{q_5\} \{q_2, q_3, q_4\}$	$(\alpha, \{q_0, q_1\}) (\alpha, \{q_5\})$
$(\alpha, \{q_0, q_1\})$	$\{q_0, q_1\} \{q_5\} \{q_2, q_3, q_4\}$	$(\alpha, \{q_5\})$
$(\alpha, \{q_5\})$	$\{q_0, q_1\} \{q_5\} \{q_2, q_3, q_4\}$	\emptyset

sure that at each step, a splitter is used and thus a term in the sum over \mathcal{L} is removed and “goes” into the sum over $\Sigma \times \mathcal{P} \setminus \mathcal{L}$ ¹⁹.

However, because of the floor operation, we still cannot guarantee that the estimate T strictly decreases²⁰. My thoughts on this are that “most of the time” there should exist at least one block in the partition that gets split in such a way that the corresponding term in the updated sum is strictly less than the term in the sum for the previous partition. The expression “most of the time” denotes a “convenient split”, i.e. a split that actually cuts off a block in half or almost. In this case indeed, the term $\log n$ gets replaced by something close to $\log \frac{n}{2} = \log n - 1$. However, in cases for example when a splitter splits off only one element from a block, the term $\log n$ gets replaced by $\log(n - 1)$ which is not necessarily strictly less than $\log n$ (recall that \log denotes $\lfloor \log_2 \rfloor$). Cases where such a split occurs for all blocks should be studied, if they exist. If no such cases exist, we would be provided with a good start of a proof that the estimate T strictly decreases. Otherwise, we would have to examine classes of automaton individually and check that the estimate T actually strictly decreases for each of them as well.

7 Further work

As mentioned above, finding a proof for a strict inequality is still a work in progress. It involves finding a suitable estimate, although the estimate T seems to be a good candidate. However, this estimate makes the formalization more complex since it involves to consider separately elements from the workset and its complement and how they are affected by a state transition.

Everything linked to the estimate left aside, the formalization is still at a very early stage since we only focussed on the abstract level. However, this was the most important part of the formalization since it is the most complex one. The next steps would be to dive into the successive refinements and refine the abstract currencies. The formalization of the concrete currencies

¹⁹This remark is very informal.

²⁰Consider $\lfloor \log_2 n \rfloor$ and $\lfloor \log_2 n + 1 \rfloor$: the bigger n is, the more likely it is that both values are equal. Such terms might show up in border cases such as when a splitter splits off only one element from a block.

will be more straightforward since it will be mostly about implementing the abstract currencies with the chosen data structures, which probably already exists.

8 Conclusion

Although I was already quite used to working with Isabelle, this project was a great opportunity to improve my skills with the tool and in particular not to develop bad habits, since I had the best source of advice and help I could have hoped for: Peter Lammich, Tobias Nipkow and the whole team at his chair at TUM. I am very grateful for their help and I thank them heartfully for their time, patience and welcome.

Not only have I learned a lot about abstract concepts of formal verification, but I have also learned a lot about the implementation of Isabelle and its guts. I am now much more confident and my proofs – be it for the *apply*-style or the *Isar*-style – are much more robust and elegant. Even though intuition is not always a good feeling to follow in formal verification, mine has been strengthened as much as my creativity in terms of problem solving.

This project has given me a new taste of academic research and it has strengthened my will to pursue in this direction. Having been able to be part of such an internationally renowned team for a few months has been a great experience and I am very grateful for it.

References

- [EB23] Javier Esparza and Michael Blondin. *Automata Theory: An Algorithmic Approach*. 2023.
- [HL19] Maximilian P. L. Haslbeck and Peter Lammich. Refinement with Time - Refining the Run-Time of Algorithms in Isabelle/HOL. In John Harrison, John O’Leary, and Andrew Tolmach, editors, *10th International Conference on Interactive Theorem Proving (ITP 2019)*, volume 141 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 20:1–20:18, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [HL22] Maximilian P.L. Haslbeck and Peter Lammich. For a few dollars more: Verified fine-grained algorithm analysis down to llvm. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 44(3):14:1–14:36, September 2022.
- [Hop71] John E. Hopcroft. *An $n \log n$ Algorithm for Minimizing States in a Finite Automaton*. Stanford University, Stanford, CA, USA, 1971.
- [Lam16] Peter Lammich. The imperative refinement framework. *Archive of Formal Proofs*, August 2016. https://isa-afp.org/entries/Refine_Imperative_HOL.html, Formal proof development.
- [LT12] Peter Lammich and Thomas Tuerk. Applying data refinement for monadic programs to hopcroft’s algorithm. pages 166–182, 08 2012.