

Julius-Maximilians-Universität Würzburg  
Institut für Informatik  
Lehrstuhl für Informatik IV  
Theoretische Informatik

## **Bachelor Thesis**

# **Simulation of Proof Systems**

Nils Wisiol

submitted on May 11, 2012

supervisor:  
Dr. Christian Glaßer

Acht Jahre nachdem die P-NP-Frage von Cook gestellt wurde [Coo71], war es auch Cook zusammen mit Reckhow, der den engen Zusammenhang dieser Frage mit Beweissystemen feststellte [CR79].

Die Bedeutung der bis heute ungeklärten P-NP-Frage ist immens. Die Gleichheit dieser Mengen würde bedeuten, dass man zu jedem Problem, dessen Lösung sich leicht überprüfen lässt, auch leicht eine Lösung finden kann. Es wäre beispielsweise genau so schwer, einen Beweis zu verifizieren, wie einen Beweis zu finden. Es würde kaum einen Unterschied machen, ob man die PIN zu einer EC-Karte herausfinden möchte, oder überprüfen möchte, ob die eingegebene Nummer die korrekte ist. Aufgrund dieser unwirklich erscheinenden Implikationen glauben die meisten Komplexitätstheoretiker, dass  $P \neq NP$ .

Der Zusammenhang der P-NP-Frage mit Beweissystemen lässt sich wie folgt formulieren. Existiert kein polynomiell beschränktes Beweissystem für TAUT, dann folgt  $P \neq NP$ . Diesen Satz werden wir im Laufe dieser Arbeit beweisen, und anschließend noch etwas tiefer in die Theorie der Beweissysteme einsteigen.

Zunächst wird der Leser in die wichtigsten Definitionen und Begriffe aus dem Gebiet der Beweissysteme eingeführt. Anschließend werden in Kapitel 3 einige wichtige Ergebnisse zusammengefasst. In Kapitel 4 wird bewiesen, dass in bestimmten superpolynomiellen Komplexitätsklassen Sprachen existieren, die keine optimalen Beweissysteme besitzen. Abschließend werden die Ergebnisse der Arbeit zusammengefasst und ein Ausblick auf weitere interessante und offene Fragen gegeben.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Preliminaries</b>	<b>5</b>
<b>3</b>	<b>A brief Overview of Proof Systems</b>	<b>7</b>
<b>4</b>	<b>A set in <math>\text{co-NEXP} \setminus \text{OPT}</math></b>	<b>10</b>
<b>5</b>	<b>Conclusion and future work</b>	<b>13</b>
	<b>Bibliography</b>	<b>14</b>

# 1 Introduction

After Cook stated 1971 the P versus NP question [Coo71], he gave the main motivation to study proof systems in 1979. Cook and Reckhow showed in their article the close relation between the separation of complexity classes and the existence of polynomially bounded proof systems [CR79].

Despite its importance, the P versus NP question remains still open. Informally speaking,  $P = NP$  means that for every problem that has a efficiently verifiable solution, we can find that solution efficiently as well. While most theoreticians assume that  $P \neq NP$ , their equality would have heavy implications. One consequence affects the security of communication in the Internet, as public-key cryptography depends on the existence of certain problems in NP that are not efficiently to decide. If there are no problems with solutions fast to verify but hard to find, as  $P = NP$  would imply, the small lock beside the URL in one's browser could not indicate a secure communication anymore [For09].

$P = NP$  would also have fundamental implications on mathematics: As mathematical proofs must by definition be efficiently verifiable,  $P = NP$  would imply that they are efficiently to find [CR79]. This argument to believe that  $P \neq NP$  is further illustrated in the following quotation, taken from Aaronsons Blog<sup>1</sup>.

“If  $P = NP$ , then the world would be a profoundly different place than we usually assume it to be. There would be no special value in ‘creative leaps’, no fundamental gap between solving a problem and recognizing the solution once it’s found. Everyone who could appreciate a symphony would be Mozart; everyone who could follow a step-by-step argument would be Gauss; everyone who could recognize a good investment strategy would be Warren Buffett.”

Closely related to the P versus NP problem is the question if  $NP = co-NP$ . If  $P = NP$ , then  $NP = co-NP$ , since P is closed under complement. In return, if one can separate NP from co-NP, then  $P \neq NP$ . To connect the field of proof systems with the P versus NP questions, we state

**Proposition 1** ([KMT03], [CR79]).  *$NP = co-NP$  if and only if a polynomially bounded proof system for TAUT exists.*

We will this proposition later in chapter 3.

In order to introduce the reader into the field of proof systems, we will first define the important notions used related to it. Subsequently, we will give an overview of important results about proof systems in chapter 3. These results will connect proof systems with the P versus NP question as mentioned above and will give a basis for the proofs in the subsequent chapter. In chapter 4 we will proof the main theorem of this thesis, showing that there are languages without optimal proof systems in all super-polynomial complexity classes. This chapter will also analyze the inner structure of these languages that do not possess an optimal proof system.

Finally, we will give a conclusion and look forward to currently unresolved problems and further questions.

---

<sup>1</sup><http://www.scottaaronson.com/blog/>

## 2 Preliminaries

As mentioned before, we will first introduce important symbols and definitions. Although some familiarity with standard notions of complexity theory is assumed, we will here define most of the notions used in this thesis. For the most important ones, we will give a short discussion.

Let  $\Sigma = \{0, 1\}$  denote the alphabet. The output of a Turing transducer  $M$  on input  $x \in \Sigma^*$  is denoted by  $M(x)$ . If the transducer  $M$  does not accept or runs forever on input  $x$ , we define  $M(x) = \perp$ . We say a Turing transducer *calculates* a partial function  $f$ , if  $M(x) = f(x)$  for all  $x \in \Sigma^*$ . We further define  $\text{time}_M(x)$  as the number of steps the transducer  $M$  runs on input  $x \in \Sigma^*$ . With  $\mathcal{FP}$  we denote the set of all partial functions  $f$  for which a Turing transducer  $M$  calculating  $f$  exists such that  $\text{time}_M(x) \leq p(|x|)$  for a polynomial  $p$ .

**Definition 1.** A function  $f \in \mathcal{FP}$  is called *proof system* for a language  $L$  if the range of  $f$  is  $L$ . A string  $w$  with  $h(w) = x$  is called an *h-proof* for  $x$ .  $f$  is called a *polynomially bounded proof system*, if there is a polynomial  $p$  such that for every  $x \in L$ , there is a *h-proof*  $w$  with  $|w| \leq p(|x|)$ .

With this definition, a proof system for  $L$  is basically a polynomial-time bounded function that enumerates  $L$ . Notice, although its time bound against the input, the shortest proof of a string  $w \in L$  can be very long. To give an example, let  $h$  be defined by

$$\text{sat}(x) = \begin{cases} \varphi & (x = \langle a, \varphi \rangle \text{ and } \alpha \text{ is an satisfying assignment for } \varphi), \\ \perp & (\text{otherwise}). \end{cases}$$

Then  $h$  is a proof system for SAT. It is an open question, whether  $\text{sat}$  is p-optimal. Köbler and Messner showed, that this question is equivalent to a variety of well studied complexity theoretic assumptions [KM00]. We will show some of the connections in the next chapter.

There may be various proof systems for a language  $L$ . In order to make them comparable, we define the notion of *simulation* of proof systems. It turns out that the notion of simulation corresponds in a certain way with the notion of many-one reducibility [KM00]. In the following definitions, we will keep this correspondence.

**Definition 2.** Let  $h$  and  $h'$  be proof systems for a language  $L$ . If there is a polynomial  $p$  and a function  $f$  such that for all  $w \in \Sigma^*$

$$h(f(w)) = h'(w)$$

and  $|f(w)| \leq p(|w|)$ , then  $h$  *simulates*  $h'$ .

Informally speaking,  $f$  translates  $h$ -proofs into  $h'$ -proofs and keeps them short. In the given definition,  $f$  could be hard or even impossible to calculate. Hence we define a stronger version of this notion, demanding  $f \in \mathcal{FP}$ .

**Definition 3.** Again, let  $h$  and  $h'$  be proof systems for a language  $L$ . If  $h$  *simulates*  $h'$  with a function  $f$  and additionally  $f \in \mathcal{FP}$ ,  $h$  *p-simulates*  $h'$ .

Notice, if  $f$  is a function that can be calculated in polynomial time  $p$ , then we obtain  $|f(w)| \leq p(|w|)$  as required in the definition of simulation. With a proof system p-simulating another, we can

translate proofs as above in polynomial short time. With the notion of simulation of proof systems, we can compare different proof systems for a language  $L$ . With respect to these notions, we will define a notion of the best proof system as follows.

**Definition 4.** *A proof system  $h$  for  $L$  is called optimal, if it simulates every proof system for  $L$ . It is called  $p$ -optimal, if it  $p$ -simulates every proof system for  $L$ .*

Like noted above, this definition corresponds with the definition of complete problems in respect of many-one-reducibility [KMT03]. We will investigate further on the connection between complete problems and optimal proof systems later in lemma 2.

The existence of optimal proof systems for a arbitrary language  $L$  is an important complexity theoretic question. For languages in  $P$  and  $NP$ , there is always a optimal proof system, as we will see in 3. For super-polynomial time complexity classes, there are languages without an optimal proof system, as we will show in chapter 4. For that reason, we will define a complexity class containing all languages possessing an optimal proof system.

**Definition 5.** *Let  $OPT$  be the complexity class of all languages that have a optimal proof system.*

Observe that for  $OPT$  we use the weaker notion of simulation. As noted above, we can easily state a proof system for languages in  $P$  or  $NP$ , therefore we obtain  $NP \subseteq OPT$ . It is an open questions whether  $OPT \subseteq NP$ .

With these notions, we will take a look at important results in the field of optimal proof systems in the next chapter. For notions not defined in this thesis, refer to a standard work of computational complexity like the one from Papadimitriou [Pap94].

is there a simulation  
notion?  
citation, con-  
sequences o  
 $OPT \subseteq NP$ , more  
info

### 3 A brief Overview of Proof Systems

After defining the important notions for this thesis, we will give a brief overview of some important results in the field of optimal proof systems.

One basic lemma that is widely used formalizes a part of the connection between optimal proof systems and polynomial many-one-reducibility. Later in this thesis, we will use it to proof corollary 9. The following proof is mainly taken from Köbler et al. We will omit the proof for optimal proof system, as it easily follows from the proof for p-optimal ones.

**Lemma 2** ([KMT03]). *If  $A$  has a (p-)optimal proof system and if  $B \leq_m^p A$ , then  $B$  has a (p-)optimal proof system, too.*

*Proof.* Let  $h$  be a p-optimal proof system for  $A$  and let  $B$  many-one reduce to  $A$  via  $f \in \mathcal{FP}$ , that is  $x \in B \Leftrightarrow f(x) \in A$ . Then  $h'$  defined by

$$h'(\langle x, w \rangle) = \begin{cases} x & (h(w) = f(x)), \\ \perp & (\text{otherwise}), \end{cases}$$

is a proof system for  $B$ , as  $h(w) = f(x) \in A$  is equivalent to  $x \in B$ . To show  $h'$  is optimal, let  $g'$  be a proof system for  $B$ . In order to obtain a proof system for  $A$ , let  $g$  be

$$g(bw) = \begin{cases} h(w) & (b = 0), \\ f(g'(w)) & (b = 1). \end{cases}$$

Since both  $h(w)$  and  $f(g'(w))$  are in  $A$  and  $h$  is a proof system for  $A$ ,  $g$  is also a proof system for  $A$ . As  $h$  is p-optimal, there is a function  $t \in \mathcal{FP}$  translating  $g$ -proofs to  $h$ -proofs implying that

$$h(t(1w)) = g(1w) = f(g'(w)).$$

This implies  $h'(\langle g'(w), t(1w) \rangle) = g'(w)$ . Hence,  $h'$  p-simulates  $g'$ . □

In contraposition to this, we can state that for  $B \leq_m^p A$ , if  $B$  has no (p-)optimal proof system, then  $A$  has not either.

The following lemma gives a partial answer to the basic question what languages do have optimal proof systems. Although Köbler et al. omitted the proof of the following theorem due to its straightforwardness, we will give a proof for the purpose of this thesis.

**Lemma 3** ([KMT03]). *(i) Every language in  $P$  has a p-optimal proof system.*

*(ii) Every language in  $NP$  has an optimal proof system.*

*Proof.* (i) Let  $L \in P$ . Then there is a function  $f \in \mathcal{FP}$  with  $f(w) = 1 \Leftrightarrow w \in L$ . To show there is a proof system, let  $h$  be defined by

$$h(w) = \begin{cases} w & (f(w) = 1), \\ \perp & (\text{otherwise}). \end{cases}$$

Then  $h$  is a proof system for  $L$ . To show  $h$  is optimal, let  $h'$  be an arbitrary proof system for  $L$ . Then  $h' \in \mathcal{FP}$  by definition and we can translate  $h'$ -proofs with  $h'$  in polynomial time into  $h$ -proofs, in formulas

$$h(h'(w)) = h'(w).$$

Therefore,  $h$  p-simulates every proof system  $h'$ .

- (ii) Let  $L \in \text{NP}$ . Then there is a nondeterministic Turing transducer  $M$  deciding  $L$  in polynomial time. Let  $f_i(x) \in \mathcal{FP}$  be the function calculating the  $i$ -th path of the nondeterministic calculation of  $M$ . Finally, let  $h$  be defined by

$$h(\langle i, w \rangle) = \begin{cases} f_i(w) & (f_i(w) \text{ accepts}), \\ \perp & (\text{otherwise}). \end{cases}$$

Then  $h \in \mathcal{FP}$  is a proof system for  $L$ . To show  $h$  is optimal, let  $h'$  be an arbitrary proof system for  $L$ . Let  $g$  be a function that maps an  $w \in L$  to an  $i$  such that  $f_i(w)$  accepts in polynomial time. Notice that  $g$  may be not in  $\mathcal{FP}$ . With these definitions, we obtain

$$h(\langle g(h'(w)), h'(w) \rangle) = h'(w).$$

Therefore,  $h$  simulates every proof system  $h'$  via the translating function

$$w \mapsto \langle g(h'(w)), h'(w) \rangle.$$

□

This implies  $\text{NP} \subseteq \text{OPT}$ . By stating different properties for P and NP, the lemma connects to the P-NP-question. If one would find a set in NP without an p-optimal proof system, one would have separated P from NP.

**Corollary 4.** *If there is no p-optimal proof system for SAT, then  $P \neq \text{NP}$ .*

In order to prove proposition 1, we will show two lemmata of which the first is taken from the work of Cook and Reckhow [CR79]. It gives an equivalent formulation of the question whether  $\text{NP} = \text{co-NP}$ .

**Lemma 5.**  *$\text{NP} = \text{co-NP}$  if and only if  $\text{TAUT} \in \text{NP}$ .*

*Proof.* Without further assumptions, we can show by using a nondeterministic Turing transducer, that an arbitrary formula is not a tautology in polynomial time by guessing and verifying an assignment for which the formula is falsified. Thus,  $\overline{\text{TAUT}} \in \text{NP}$ .

For every  $L \in \text{NP}$ , there is a function  $f \in \mathcal{FP}$  such that  $x \in L \Leftrightarrow f(x) \in \overline{\text{TAUT}}$ . We will show this as follows. As SAT is  $\leq_m^p$ -complete for NP, we can reduce every  $L \in \text{NP}$  and obtain an  $f' \in \mathcal{FP}$  with  $x \in L \Leftrightarrow f'(x) \in \text{SAT}$ . Hence, there is a truth assignment for  $f'(x)$  such that  $f'(x)$  yields true. By negating, we obtain the existence of a truth assignment such that  $\neg f'(x)$  is false. Consequently,  $\neg f'(x) \in \overline{\text{TAUT}}$ . Notice that the negation of  $f'(x)$  can be done efficiently. By choosing  $f(x) = \neg f'(x)$ , we obtain  $f \in \mathcal{FP}$  with the stated properties.

Assume  $\text{TAUT} \in \text{NP}$ , and let  $L$  be an arbitrary language with  $L \in \text{NP}$ . As shown above, there is a function  $f \in \mathcal{FP}$  such that  $x \in L \Leftrightarrow f(x) \in \overline{\text{TAUT}}$  respectively  $x \in \overline{L} \Leftrightarrow f(x) \in \text{TAUT}$ . Since  $\text{TAUT} \in \text{NP}$ , for any given  $x \in \Sigma^*$  we can in nondeterministic polynomial time decide whether  $f(x)$  is in TAUT. Thus we also can decide in nondeterministic polynomial time whether  $x \in \overline{L}$ , as  $f \in \mathcal{FP}$ . It follows that  $\overline{L} \in \text{NP}$ . As this proves that NP is closed under complement, we obtain



$\text{NP} \subseteq \text{co-NP}$ . To see that  $\text{co-NP} \subseteq \text{NP}$ , let an arbitrary language  $\bar{L}$  be in  $\text{co-NP}$ . By definition we obtain  $L \in \text{NP}$ . As  $\text{NP}$  is closed under complement,  $\bar{L} \in \text{NP}$ . Thus,  $\text{co-NP} \subseteq \text{NP}$ .

Assume  $\text{TAUT} \notin \text{NP}$ , then  $\overline{\text{TAUT}} \notin \text{co-NP}$ . As we have seen above,  $\overline{\text{TAUT}} \in \text{NP}$ . Hence  $\text{NP} \neq \text{co-NP}$ .  $\square$

The second lemma connects with the theory of proof systems by formulating a necessary and sufficient condition for a language being in  $\text{NP}$  [CR79].

**Lemma 6.** *A set  $L \neq \emptyset$  is in  $\text{NP}$  if and only if  $L$  has a polynomially bounded proof system.*

*Proof.* Assume  $L \in \text{NP}$ , then some nondeterministic Turing transducer  $M$  accepts  $L$  in polynomial time. Let  $f_i(x) \in \mathcal{FP}$  be the function calculating the  $i$ -th path of the nondeterministic calculation of  $M$ . We define  $f$  by

$$f(\langle i, x \rangle) = \begin{cases} x & (f_i(x) \text{ accepts}), \\ \perp & (\text{otherwise}). \end{cases}$$

Then  $f$  is a polynomially bounded proof system for  $L$ .

Conversely, assume  $f$  is a polynomially bounded proof system for  $L$ . Then a nondeterministic Turing transducer on input  $y$  can guess a proof  $x$  and verify  $f(x) = y$ .  $\square$

Putting lemma 5 and 6 together, we obtain proposition 1,

$$\text{NP} = \text{co-NP} \Leftrightarrow \text{there is a polynomially bounded proof system for TAUT.}$$

Using this theorem, one tried to separate  $\text{NP}$  from  $\text{co-NP}$  by studying more and more powerful proof systems, showing that they are not polynomially bounded [KMT03]. As mentioned before, there was no success on answering this questions until now. To take the notion of optimal proof systems into account, one could ask if there is an optimal or even  $p$ -optimal proof system for  $\text{TAUT}$ . If that were the case, then the existence of one specific proof system that is not polynomially bounded would suffice to proof that  $\text{NP} \neq \text{co-NP}$  and hence  $\text{P} \neq \text{NP}$  [KMT03].

Krajíček and Pudlák proved a sufficient condition for the existence of optimal proof systems for  $\text{TAUT}$  [KP89].

**Theorem 7.** *If  $\text{NE} = \text{co-NE}$  then optimal proof systems for  $\text{TAUT}$  exist. If  $\text{E} = \text{NE}$  then  $p$ -optimal proof systems for  $\text{TAUT}$  exists.*

We will omit their proof in this thesis, since it uses many notions of formal logics and a huge equivalence theorem not introduced here.

With these results, we will further investigate the question what languages possess optimal proof systems.

## 4 A set in $\text{co-NEXP} \setminus \text{OPT}$

In the main theorem of this thesis, we will show that in certain complexity classes above of NP, there are always languages that possess no optimal proof system. To formalize this, we say a function is *time-constructible*, if there is a Turing transducer that stops on input  $n$  after  $t(n)$  steps.

**Theorem 8.** *Let  $t : \mathbb{N} \rightarrow \mathbb{N}$  be a time-constructible function such that for every polynomial  $p$  there is a number  $n$  with  $p(n) \leq t(n)$ . Then there is a language  $L \in \text{co-NTIME}(t(n))$  that has no optimal proof system.*

Notice, if for function  $f$  there is an  $n \in \mathbb{N}$  with  $f(n) \geq p(n)$  for every polynomial  $p$ , then there are for every polynomial  $p$  infinitely many  $n \in \mathbb{N}$  for which  $f(n) \geq p(n)$  is satisfied. Otherwise, we could construct a polynomial  $q$  that is greater than  $f$  for every  $n \in \mathbb{N}$ . To see this, let  $p$  be a polynomial for which  $f(n) \geq p(n)$  for only finitely many  $n \in \mathbb{N}$ . Then we can determine the greatest difference between  $f$  and  $p$  by taking the maximum of the set  $\eta = \{f(n) - p(n), f(n) \geq p(n)\}$ . Then for the polynomial  $p'(n) = \eta + p(n) + 1$  there is no  $n$  with  $f(n) \geq p'(n)$ , which contradicts the assumption that for every polynomial  $p$  there is a natural number such that  $f(n) \geq p(n)$ .

Messner showed that under the same presumptions as in our theorem, there is a language  $L \in \text{co-NTIME}(t(n))$  without an optimal acceptor [Mes99]. He also proofed that the existence of a optimal acceptor is equivalent to the existence of a optimal proof system for every p-cylinder  $L$ . We will here give a proof that is based on the work of Messner, but stays in the notion of proof systems.

Before formally proofing theorem 8, let us briefly discuss the main idea of the proof. As a first observation, we notice that we can enumerate all  $\mathcal{FP}$ -functions  $f_i$  such that  $\text{time}(f_i) \leq n^i + i$ . We then construct a language  $L$  by uniting languages  $L_i$ . We will construct them in a way such that there are only long  $f_i$ -proofs for all strings in  $L_i$ . Afterwards we assume there is an optimal proof system  $g$  for  $L$ . As for every  $\mathcal{FP}$ -function there is a subset that has only long proofs, there is such a subset  $L_g$  for  $g$ . But as it turns out we can state a proof system for  $L_g$  that has short proofs, we can combine both proof systems to a system that runs polynomially fast on nearly every string in  $L$ , but way faster on strings in  $L_g$ . Hence,  $g$  cannot be an optimal proof system.

*Formal proof of theorem 8.* Let  $f_1, f_2, \dots$  be an enumeration of all  $\mathcal{FP}$ -functions with  $\text{time}(f_i) \leq n^i + i$ . We obtain this enumeration by enumerating all Turing transducers (using Gödel's enumeration) and combine each transducer with a clock that ensures a maximum runtime of  $n^i + i$  steps for a transducer  $M_i$ . For any  $i > 0$ , let  $L_i$  be the regular language described by the expression  $0^i 10^*$ . Define

$$L'_i = \{x \in L_i \mid \forall y \in \Sigma^* |y|^{2i} \leq t(|x|) \implies f_i(y) \neq x\}.$$

That is, as long as you put strings of length  $|y|^{2i} \leq t(|x|)$  into  $f_i$ , you will not obtain  $x$ . Let  $L = \bigcup_{i>0} L_i$ .

First, we obtain  $L \in \text{co-NTIME}(t(n))$ . To show this, one considers

$$L \in \text{co-NTIME} \Leftrightarrow \overline{L} = \overline{\bigcup_{i>0} L'_i} = \bigcap_{i>0} \overline{L'_i} \in \text{NTIME}.$$

By negating the condition for  $L'_i$ , we get

$$\overline{L'_i} = \{x \in \Sigma^* \mid x \notin L_i \vee (\exists y \in \Sigma^* \mid |y| \leq t(|x|) \wedge f_i(y) = x)\}.$$

For any given  $x$ , we can decide in polynomial time whether it is in any  $L_i$  or not. If it is not, then  $x$  is in  $\overline{L'_i}$  for all  $i > 0$  and therefore  $x \in \overline{L}$ , so we are done. If it is in any  $L_i$ , it is in exactly one  $L_i$ . Let  $i^*$  be the set with  $x \in L_{i^*}$ . We can simulate a polynomial-time machine calculating  $f_{i^*}(y)$  on every input  $y \in \Sigma^*$  with  $|y|^{2i} \leq t(|x|)$ . If, and only if, there is a path with  $f_{i^*}(y) = x$ , then  $x \in \overline{L}$ . In both cases,  $\overline{L} \in \text{NTIME}(t(n))$ .

have a look at simulation runtime

For a proof system  $f_i$  with  $f_i(\Sigma^*) = L$ , we observe that  $L'_i = L_i$ . Assume there is an  $x = 0^i 1z \in L_i$  that is not in  $L'_i$ . Then there is an  $y$  with  $|y|^{2i} \leq t(|x|)$  and  $f_i(y) = x$ . Since  $f_i$  is a proof system for  $L$ , this yields  $x = 0^i 1z \in L$  and so  $x \in L'_i$ , which contradicts the assumption. Therefore, for any  $y$  with  $f_i(y) = x \in L_i$  we know that  $|y|^{2i} > t(|x|)$ . Speaking informally, every proof system  $f_i$  for  $L$  has long proofs on  $L'_i \subset L$ .

Assume now, for contradiction, that  $f_i$  is a optimal proof system for  $L$ . Let  $g$  be a function defined as

$$g(bx) = \begin{cases} f_i(x) & (b = 0), \\ x & (b = 1 \text{ and } x = 0^i 10^* \in L_i = L'_i). \end{cases}$$

$g$  is a proof system for  $L$  with polynomial length-bounded proofs for all  $x \in L_i$ . As  $f_i$  is optimal, there is a function  $f^*$  such that for all  $x \in L'_i$ ,  $f_i(f^*(x)) = g(x)$  and  $|f^*(x)| \leq p(|x|)$  for a polynomial  $p$ . Let  $q$  be the polynomial  $q(n) = p(n)^{2i}$ . As  $p(|x|)$  is positive,  $p(|x|) \leq p(|x|)^{2i}$ . As there is an  $n$  with  $q(n) \leq t(n)$ , there is an  $x$  in  $L_i$  such that  $|f^*(x)| \leq p(|x|) \leq q(|x|) = p(|x|)^{2i} \leq t(|x|)$ . According to the definition of  $L'_i$ , this yields  $f_i(f^*(x)) \neq x$ . Therefore,  $f_i$  is not optimal on  $L'_i$ , which contradicts the assumption that  $f_i$  is optimal on  $L$ .  $\square$

Using the relation to many-one-hard reducible sets proofed in lemma 2, we obtain

**Corollary 9.** *No set  $\leq_m^p$ -hard for co-NE has an optimal proof system.*

*Proof.* With  $t(n) = 2^n$ , we can get an  $L \in \text{co-NE}$  that has no optimal proof system. Any  $\leq_m^p$ -hard set  $A$  for co-NE is  $L \leq_m^p A$ . Together with lemma 2 we obtain, that  $A$  cannot have optimal proof system.  $\square$

Now, let us take a closer look at this set  $L$  that has no optimal proof system. One first observation is that  $L$  is sparse. As every  $L'_i$  only contains strings that are of the form  $0^i 10^*$ ,  $L$  is a subset of the regular language  $L_R = 0^* 10^*$ . Therefore, the density of  $L_R$  is an upper bound for the density of  $L$ . In  $L_R$ , there are exactly  $n$  words of length  $n$ . As a consequence,  $\text{dens}_{L_R}(n) = \sum_{i=1}^n i = \frac{n^2+n}{2}$ . Hence,  $L_R$  and  $L$  are both sparse. We can construct even less dense languages by constructing tally sets without optimal proof systems.

**Theorem 10.** *If  $L \subseteq \{x \in \Sigma^*, x = 0^* 10^*\}$  has no optimal proof system, then there is a  $T \in \text{TALLY}$  that has no optimal proof system.*

*Proof.* Using the Cantor pairing function, we can construct a bijective function  $t \in \mathcal{FP}$  that maps each element of  $L$  one-to-one to an element of  $T \subseteq \text{TALLY}$ . Since all strings in  $L$  are almost TALLY—namely every string in  $L$  contains only one 1— $t$  is a polynomial time function. Let  $g$  be an optimal proof system for  $T = \{t(x), x \in L\}$ . Then  $t^{-1} \circ g$  is a proof system for  $L$ . Let  $h$  be an arbitrary proof system for  $L$ . Then  $t \circ h$  is a proof system for  $T$ . As  $g$  is optimal, there is a  $f \in \mathcal{FP}$  with  $g(f(x)) = t(h(x))$ . It follows that  $t^{-1}(g(f(x))) = h(x)$  for all proof systems  $h$  for  $L$ . Hence,  $t^{-1} \circ g$  is a optimal proof system for  $L$ , which contradicts the assumption.  $\square$

As with  $t \in \mathcal{FP}$  from this proof, we can state  $x \in L \Leftrightarrow t(x) \in T$ . Hence  $L \leq_m^p T$ . As  $t$  is bijective, we further obtain  $T \leq_m^p L$ . This yields the conjecture that  $L$  as constructed in our proof is not NP-complete. Otherwise,  $P = NP$  as we would have obtained an tally language that is NP-complete.

Using this theorem, we can further improve our results on the density of language not possessing an optimal proof system. For any tally language  $p(n) = n$  is an upper bound of its density.

**Corollary 11.** *There is a language  $L$  with  $\text{dens}_L(n) \leq n$  that possesses no optimal proof system.*

## 5 Conclusion and future work

What a great work!

# Bibliography

- [Coo71] Stephen A. Cook, *The complexity of theorem-proving procedures*, Proceedings of the third annual ACM symposium on Theory of computing (New York, NY, USA), STOC '71, ACM, 1971, pp. 151–158.
- [CR79] Stephen A. Cook and Robert A. Reckhow, *The relative efficiency of propositional proof systems*, Journal of Symbolic Logic **44** (1979), 36–50.
- [For09] Lance Fortnow, *The status of the P versus NP problem*, Commun. ACM **52** (2009), no. 9, 78–86.
- [KM00] Johannes Köbler and Jochen Messner, *Is the standard proof system for SAT p-optimal?*, Proceedings of the 20th Conference on Foundations of Software Technology and Theoretical Computer Science (London, UK, UK), FST TCS 2000, Springer-Verlag, 2000, pp. 361–372.
- [KMT03] Johannes Köbler, Jochen Messner, and Jacobo Torán, *Optimal proof systems imply complete sets for promise classes*, Inf. Comput. **184** (2003), no. 1, 71–92.
- [KP89] Jan Krajíček and Pavel Pudlák, *Propositional proof systems, the consistency of first order theories and the complexity of computations*, J. Symb. Log. **54** (1989), no. 3, 1063–1079.
- [Mes99] Jochen Messner, *On optimal algorithms and optimal proof systems*, Proceedings of the 16th annual conference on Theoretical aspects of computer science (Berlin, Heidelberg), STACS'99, Springer-Verlag, 1999, pp. 541–550.
- [Pap94] Christos H. Papadimitriou, *Computational complexity*, Addison-Wesley, 1994.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen Hilfsmittel und Quellen als die angegebenen benutzt habe. Weiterhin versichere ich, die Arbeit weder bisher noch gleichzeitig einer anderen Prüfungsbehörde vorgelegt zu haben.

Würzburg, den \_\_\_\_\_, \_\_\_\_\_  
(Nils Wisiol)