# Twelve Problems in Proof Complexity

Pavel Pudlák*

Mathematical Institute, Prague

## 1 Introduction

Proof complexity is a research area that studies the concept of complexity from the point of view of logic. Although it is very much connected with computational complexity, the goals are different. In proof complexity we are studying the question *how difficult is to prove a theorem?* There are various ways how one can measure the *"complexity"* of a theorem. We may ask what is *the length of the shortest proof* of the theorem in a given formal system. Thus the complexity is the size of proofs. This corresponds to questions in computational complexity about the size of circuits, the number of steps of Turing machines etc. needed to compute a given function. But we may also ask how strong theory is needed to prove the theorem. This also has a counterpart in computational complexity—the questions about the smallest complexity class to which a given set or function belongs.

Often the best way to find out what is going in some field of research is to look at open problems. Therefore my aim in this paper is to compile a list of problems in proof complexity that I consider to be important, but which also seem to be within the reach of our methods. With each problem, I shall define the necessary concepts and mention some related results.

The paper is intended for researchers in computational complexity who want to know what is going on in proof complexity and, perhaps, want to try some open problem there. Essentially all problems have already been stated before, sometimes in different forms. The reader interested in problems should consult monographs [6, 20], survey articles [11, 36] and other lists of problems [11, 21].

## 2 Frege systems

Most of my problems will be about propositional proof systems. I shall consider classical propositional logic, but I shall also mention some results about nonclassical propositional calculi. The general definition of a propositional proof system [12] is based on the following three conditions:

1. soundness;
2. completeness;
3. polynomial time decidability of the relation: *D is a proof of proposition* $\phi$.

* Supported by grants A1019401 and 1M002162080

Since the set of propositional tautologies is **coNP**-complete, there exists a proof system $P$ such that every tautology has a proof of polynomial length in $P$ if and only if $\mathbf{NP} = \mathbf{coNP}$. If $P$ has that property, I shall say that $P$ *is polynomially bounded.*

Although $\mathbf{NP} \neq \mathbf{coNP}$ is considered to be very likely true, we are not able to prove that some very basic proof systems are not polynomially bounded. These proof systems are called *Frege systems.* They are the well-known systems used in most textbooks and, in fact, fairly close to natural reasoning of mathematicians. A Frege system $P$ is based on a finite number of axiom schemas and deduction rules. A proof in $P$ is a string of propositions which are either instances of the axiom schemas or follow from previous ones by deduction rules.

There are two basic measures of complexity of Frege proofs. First, we may count the number of propositions in the proof; second, we may count the total length of an encoding of the proof as a binary string. This determines two measures of proof complexity of a tautology—the least *number of steps* in a proof and the *length* of the shortest proof. It has been shown that for every two Frege systems the the numbers of steps differ by at most a polynomial; the same holds for the length, [12]. In fact, when both systems use the same language (the same basis of connectives), then the lengths, resp. the numbers of steps, differ by at most a linear factor and the proof is trivial.

*Problem 1.* Prove a superpolynomial lower bound on the length of proofs for a Frege system (or prove that it is polynomially bounded).[1]

Essentially the only lower bound on the lengths of proofs in a Frege system is based on the simple observation that in a proof of an irreducible tautology $\tau$ [2] all subformulas must occur. Thus if the depth of $\tau$ is $n$ the size of every proof of $\tau$ is $\Omega(n^2)$. No lower bounds are known for the number of steps!

This is the most difficult of all problems I am going to state in this paper. As a matter of fact, I doubt that it is within the reach of the current methods, but it is worth mentioning it, before talking about its weaker versions and other related problems. The number of steps in a proof is always at most its length. It seems possible that there are tautologies with proofs that have only polynomial number of steps while they have only proofs of exponential length. So the problem to prove superpolynomial lower bounds on the number of steps for Frege proofs is even harder.

One possible weakening is to prove superpolynomial lower bounds using some complexity-theoretical assumptions. Of course, the assumptions must not imply $\mathbf{NP} \neq \mathbf{coNP}$, as that assumption implies that there is no polynomially bounded propositional proof system.

*Problem 2.* Prove a superpolynomial lower bound on the length of proofs for a Frege system *using a conjecture that does not imply* $\mathbf{NP} \neq \mathbf{coNP}$.

[1] I put the alternative into parenthesis, because I believe it is very unlikely. In the rest of the paper I shall omit such alternatives.

[2] This means that all subformulas of $\tau$ are essential for $\tau$ being a tautology.

One natural place to look for such conjectures is in theoretical cryptography. Several conjectures in that field are stronger than $\mathbf{P} \neq \mathbf{NP}$ but they are not known to imply $\mathbf{NP} \neq \mathbf{coNP}$. Solving this problem requires finding some computational consequences of the existence of short Frege proofs. I shall say more about it in the next section.

The difficulty of proving lower bounds for Frege proofs is caused not only by the lack of suitable methods, but also by the lack of suitable candidates for hard tautologies. Most of the tautologies based on simple combinatorial principles and theorems, such as the Pigeon-Hole Principle and the finite Ramsey theorem, have been shown to have polynomial size proofs. On the opposite end of the spectrum of various tautologies, there are tautologies that are almost surely hard. These are tautologies expressing the consistency of strong theories.[3] But for this kind of tautologies our combinatorial methods do not work. The method of diagonalization, which is so useful in predicate logic, completely fails in propositional logic. Therefore we need tautologies that are based on natural and sufficiently hard combinatorial principles.

A class of candidates for hard tautologies was proposed in [24, 2]. Let $F_n$ be a mapping from binary strings of length $n$ to binary strings of length $n + 1$. Then there is a string $b$ of length $n + 1$ which is not in the range of $F_n$. If $F_n$ is computable by a polynomial size circuit, we can define a polynomial size tautology $\tau_{F_n,b}$, for every $b \notin Rng(F_n)$, that expresses this property of $b$. The hope is that if $F$ is sufficiently pseudorandom, then $\tau_{F_n,b}$ is hard for every $b \notin Rng(F_n)$. We know that it does not suffice to assume that $F_n$ be a pseudorandom generator. In [45] Razborov stated specific properties of $F$ and conjectured that the tatuologies based on such functions are hard.

Furthermore, the hardness of these formulas, for some proof systems, has been conjectured for the following specific function $T$. Let $s(x)$ be a numeric functions such that $s(x) = o(x)$. Given a number $k$, interpret binary strings of length $2^{s(k)}$ as codes of boolean circuits defining functions of $k$ variables; interpret binary strings of length $2^k$ as truth tables of boolean functions of $k$ variables. Then $T$ is the mapping that, given a string $c$ of length $2^{s(k)}$, maps $c$ onto the truth table of the function computed by the circuit encoded by $c$. So, roughly speaking, the conjecture is that it is hard to prove lower bounds on circuit complexity.

Another specific function was studied in [26].

The connection to Frege proofs is not quite clear to me. In fact, it is conceivable that for suitable sequences of functions $\{F_n\}$ and strings $\{b_n\}$, the tautologies $\tau_{F_n,b_n}$ are hard for *all* proof systems. This conjecture is buttressed by the fact that for all proof systems for which superpolynomial lower bounds have been obtained, also superpolynomial lower bounds have been proved for formulas of this type.

[3] For a theory $T$ it is a sequence of tautologies $\{\tau_n\}$, where $\tau_n$ expresses that no string of length $n$ is a proof of contradiction from the axioms of $T$.

The problem of proving lower bounds on the size and the number of steps for Frege proofs has been studied for nonclassical logics too. Quite recently Pavel Hrubeš proved exponential lower bounds on the number of steps in propositional intuitionistic logic and in several modal logics [16, 17].

Hrubeš's results cover a lot of modal logics, still there are other nonclassical logics for which no lower bounds on the lengths of proofs are known. One that I find particularly interesting is *orthomodular logic.*

*Problem 3.* Prove lower bounds on proofs in a Frege system for orthomodular logic.

Orthomodular logic is, roughly speaking, classical logic with distributivity replaced by the weaker law of modularity [13]. This is one of the logics studied in the field of quantum logic, but quantum physics and quantum computation are not my motivations. The reason why I think this problem should be studied is its connection to proofs of lower bounds on classical Frege systems. Connections with structures studied in quantum mechanics were mentioned already in [23]. One can show that if a Frege proof system for orthomodular logic is polynomially bounded, then so is every Frege proof system for classical logic.[4] I am proposing Problem 3 as a weaker version of the central Problem 1, but it may turn out that they are equivalent.

The weaker *orthologic* is also interesting in connection with lower bounds on classical Frege systems, however notice that it has polynomially bounded proof systems [15].

## 3 Feasible interpolation

Feasible interpolation, a.k.a. effective interpolation, was invented by Jan Krajíček [19]. It is a way to obtain, from a short proof, some effective computation. In particular, to obtain a polynomial size circuit computing a function related to a suitable tautology from its polynomial size proof. This enables one to reduce the problem of proving lower bounds on the size of proofs to proving lower bounds on the size of circuits.

Let $\alpha(\bar{p}) \vee \beta(\bar{q})$ be a tautology where $\bar{p}$ and $\bar{q}$ are disjoint sets of propositional variables. Then either $\alpha(\bar{p})$, or $\beta(\bar{q})$, or both are tautologies. More generally, if $\alpha(\bar{r}, \bar{p}) \vee \beta(\bar{r}, \bar{q})$ is a tautology, then for every assignment of truth values $\bar{a}$ to $\bar{r}$, either $\alpha(\bar{a}, \bar{p})$ or $\beta(\bar{a}, \bar{q})$ or both are tautologies.

**Definition 1.** *A proof system $P$ has the* feasible interpolation property*, if there exists a polynomial time algorithm $A$ which outputs either $0$ or $1$ and such that given a proof $D$ of $\alpha(\bar{r}, \bar{p}) \vee \beta(\bar{r}, \bar{q})$ and a truth assignment $\bar{a}$,*

1. *if $A$ outputs $0$, then $\alpha(\bar{a}, \bar{p})$ is a tautology;*
2. *if $A$ outputs $1$, then $\beta(\bar{a}, \bar{q})$ is a tautology.*

[4] Thomas Vetterlein, personal communication.

It should be noted that the feasible interpolation property is also equivalent to the following property. *There exists a polynomial time algorithm which from a proof $D$ of $\alpha(\bar{r},\bar{p}) \vee \beta(\bar{r},\bar{q})$ constructs a circuit $C(\bar{r})$ such that for every truth assignment $\bar{a}$*

1. *if $C(\bar{a}) = 0$, then $\alpha(\bar{a},\bar{p})$ is a tautology;*
2. *if $C(\bar{a}) = 1$, then $\beta(\bar{a},\bar{q})$ is a tautology.*

A number of propositional proof systems posses this property. The first one for which this was established was the cut-free sequent calculus, soon after for the propositional Resolution system and others. This property was also shown for Frege systems for some nonclassical logics, including intuitionistic logic. For nonclassical logics, however, one has to modify it a little. For instance, in case of intuitionistic logic one has to consider tautologies of the form

$$(r_1 \vee \neg r_1) \wedge \ldots \wedge (r_n \vee \neg r_n) \rightarrow \alpha(\bar{r},\bar{p}) \vee \beta(\bar{r},\bar{q}).$$

For Frege systems for classical logic it has been shown that the property fails, assuming some likely conjectures, eg., that factoring integers is hard (ie., not solvable in polynomial time) [29, 5].

It is not difficult to prove that the feasible interpolation property is equivalent to separation of some disjoint **NP** sets in the following sense (which I state in a bit informal way).

**Proposition 1.** *A proof system $P$ has the feasible interpolation property if and only if whenever $P$ proves that two* **NP** *sets $A$ and $B$ are disjoint using a sequence of polynomial size proofs, then $A$ and $B$ can be separated by a set in* **P/poly**, *(ie., $\exists C \in \mathbf{P/poly}(A \subseteq C \wedge B \cap C = \emptyset)$.*

Consequently, if $\mathbf{NP} \cap \mathbf{coNP} \not\subseteq \mathbf{P/poly}$, then the feasible interpolation property implies that the proof system is not polynomially bounded. This assumption is not known to imply $\mathbf{NP} \neq \mathbf{coNP}$. Thus we get conditional superpolynomial lower bounds using a condition different from $\mathbf{NP} \neq \mathbf{coNP}$. In many cases, however, after proving the feasible interpolation property also unconditional exponential lower bounds have been proved.

Since this method proved to be extremely useful for proving lower bounds, my questions concern the possibility of extending it to stronger systems. I shall state the problems only for Frege systems, but they are meaningful for every system for which we do not have the feasible interpolation property. The first problem is about the possibility to replace the separation using sets in **P/poly** by separation using more complex sets.

*Problem 4.* Prove that Frege systems have the feasible interpolation property in a more general sense, namely, with the separation using sets in **P/poly** replaced by separation using sets in a larger complexity class.

As shown in [37], it suffices to determine how difficult is to separate the canonical pair of **NP** sets associated with a Frege system. Recall that the canonical pair of a proof system $P$ is the pair of the following two **NP** sets [41]:

$$Prov(P) := \{(\phi, 0^n)\ ;\ \phi \text{ has a proof of length at most } n \text{ in } P\}$$
$$NegSat := \{(\phi, 0^n)\ ;\ \neg\phi \text{ is satisfiable}\}.$$

(The string $0^n$ only serves for padding the input to make it of length at least $n$.) Thus the Problem 4 asks for a nontrivial upper bound on the complexity of sets that separate the canonical pair of a Frege system. This connection also shows an important fact about the problem: *we do not not have to consider general proofs, but only certain very concrete ones.* In Frege systems the propositional translations of the sentence

$$Prov(\mathbf{F}) \cap NegSat = \emptyset$$

have polynomial size proofs, where **F** denotes some fixed Frege system.[5] These are the proofs that we only need to consider.

Notice that a solution of Problem 4 would also be a solution of Problem 2. There may be other ways one can generalize feasible interpolation so that it also holds for Frege systems, the form of which may eventually have little to do with the original concept of interpolation. Therefore I shall state a version of the previous problems in a very general form.

*Problem 5.* Derive any nontrivial computational consequences from the existence of a small Frege proof.

There is another reason for posing the problem in this way. It is well known that intuitionistic logic and some other related logics are *constructive.* This means, roughly speaking, that one can interpret proofs as algorithms. This mainly concerns predicate logic, but there are results of this kind also for propositional logic [8]. So the above problem can be paraphrased: *Does classical propositional logic have any constructive properties?*

Problems 2,4 and 5 are very much related, one should view them as possible ways of attacking the central Problem 1.

Since all problems about Frege systems seem to be very hard, one should start with some special cases. In case of the problems about the feasible interpolation, one should start with bounded depth Frege systems. Assuming likely conjectures, they do not have the feasible interpolation property already for small depths [5]. It would be interesting to find some generalized interpolation for as weak a system as $Res(log)$, which is a generalization of Resolution, based on disjunctions (clauses) of conjunctions of logarithmic lengths.

## 4 The Bounded Arithmetic Hierarchy

I shall start with a topic that at first will seem completely unrelated to previous problems. For $n \geq 0$, $T_2^n$ denotes the theory axiomatized by induction axioms restricted to $\Sigma_n^b$ formulas. The class of $\Sigma_n^b$ defines precisely the class of sets $\Sigma_n^p$ of the *Polynomial Hierarchy*. The *Bounded Arithmetic Hierarchy* is the sequence of theories $\{T_2^n\}$. Roughly speaking, $T_2^n$ formalizes reasoning that uses

[5] These sentences are also known as the *Reflection Principle for the Frege System,* see [37].

only concepts from the $n$-th level of the Polynomial Hierarchy. (See [6, 20] for definitions.)

We proved that if the Polynomial Hierarchy is strictly increasing, then so is the Bounded Arithmetic Hierarchy [27]. Furthermore, we proved that the relativized Bounded Arithmetic Hierarchy is strictly increasing. That results, however, only show that there are $\Sigma^b_{n+2}$ sentences that separate $T_2^{n+1}$ from $T_2^n$, resp. the same for the relativized case.[6] It is an open problem whether one can prove similar results for sentences of fixed complexity.

*Problem 6.* (I) Assuming some reasonable conjecture in computational complexity theory, for some $k$, prove or disprove that for all $n \geq 0$, $T_2^{n+1}$ proves more $\Sigma^b_k$ sentences than $T_2^n$.

(II) The same for relativized theories $T_2^n[R]$, without using unproven assumptions.

The relativized theories $T_2^n[R]$ are extensions of $T_2^n$ obtained by adding a new predicate $R$ and extending the induction axioms to $\Sigma^b_n[R]$; there are no specific axioms for $R$. The predicate $R$ plays a similar role as oracles in relativized complexity classes and this connection is actually used for separation results.

Recently, most research activities focused on the $\Sigma^b_1$ sentences provable in theories $T_2^n$. These sentences are related to a very natural concept in computational complexity theory.

**Definition 2.** *A* total **NP** search problem *is determined by a relation $R \in$* **P** *and a polynomial $p$ such that*

$$\forall x \exists y (|y| \leq p(|x|) \wedge R(x, y)). \tag{1}$$

The sentence (1), which expresses that the search problem is total, is a $\Sigma^b_1$ sentence. For $T_2^0$, the total search problems corresponding to the provable $\Sigma^b_1$ sentences, are solvable in polynomial time. For $T_2^1$ the search problems belong to the well-known class *Polynomial Local Search,* **PLS**. Characterizations of the search problems of higher levels of the Bounded Arithmetic Hierarchy were obtained quite recently [38, 31, 46]. I shall describe the simplest characterization, which is due to Skelley and Thapen [46].

An *$n$-game* is an $n$-ary relation $G(x_1, \ldots, x_n)$. We think of it as played by two players, A is starting and B playing as the second. The players alternate in picking $x_i$'s; B wins if $G(x_1, \ldots, x_n)$ holds true, otherwise A wins. The concept of a winning strategy is well-known. Further, we need the concept of a *reduction of an $n$-game $G$ to an $n$-game $H$*. It is a string of functions $f_1, \ldots, f_n$ such that for every $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ such that $y_i = f_i(x_1, x_3, \ldots, x_i)$ for $i$ odd, and $x_i = f(y_2, y_4, \ldots, y_i)$ for $i$ even, if $H(y_1, \ldots, y_n)$, then $G(x_1, \ldots, x_n)$.

If we have a wining strategy for B in $H$ and a reduction of $G$ to $H$, then we obtain a winning strategy for B in $G$ by simply composing the strategy with the reduction.

[6] More precisely, we should denote these sentences by $\forall\Sigma_{n+2}$, as we are talking about the universal closures of $\Sigma_{n+2}$ sentences.

The principle $GI_n$ says that the following is *impossible:*

> *There are games $G_0, \ldots, G_a$, a winning strategy $\alpha$ for A in game $G_0$, reductions $\rho_i$ of $G_{i+1}$ to $G_i$ for $i = 0, \ldots, a-1$ and a winning strategy $\beta$ for B in $G_a$.*

Indeed, if we compose $\beta, \rho_{a-1}, \rho_{i-2}, \ldots, \rho_0$, we obtain a winning strategy for B in $G_0$ contradicting to the existence of a winning strategy for A in $G_0$.

The total **NP** search problem associated with $GI_n$ is defined using circuits. The games $G_0, \ldots, G_a$ are given by a circuit $C(z, x_1, \ldots, x_n)$, where for the binary string $\bar{i}$ representing index $i$, $0 \leq i \leq a$, $C(\bar{i}, x_1, \ldots, x_n)$ defines game $G_i$. Similarly, the reductions $\rho_i$, $0 \leq i \leq a$, are given by one circuit. Further, we have a circuit defining strategy $\alpha$ and a circuit defining strategy $\beta$. Notice that the number of games $a$ is, in general, exponential in the size of input.

The task of the search problem is, given the circuits, to find out what is wrong. Namely, we should find

1. either $x_1, \ldots, x_n$ that show that $\alpha$ is not a winning strategy for A in $G_0$,
2. or $i$, $0 \leq i < a$, $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ that show that $\rho_i$ is not a reduction of $G_{i+1}$ to $G_i$,
3. or $y_1, \ldots, y_n$ that show that $\beta$ is not a winning strategy for B in $G_a$.

This is also the way in which $GI_n$ is formalized as a $\Sigma_1^b$ sentence. I shall use the same notation for the principles, their formalizations and the associated search problems.

**Theorem 1 ([46]).** *For $n \geq 1$, $GI_n[R]$ characterizes $\Sigma_1^b[R]$ consequences of $T_2^n[R]$ (hence also $GI_n$ characterizes $\Sigma_1^b$ consequences of $T_2^n$).*

A search problem $S$ is *polynomially reducible* to a search problem $S'$, if we can solve $S$ in polynomial time using queries to an oracle that produces solutions of the queried instances $S'$. If $\Sigma_1^b$ theorems of $T_2^{n+1}[R]$ are the same as $\Sigma_1^b$ theorems of $T_2^n[R]$, then for every oracle $A$, $GI_{n+1}^A$ is polynomially reducible to $GI_n^A$ (using also the oracle $A$). In other words, that assumption implies that $GI_{n+1}$ is polynomial reducible to $GI_n$ and the proof relativizes. This enables us to reduce the Problem 6 to a purely computational one.

*Problem 7.* For $n = 1, 2, \ldots$, find an oracle $A$ such that the search problem $GI_{n+1}^A$ is not reducible to the search problem $GI_n^A$.

I have stated this problem for a specific characterization, but one can try other characterizations of $\Sigma_1^b$ theorems of theories $T_2^n$. This is only a matter of convenience, all these problems are equivalent.

One should not forget about the unrelativized case.

*Problem 8.* For $n = 1, 2, \ldots$, find a reasonable conjecture in complexity theory which implies that the search problem $GI_{n+1}$ is not reducible to the search problem $GI_n$.

A solution of this problem may be a clue for solving the previous problem. Specifically, if the conjecture used in a solution to Problem 8 can be proved when relativized by an oracle, then we get a solution to Problem 7.

## 5 Bounded depth Frege systems

Bounded depth circuits are an important class of circuits studied in computational complexity, and exponential lower bounds on the size of such circuits computing explicitly defined functions have been proved. A related concept has been studied in proof complexity. A *depth d Frege system* is a Frege system in which only formulas of depth $d$ are allowed. As in case of bounded depth circuits, the depth of Frege proofs is the the maximal number of alternations of $\wedge, \vee$ and $\neg$ in a formula of the proof (we assume that no other connectives are used).

Let $\mathbf{F}$ denote some Frege systems and $\mathbf{F}_d$ its depth $d$ restrictions. Following the brake-through superpolynomial lower bound of Ajtai [1], exponential lower bounds on bounded depth Frege proof have been proved [19, 28, 34]. Specifically, for every fixed $d$, the tautology expressing the Pigeon-Hole Principle has only exponentially long proofs in $\mathbf{F}_d$. Superpolynomial separations of $\mathbf{F}_d$ from $\mathbf{F}_{d+1}$ was proved in [30] using padded Pigeon-Hole tautologies, which are of depth 2. Also exponential separation of $\mathbf{F}_d$ from $\mathbf{F}_{d+1}$ is known [19], but it uses tautologies of maximal depth that is possible in these systems.[7] The following is still an open problem.

*Problem 9.* Does there exist a $k$ such that for every $d \geq k$ there exists a sequence of tautologies of depth $k$ that have polynomial size proofs in $\mathbf{F}_{d+1}$, but which do not have proofs of size $2^{(\log n)^{O(1)}}$ in $\mathbf{F}_d$?

We believe that the answer to this problem is positive with $k = 2$ and with a lower bound $2^{n^{\Omega(1)}}$. But why do we need a lower bound $2^{(\log n)^{\omega(1)}}$? It is because such a lower bound would help us solve Problem 7. The statement of Problem 9 does not formally imply the statement of Problem 7. To get such a relationship we would have to insist that the sequences of tautologies are uniform in a certain well defined sense. Namely, the tautologies should be propositional translations of $\Sigma_k^b$ sentences. However, it seems unlikely that the use of nonuniform sequences of families could help.

The candidate tautologies are the translations of the $\Sigma_1^b$ sentences that characterize sentences provable in $T_2^d$. But even the simplest ones, the $GI_d$ are fairly complicated which is the reason why researcher have not studied their propositional translations. For a few small depths we have simpler candidates. In particular, the minimal depth in which one can prove the finite Ramsey theorem is an open problem.

The only general lower bound technique for bounded depth Frege proofs that we have is based on Switching Lemmas [3]. Formally they look very much like the classical Switching Lemma of Yao and Håstad and the proof techniques are similar, but there are additional technical complications. Every tautology requires a lemma of a specific form, thus the more complicated the tautology is,

[7] The depth of these tautologies is $d + 1$, therefore one has to use $\mathbf{F}_d$ as a refutation system, in order to be able to prove such tautologies, or one can formalize $\mathbf{F}_d$ as a sequent system with arbitrary formulas and the cut rule restricted to depth $d$ formulas.

the more complicated the lemma is. Another problem is that these lemmas do not have interpretations in finite domains, thus we have to treat them purely syntactically, or use nonstandard models.

In boolean complexity theory exponential lower bounds have been proved for a larger class of circuit. A $MOD_q^n$ gate is a boolean function of $n$ variables, whose value is 1 if and only if the number of ones in the input string is divisible by $q$. Razborov and Smolensky considered bounded depth circuits with ANDs, ORs, NOTs and gates $MOD_p^n$ with $p$ prime, and proved exponential lower bounds on the circuit size of some explicitly defined functions [40, 47]. After the method of random restrictions had been adapted for bounded depth Frege proofs and exponential lower bounds had been proved, researchers in proof complexity attempted to prove lower bounds on the more general type of bounded depth Frege system in which $MOD_p^n$ gates, for $p$ prime, were allowed. But an adaptation of the approximation method turned out to be much harder, if not impossible. So the following is still an open problem. Let $\mathbf{F}_d[p]$ denote a suitable depth $d$ Frege system that uses gates AND, OR, NOT and $MOD_p$.

*Problem 10.* Prove superpolynomial lower bounds on $\mathbf{F}_d[p]$ proofs for $p$ prime.

I am not considering $\mathbf{F}_d[q]$ systems with $q$ composite, although such systems can be defined, because superpolynomial lower bounds for bounded depth circuits with $MOD_q$, $q$ composite, is a widely open problem, and we expect that the corresponding problem in proof complexity will be even harder to solve.

I will now explain what is the obstacle to adapting the method of approximation to $\mathbf{F}_d[p]$. Let us first recall how the lower bounds on bounded depth circuits with modular gates are proved. The basic idea is to approximate functions computed at the gates of the circuit by low degree polynomials. Then one shows that the precision of the approximation deteriorates slowly, thus the output function should be approximated well, assuming the circuit is small. Finally, one proves that such an approximation does not exist for the given function.

Given an $\mathbf{F}_d[p]$ proof, we would like to mimic the above reasoning. So we would like to associate low degree polynomials with formulas in the proof and show that the polynomials approximate axioms very well and the precision of the approximation decreases slowly in the course of the proof. But what does it mean to approximate a formula in a proof? If we count truth assignments as in the proof for bounded depth circuits we get nowhere. Each formula in the proof is a tautology, hence it is trivially approximated by the constant 1 (which is a zero degree polynomial). According to our experience from the proofs for $\mathbf{F}_d$, we have to consider "imaginary" truth assignments that falsify the tautology $\phi$ for which we want to prove a lower bound. Such assignments do not exist in real world (since $\phi$ is a tautology), we can only imagine them, or we have to use nonstandard models. If we use a nonstandard model $\mathcal{M}$, then the imaginary assignments are represented by real objects, but they have to be *external* to $\mathcal{M}$. Then, for a polynomial $p$, we need to express in $\mathcal{M}$ on how many of these external assignments $p$ vanishes. That is the problem, because there is no natural way how to count external objects inside of $\mathcal{M}$.

This is the main stumbling block when one tries to translate the approximation method in a straightforward way. Researchers tried several other ways, but they only obtained partial results. The *Polynomial Calculus* was proposed as the most rudimentary special case of $\mathbf{F}_d[p]$ proofs [10]. Exponential lower bounds have been proved for this system [44] and a reduction to lower bounds for the Polynomial Calculus extended by certain axioms has been found [7].[8] Lower bounds for a system that combines $\mathbf{F}_d$ with the Polynomial Calculus were proved in [22].

## 6 Integer Linear Programing

The general form of an Integer Linear Programing problem is: *for a given set of inequalities with rational coefficients, find solutions in the domain of integers.* If we want to study the complexity of Integer Linear Programing, we can simplify it by considering only the decision problem: *does the system of inequalities have an integral solution?* It is well-known that this problem is **NP**-complete.

From the point of view of proof complexity, the most interesting problem is: *how difficult is to prove that a given set of inequalities does not have an integral solution?* Since it is a **coNP**-complete problem, we believe that proofs in any proof system must be exponentially large. Since we are not able to prove this conjecture in general, we would like to prove it at least in some special cases, ie., for particular proof systems.

Exponential lower bounds have been proved for two systems [35, 14]. Furthermore, exponential lower bounds have been obtained for several other systems for *tree-like proofs,* see [18], and [4] combined with the recent bounds on multiparty communication complexity of disjointness [32, 9]

I shall describe in more detail one proof system which seems within the reach of our methods; it is the *Lovász-Schrijver system* [33]. We want to prove the unsatisfiability of a system of linear inequalities $\{L_i \geq 0\}_{i=1}^m$ by integers. The initial inequalities are:

1. $L_i \geq 0$, $i = 1, \ldots, m$;
2. $x_j^2 - x_j \geq 0$, for any variable $x_j$ used in $\{L_i \geq 0\}_{i=1}^m$.

A proof is a sequence of inequalities derived from the initial inequalities by the rules of the system, ending with the contradictory inequality $-1 \geq 0$. The inequalities in the proof are of degree at most 2. The rules are:

1. we can derive any positive linear combination of established inequalities;
2. from a linear inequality $L \geq 0$, we can derive $x_j L \geq 0$, for any variable $x_j$ used in $\{L_i \geq 0\}_{i=1}^m$;
3. from a linear inequality $L \geq 0$, we can derive $(1 - x_j)L \geq 0$, for any variable $x_j$ used in $\{L_i \geq 0\}_{i=1}^m$.

[8] More precisely, it is a reduction to an extension of a weaker system called the *Nullstellensatz System.*

Exponential lower bounds on tree-like proofs follow from the aforementioned results, for general proofs (DAG-like), however, it is an open problem.

*Problem 11.* Prove superpolynomial lower bounds on Lovász-Schrijver proofs.

For the Lovász-Schrijver system the feasible interpolation property has been proved [35], thus we know that the system is weak. As I have mentioned, in many cases unconditional lower bounds were found after the feasible interpolation property had been established. These lower bounds are based on monotone versions of the feasible interpolation property, in which monotonic computational models are used instead of boolean circuits. For the Lovász-Schrijver proof system the following monotonic model is needed. I call it *monotone linear programs for computing boolean functions.* Such a program $P$ is given by a set of inequalities of the form:

$$\sum_j a_{ij} z_j \leq \sum_k b_{ik} x_k + c_i$$

where $a_{ij}, b_{i,k}, c_i \in \mathbf{Q}$ are constants, $b_{i,k} \geq 0$, and $z_j, x_k$ are variables. Variables $x_k$ are used for 0–1 inputs. $P$ computes the boolean function $f(\bar{x})$ that for every string of zeros and ones $\bar{d}$ satisfies:

$$P_{\bar{x}:=\bar{d}} \text{ has a solution, iff } f(\bar{d}) = 1.$$

The solution is for the variables $z_j$ and we require $z_j \geq 0$. Notice that $P$ computes a monotone boolean function because of the condition $b_{i,k} \geq 0$. Without this condition the model would be as efficient as general boolean circuits.

Solving the following problem positively would be a major step towards proving superpolynomial lower bounds on Lovász-Schrijver proofs.

*Problem 12.* Prove a superpolynomial lower bound on the size of a monotone linear program computing an explicitly defined monotone boolean function.

The problem is important also for computational complexity, since monotone linear programs are the strongest monotonic computational model that has been defined.

**Acknowledgment**

I thank Jan Krajíček for discussing the problems and his remarks to the draft of this paper. It should be noted, however, that we do not fully agree on which problems are the most important in proof complexity.

## References

1. M. Ajtai, *The complexity of the pigeonhole principle.* In: Proc. IEEE 29th Annual Symp. on Foundation of Computer Science, 1988, 346-355.

2. M. Alekhnovich, E. Ben-Sasson, A. A. Razborov and A. Wigderson, *Pseudorandom Generators in Propositional Proof Complexity.* SIAM Journal on Computing, Vol. 34, No 1, 2004, 67-88.
3. P. Beame, *A switching lemma primer.* Technical Report UW-CSE-95-07-01, Department of Computer Science and Engineering, University of Washington, November 1994.
4. P. Beame, T. Pitassi and N. Segerlind, *Lower Bounds for Lovsz-Schrijver Systems and Beyond Follow from Multiparty Communication Complexity.* ICALP 2005: 1176-1188
5. Bonet, M., Pitassi, T. and Raz, R., *No feasible interpolation for $TC^0$-Frege proofs,* Proc. 38-th FOCS, 1997, 254-263.
6. S.R. Buss, *Bounded Arithmetic,* Bibliopolis, 1986.
7. S. Buss, R. Impagliazzo, J. Krajíček , P. Pudlák, A.A. Razborov and J. Sgall, *Proof complexity in algebraic systems and constant depth Frege systems with modular counting,* Computational Complexity 6, (1996/1997), 256-298
8. S. Buss and P. Pudlák, *On the computational content of intuitionistic propositional proofs.* Annals of Pure and Applied Logic 109 (2001), 49-64.
9. A. Chattopadhyay and A. Ada, *Multiparty Communication Complexity of Disjointness.* arXiv e-print (arXiv:0801.3624)
10. M. Clegg, J. Edmonds, R. Impagliazzo, *Using the Groebner basis algorithm to find proofs of unsatisfiability,* Proc. 28-th ACM STOC, 1996, 174-183.
11. P. Clote and J Krajíček, *Open Problems.* In: Arithmetic, Proof Theory and Computational Complexity, eds. P. Clote and J. Krajíček, Oxford Press, (1993), 1-19.
12. S.A.Cook and R.A. Reckhow, *The relative efficiency of propositional proof systems,* Journ. Symbolic Logic 44, 1987, 25-38.
13. M. L. Dalla Chiara and R. Giuntini, *Quantum Logics.* In: Handbook of Philosophical Logic, Gabbay and Guenthner eds., Kluwer, 2002, 129-228.
14. S. Dash, *An Exponential Lower Bound on the Length of Some Classes of Branch-and-Cut Proofs.* IPCO 2002: 145-160
15. U. Egly and H. Tompits, *On different proof-search strategies for orthologic.* Stud. Log. 73 (2003), 131-152.
16. P. Hrubeš, *A lower bound for intuitionistic logic.* Ann. Pure Appl. Logic 146(1), 2007, 72-90.
17. P. Hrubeš, *Lower bounds for modal logics.* To appear in Journ. Symbolic Logic.
18. A. Kojevnikov, D. Itsykson, *Lower Bounds of Static Lovsz-Schrijver Calculus Proofs for Tseitin Tautologies.* ICALP (1) 2006: 323-334
19. J. Krajíček, Lower Bounds to the Size of Constant-Depth Propositional Proofs", J. of Symbolic Logic, 59(1), (1994), pp.73-86.
20. J. Krajíček, *Bounded Arithmetic, Propositional Logic, and Complexity Theory.* Encyclopedia of Mathematics and its Applications 60, Cambridge Univ. Press, 1995.
21. J. Krajíček, *A fundamental problem of mathematical logic.* Collegium Logicum, Annals of Kurt Gödel Society Vol. 2, 1996, 56-64.
22. J. Krajíček, *Lower bounds for a proof system with an exponential speed-up over constant-depth Frege systems and over polynomial calculus.* In: Eds. I. Prívara, P. Růžička, 22nd Inter. Symp. Mathematical Foundations of Computer Science, LNCS 1295, Springer-Verlag, (1997), 85-90.
23. J. Krajíček, *On methods for proving lower bounds in propositional logic.* M.L. Dalla Chiara et al. (eds.), Logic and Scientific Methods, Kluwer Acad. Publ., 69-83.
24. J. Krajíček, *On the weak pigeonhole principle.* Fundamenta Mathematicae, Vol.170(1-3), (2001), 123-140.

25. J. Krajíček, *Proof complexity.* In: Laptev, A. (ed.), European congress of mathematics (ECM), Stockholm, Sweden, June 27–July 2, 2004. Zurich: European Mathematical Society, (2005), 221-231.
26. J. Krajíček, *A proof complexity generator.* In: Proc. 13th Int. Congress of Logic, Methodology and Philosophy of Science (Beijing, 2007), King's College Publications, London, ser. Studies in Logic and the Foundations of Mathematics. Eds. C.Glymour, W.Wang, and D.Westerstahl. To appear.
27. J. Krajíček, P. Pudlák and G. Takeuti, *Bounded Arithmetic and the Polynomial Hierarchy.* Annals of Pure and Applied Logic 52, 1991, 143-153.
28. J. Krajíček, P. Pudlák and A. Woods, *Exponential lower bound to the size of bounded depth Frege proofs of the pigeonhole principle.* Random Structures and Algorithms, 7(1), (1995), 15-39.
29. J. Krajíček and P. Pudlák, *Some consequences of cryptographical conjectures for $S^1_2$ and $EF$.* Information and Computation 140, 1998, 82-94.
30. J. Krajíček and R. Impagliazzo, *A note on conservativity relations among bounded arithmetic theories.* Mathematical Logic Quarterly, 48(3), (2002), 375-377.
31. J. Krajíček A. Skelley and N. Thapen, *NP search problems in low fragments of bounded arithmetic.* J. of Symbolic Logic, 72(2), (2007), 649-672.
32. T. Lee and A. Schraibman, *Disjointness is hard in the multi-party number on the forehead model.* arXiv e-print (arXiv:0712.4279)
33. L. Lovász and A. Schrijver, *Cones of matrices and set-functions and 0-1 optimization.* SIAM J. Optimization 1(2), 1991, 166-190.
34. T. Pitassi, P. Beame and R. Impagliazzo, *Exponential Lower Bounds for the Pigeonhole Principle.* Computational Complexity 3, 1993, 97-140.
35. P. Pudlák, *Lower bounds for resolution and cutting planes proofs and monotone computations.* J. Symbolic Logic 62/3, 1997, 981-998.
36. P. Pudlák, *The lengths of proofs.* In: Handbook of Proof Theory, Elsevier, 1998, 547-637.
37. P. Pudlák, *On reducibility and symmetry of disjoint NP-pairs.* Theor. Comput. Science 295, (2003), 323-339,
38. P. Pudlák, *Consistency and games—in search of new combinatorial principles.* In Proc. Logic Colloquium'03, Helsinki, eds. V. Stoltenberg-Hansen and J. Vaananen, Assoc. for Symbolic Logic 2006, 244-281.
39. P. Pudlák and J. Sgall, *Algebraic models of computation and interpolation for algebraic proof systems.* In: Proof Complexity and Feasible Arithmetics, P.W. Beame and S.R. Buss eds., DIMACS Series in Discrete Mathematics and Theoretical Computer Science 39, 279-295
40. A.A. Razborov, *Lower bounds on the size of bounded-depth networks over a complete basis with logical addition.* (Russian), in Matematicheskie Zametki, Vol. 41, No 4, 1987, pages 598-607. English translation in Mathematical Notes of the Academy of Sci. of the USSR, 41(4):333-338, 1987.
41. A.A. Razborov, *On provably disjoint NP-pairs.* BRICS Report Series RS-94-36, 1994, http://www.brics.dk/RS/94/36/index.html.
42. A.A. Razborov, *Unprovability of lower bounds on the circuit size in certain fragments of Bounded Arithmetic,* Izvestiya of the R.A.N. 59(1), (1995), 201-222; see also Izvestiya: Mathematics 59:1, 205-227.
43. A. A. Razborov, *Bounded Arithmetic and Lower Bounds in Boolean Complexity.* In: Feasible Mathematics II, Birkhuser Verlag, 1995, 344-386.
44. A.A. Razborov, *Lower bound for the polynomial calculus.* Computational Complexity, Vol. 7, No 4, 1998, 291-324.

45. A.A. Razborov, *Pseudorandom Generators Hard for k-DNF Resolution and Polynomial Calculus Resolution.* 2002-2003, http://www.mi.ras.ru/ razborov/res k.ps
46. A. Skelley and N. Thapen, *The provably total search problems of Bounded Arithmetic.* Preprint 2007.
47. R. Smolensky, *Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity.* STOC 1987: 77-82.