# Lower bounds for resolution and cutting plane proofs and monotone computations*

Pavel Pudlák†
Mathematical Institute
Academy of Sciences of Czech Republic

11.6.96

**Abstract**

We prove an exponential lower bound on the length of cutting plane proofs. The proof uses an extension of a lower bound for monotone circuits to circuits which compute with real numbers and use nondecreasing functions as gates. The latter result is of independent interest, since, in particular, it implies an exponential lower bound for some arithmetic circuits.

## 1 Introduction

The problem of proving lower bounds on the length of propositional proofs is generally recognized as one of the most important problems in logic and complexity theory. Its importance stems from its relation to independence results in bounded arithmetic and to the problem $\mathcal{NP} \stackrel{?}{=} co\mathcal{NP}$, see [17]. There is no general technique available, which would work for all propositional proof systems (this is needed to prove that $\mathcal{NP} \neq co\mathcal{NP}$). Therefore we are trying to extend our current methods to stronger and stronger systems.

The first important step was Haken's exponential lower bound on the length of resolution proofs [13]. Using a different technique (so called *random restrictions*) exponential lower bounds were obtained for bounded depth Frege systems, which can be considered as extensions of the resolution system [2, 16, 3]. Another strengthening of resolution is *the cutting plane proof system.* It originated in integer linear programming [12]. As a propositional proof system, it was first considered in [9]. The basic idea is to prove using a few elementary rules that a system of linear equations with integer coefficients does not have a 0–1 solution.

*to appear in Journal of Symbolic Logic

†The author was partially supported by grant A119107 of the Academy of Sciences of the Czech Republic, Humboldt Stiftung at Dortmund University and by joint US-Czechoslovak Science and Technology Program grant No. 93025.

The system will be defined below. The main result of this paper is an exponential lower bound on the length of proofs in the cutting plane proof system.

The idea of the lower bound proof goes back to Krajíček's paper [16]. There he proposed to use interpolation theorem to reduce lower bounds on the lengths of propositional proofs to lower bounds on circuits. Such an idea was implicitly used by Bonet, Pitassi and Raz in [4]. Razborov used interpolation in the context of bounded arithmetic [23]. In [18] Krajíček stated explicitly and proved interpolation theorems for certain proof systems using which it is possible to reduce the problem of proving lower bounds on the length of propositional proofs to lower bounds on the circuit size of explicit boolean functions. In particular he proved that for the resolution system and cutting plane system with a restriction on the size of coefficients (the restriction is that the absolute value of the coefficients in the inequalities used in the proof is bounded by a polynomial in the size of the proof). For such cutting plane proofs this generalizes the earlier result of Bonet, Pitassi and Raz [4].

In both cases, resolution and cutting planes, the reduction is to lower bounds on monotone boolean circuits. In this paper we give different proofs of the interpolation theorems of Krajíček. Our proofs provide more information about the connection between the proofs and the interpolating circuits. In this way we extend the theorem about cutting plane proofs to the general case where, in contrast to the previous results, there is no restriction on the coefficients.

Our reduction for cutting plane proofs does not lead to a class of circuits which has been considered before. We need circuits that compute with arbitrary integers and use certain monotone operations. It turns out that the original proof of Razborov [22] can be easily extended not only to such circuits, but in fact to *monotone circuits over real numbers*, by which we mean circuits which use real numbers and arbitrary nondecreasing real functions of bounded fan-in as gates. In particular such circuits may use the standard arithmetic operations $+, \times, \exp$, monotone threshold functions etc. Exponential lower bounds on the size of monotone arithmetic circuits, i.e., circuits with only $+$ and $\times$, had been proved before [24], but they seem to depend heavily on the restricted set of operations.

Independently of us S.A. Cook has observed that another proof of an exponential lower bound on monotone boolean circuits can be extended to monotone real circuits which gives, using our reduction, an exponential lower bound on the lengths of cutting planes proofs for another sequence of tautologies. The proof on monotone circuits, due to A. Haken [14], uses a different boolean function (*Broken Mosquito Screen*) and is simpler than the previous proofs. Let us note that neither of the two exponential lower bounds for monotone real circuits implies the other one.

# 2 Interpolation

Our discussion of this subject will be rather brief as recently this subject has been dealt with in length in [18].

The propositional version of a classical theorem of Craig [10, 11] states that for a given valid implication $\Phi(\bar{p},\bar{q}) \rightarrow \Psi(\bar{p},\bar{r})$, where $\bar{p},\bar{q},\bar{r}$ are disjoint sets of propositional variables, there exists a formula $I(\bar{p})$, which contains only the common variables $\bar{p}$, such that both $\Phi(\bar{p},\bar{q}) \rightarrow I(\bar{p})$ and $I(\bar{p}) \rightarrow \Psi(\bar{p},\bar{r})$ are valid. In this form it is a trivial fact, but Craig proved more. He gave a way of constructing the *interpolant* $I(\bar{p})$ from a proof of $\Phi(\bar{p},\bar{q}) \rightarrow \Psi(\bar{p},\bar{r})$. This implies that the complexity of the interpolant is bounded in some way by the complexity of the proof. This suggest the following way of proving lower bounds: suppose we can show that $\Phi(\bar{p},\bar{q}) \rightarrow \Psi(\bar{p},\bar{r})$ does not have a simple interpolant, then it cannot have a simple proof.

The original proof uses cut-elimination, hence the resulting interpolant is exponentially large in general. Then we can hardly deduce any limitations on the complexity of the interpolant. In fact, in order to be able to get bounds on the complexity of interpolants one has to consider $I(\bar{p})$ as a *boolean circuit.* More generally, we can think of an interpolant as a procedure which for a given input $\bar{a}$ decides whether $\Phi(\bar{a},\bar{q})$ is false or $\Psi(\bar{a},\bar{r})$ is true (one of these two possibilities is always true).

Now it is more natural to consider a valid *disjunction* $A(\bar{p},\bar{q}) \vee B(\bar{p},\bar{r})$ and look for a procedure which decides which of the two $A(\bar{a},\bar{q}), B(\bar{a},\bar{r})$ is true for a given input $\bar{a}$. It is possible that both disjuncts are true in which case the procedure can choose any of them. Equivalently we can search for an unsatisfiable disjunct. We shall call such a procedure an interpolant of $A(\bar{p},\bar{q}) \vee B(\bar{p},\bar{r})$. This is justified by the fact that a polynomial time procedure can always be converted to a polynomial size circuit $I(\bar{p})$.

Krajíček [18] has proved that for $A(\bar{p},\bar{q}), B(\bar{p},\bar{r})$ sets of clauses one can construct an interpolant $I(\bar{p})$ from a resolution refutation of $A(\bar{p},\bar{q}) \vee B(\bar{p},\bar{r})$ such that $I(\bar{p})$ is a circuit whose size is polynomial in the length of the proof. He actually gets a polynomial time decision procedure which can be equivalently stated as having a polynomial time algorithm for constructing the interpolating circuits. Furthermore he proves a similar theorem for cutting plane proofs, however the function which bounds the size of the interpolant circuit depends also on the maximal absolute value of the coefficients used in the proof.

The most striking fact is that, using a very mild assumption about the sets of clauses $A(\bar{p},\bar{q}), B(\bar{p},\bar{r})$, one can strengthen the conclusion so that the interpolating circuits are *monotone boolean circuits* i.e., circuits in basis $\{0,1,\wedge,\vee\}$. The assumption is that in one of the sets $A(\bar{p},\bar{q})$ or $B(\bar{p},\bar{r})$ all $\bar{p}$ occur positively or only negatively. This is very important, since there are exponential lower bounds on the size of monotone boolean functions computing explicitly given monotone boolean function e.g. [22], while in the nonmonotone case the best bounds are only linear. In this way the lower bound problem for resolution can be reduced to the known lower bounds in boolean complexity.

To be more precise we should note that a lower bound for monotone circuits computing an explicitly given boolean function does not automatically give a lower bound for a propositional proof system for which such an interpolation theorem is provable. Consider the sets of 0–1 inputs

$$\bar{A} =_{def} \{\bar{a}; \exists \bar{b}\ \neg A(\bar{a},\bar{b})\},$$

$$\bar{B} =_{def} \{\bar{a}; \exists \bar{c}\ \neg B(\bar{a},\bar{c})\}.$$

If $A(\bar{p},\bar{q}) \vee B(\bar{p},\bar{r})$ is a tautology, then $\bar{A} \cap \bar{B} = \emptyset$. For showing that $A(\bar{p},\bar{q}) \vee B(\bar{p},\bar{r})$ does not have short proofs, we need to know that $\bar{A}$ cannot be separated from $\bar{B}$ by a small monotone circuit. Also not every pair of disjoint sets which cannot be separated by a small monotone circuit can be used for such a purpose. We need that $\bar{A}$ and $\bar{B}$ be defined existentially by small formulas as above. Fortunately the known lower bounds for monotone circuits actually give us precisely what we need. For instance in [22], $\bar{A}$ is the set of $m$-cliques on an $n$-element set and $\bar{B}$ is the set of complete $(m-1)$-colorable graphs on the same set, where $m$ is a suitable parameter. The tautology obtained from this particular pair has been used in [18, 4], and we shall use it also for our lower bound.

There are several parallels of propositional proof complexity theory and boolean complexity theory. In the previous known relations there was always some apparent similarity, but the relation of resolution proofs to monotone circuits is a surprise. Now you can easily include a lower bound for resolution proofs in your course on circuit complexity.

Krajíček [18] gives two proofs of the interpolation theorem for resolution. We shall give yet another proof. The reason is that our proof gives more insight in the matter, in particular, the same idea enables us to prove our main result which is an exponential lower bound for unrestricted cutting plane proofs.

# 3 Resolution

Propositional resolution proof system is one of the weakest proof systems, however it is very important in artificial intelligence, as it constitutes an essential part of many automated theorem provers. We shall define it as the proof system which uses elementary disjunctions i.e., disjunctions of literals, as formulas, and the cut rule as the only rule

$$\frac{\Gamma \vee p,\ \ \Delta \vee \neg p}{\Gamma \vee \Delta},$$

(where $\Gamma, \Delta$ are elementary disjunctions). Recall that *a literal* is a variable or a negated variable and the elementary disjunctions are called *clauses.* We shall, of course, assume that two clauses are equal, if they have the same sets of literals. Resolution proof system is a refutation system which means that we are refuting initial clauses by deriving a contradiction represented as the empty clause. In resolution we can thus prove only formulas which are DNF's. This is not such a severe restriction as it may look like on the first glance. A lot of interesting statements have a natural representation as DNF's, others can be easily coded by using additional propositional variables. (For instance the proof of Cook's theorem is based on encoding Turing machine computations by a sets of clauses.)

The basic idea of our approach to interpolation theorems is very simple; we shall explain it on the resolution system. Suppose that we have a proof $P$ of the empty clause from clauses $A_i(\bar{p},\bar{q}), i \in I$, $B_j(\bar{p},\bar{r}), j \in J$ where $\bar{p},\bar{q},\bar{r}$ are disjoint sets of propositional variables. An interpolant is a circuit $C(\bar{p})$ such that for an assignment $\bar{a}$ of 0's and 1's for $\bar{p}$, if $C$ outputs 0, then $A_i(\bar{a},\bar{q}), i \in I$ are unsatisfiable, and if $C$ outputs 1, then $B_j(\bar{a},\bar{r}), j \in J$ are unsatisfiable. Clearly, instead of constructing $C$, it suffices to describe a polynomial time algorithm which on $\bar{a}$ produces 0 if $A_i(\bar{a}, \bar{q}), i \in I$ are unsatisfiable, and 1 if $B_j(\bar{a}, \bar{r}), j \in J$ are unsatisfiable. The trick is to try to get more: a refutation proof from $A_i(\bar{a}, \bar{q}), i \in I$ or a refutation proof from $B_j(\bar{a}, \bar{r}), j \in J$. Actually, this can be done quite easily. Observe that variables $\bar{q}$ can be mixed with variables $\bar{r}$ in $P$ only using a resolution along a variable $p_k$. Since we have substituted constants for variables $\bar{p}$, the resolutions along $p_k$'s can be omitted and thus a clause containing both kinds of variables will never appear. Thus $P$ splits into components and we take the component which contains the final empty clause of the original proof. This is the proof that we wanted.

A closer analysis, which we shall present below, shows that in fact we can use the original proof as a "skeleton" for the circuit (provided we choose a suitable basis of connectives for the circuit). Thus we get a very close relation between the proof and the interpolating circuit.

The ternary connective *sel* (selector) is defined by $sel(0, x, y) = x$ and $sel(1, x, y) = y$.

**Theorem 1** *Let $P$ be a resolution proof of the empty clause from clauses $A_i(\bar{p}, \bar{q}), i \in I$, $B_j(\bar{p}, \bar{r}), j \in J$ where $\bar{p}, \bar{q}, \bar{r}$ are disjoint sets of propositional variables. Then there exists a circuit $C(\bar{p})$ such that for every 0-1 assignment $\bar{a}$ for $\bar{p}$*

$$C(\bar{a}) = 0 \quad \Rightarrow \quad A_i(\bar{a}, \bar{q}), i \in I \text{ are unsatisfiable, and}$$

$$C(\bar{a}) = 1 \quad \Rightarrow \quad B_j(\bar{a}, \bar{r}), j \in J \text{ are unsatisfiable;}$$

*the circuit $C$ is in basis $\{0, 1, \wedge, \vee, sel\}$ and its underlying graph is the graph of the proof $P$.*

*Moreover, we can construct in polynomial time a resolution proof of the empty clause from clauses $A_i(\bar{a}, \bar{q}), i \in I$ if $C(\bar{a}) = 0$, respectively $B_j(\bar{a}, \bar{r}), j \in J$ if $C(\bar{a}) = 1$; the length of this proof is less than or equal to the length of $P$.*

**Theorem 2** *Suppose moreover that either all variables $\bar{p}$ occur in $A_i(\bar{p}, \bar{q}), i \in I$ only positively or all variables $\bar{p}$ occur in $B_i(\bar{p}, \bar{r}), j \in J$ only negatively, then one can replace the selector connective sel by a monotone ternary connective.*

As consequences of the relation between the proof and the circuit we get that the size of the circuit is at most the size of the proof and that the circuit is a formula, if the proof is in a tree form (the last statement had also been proved by Krajíček [18]). Note also that the monotone circuit from Theorem 2 can be transformed into a usual monotone circuit, i.e. circuit in basis $\{0, 1, \wedge, \vee\}$, by increasing the size of the circuit only linearly.

*Proof of Theorem 1.* First we describe the transformation of the proof for a given assignment $\bar{p} \mapsto \bar{a}$. Once we get an assignment for $\bar{p}$, we can eliminate all these variables from the proof, but we shall do it in two stages, as we shall need more information about this process when constructing the circuit explicitly.

Let us call a clause *q-clause,* resp. *r-clause,* if it contains only variables $\bar{p}, \bar{q}$ resp. $\bar{p}, \bar{r}$. We shall call a clause q-clause, resp. r-clause also in the case that the clause contains only variables $\bar{p}$ or is empty, but its ancestors are q-clauses, resp. r-clauses.

1. In the first stage we replace each clause of $P$ by a subclause so that each clause in the proof is either q-clause or r-clause. We start with the initial clauses, which are left unchanged and continue along the derivation $P$.

*Case 1.*

$$\frac{\Gamma \vee p_k,\ \ \Delta \vee \neg p_k}{\Gamma \vee \Delta}$$

and we have replaced $\Gamma \vee p_k$ by $\Gamma'$ and $\Delta \vee \neg p_k$ by $\Delta'$. Then we replace $\Gamma \vee \Delta$ by $\Gamma'$ if $p_k \mapsto 0$ and by $\Delta'$ if $p_k \mapsto 1$.

*Case 2.*

$$\frac{\Gamma \vee q_k,\ \ \Delta \vee \neg q_k}{\Gamma \vee \Delta}$$

and we have replaced $\Gamma \vee q_k$ by $\Gamma'$ and $\Delta \vee \neg q_k$ by $\Delta'$. If one of $\Gamma'$, $\Delta'$ is an r-clause, then it does not contain $q_k$, and we replace $\Gamma \vee \Delta$ by this clause. If both $\Gamma'$ and $\Delta'$ are q-clauses, we resolve along $q_k$, or take one, which does not contain $q_k$.

*Case 3.*

$$\frac{\Gamma \vee r_k,\ \ \Delta \vee \neg r_k}{\Gamma \vee \Delta}.$$

This is the dual case to case 2.

2. Now delete the clauses which contain a $\bar{p}$ literal with value 1, and remove all $\bar{p}$ literals from the remaining clauses. Thus we get a valid derivation of the final empty clause from the reduced initial clauses. If this final clause is a q-clause, the proof contains a subproof using only the reduced clauses $A_i, i \in I$; if it is an r-clause, the proof contains a subproof using only the reduced clauses $B_j, j \in J$.

3. The circuit $C$ is constructed so that the value computed at a gate corresponding to a clause $\Gamma$ will determine if it is transformed into a q-clause or an r-clause. We assign 0 to q-clauses and 1 to r-clauses.

Thus the circuit is constructed as follows. Put constant 0 gates on clauses $A_i, i \in I$ and constant 1 gates on clauses $B_j, j \in J$. Now consider three cases as above.

*Case 1.* If the gate on $\Gamma \vee p_k$ gets value $x$ and the gate on $\Delta \vee \neg p_k$ gets value $y$, then the gate on $\Gamma \vee \Delta$ should get the value $z = sel(p_k, x, y)$. Thus we place the $sel$ gate on $\Gamma \vee \Delta$.

*Case 2.* If the gate on $\Gamma \vee q_k$ gets value $x$ and the gate on $\Delta \vee \neg q_k$ gets value $y$, then the gate on $\Gamma \vee \Delta$ should get the value $z = x \vee y$. Thus we place the $\vee$ gate on $\Gamma \vee \Delta$.

*Case 3.* This is dual to case 2, so we place the $\wedge$ gate on $\Gamma \vee \Delta$. □

*Proof of Theorem 2.* W.l.o.g. assume that all $\bar{p}$'s are positive in clauses $A_i, i \in I$. This property is then inherited to the transformed q-clauses. Hence in case 1, if $\Delta'$ is a q-clause, it cannot contain $\neg p_k$, hence we can take it for $\Gamma \vee \Delta$, even if $p_k \mapsto 0$. Thus we can replace $sel(p_k, x, y)$ by $(p_k \vee x) \wedge y$ which is monotone and differs from the selector exactly on one input ($p_k = 0, x = 1, y = 0$) which corresponds to the above situation. □

# 4 Cutting planes

Using a result of Gomory [12] Chvátal [6, 7] proposed a method, called *cutting plane proofs,* for solving integer programing problems. He conjectured that for integer programing versions of $\mathcal{NP}$-complete problems there are no proofs of polynomial length. We shall consider only the propositional version of this system, where admissible solutions are only zeros and ones.

In cutting plane proofs we use the usual propositional variables $\bar{p}$ with the interpretation $0 =$ *false* and $1 =$ *true.* A proof line is an inequality

$$\sum_k c_k p_k \geq C,$$

where $c_k$ and $C$ are integers.

The **axioms** are $p_k \geq 0$ and $-p_k \geq -1$ (i.e. $0 \leq p_k \leq 1$) for every propositional variable $p_k$.

The rules are

1. **addition:** from $\sum_k c_k p_k \geq C$ and $\sum_k d_k p_k \geq D$ derive $\sum_k (c_k + d_k) p_k \geq C + D$;

2. **division:** from $\sum_k c_k p_k \geq C$ derive $\sum_k \frac{c_k}{d} p_k \geq \left\lceil \frac{C}{d} \right\rceil$, provided that $d > 0$ is an integer which divides each $c_k$;

3. **multiplication:** from $\sum_k c_k p_k \geq C$ derive $\sum_k d c_k p_k \geq dC$, where $d$ is an arbitrary positive integer.

Cutting plane system is also a refutation system: we want to refute a set of inequalities by deriving a contradiction, represented as $0 \geq 1$. The completeness of this system was proved by Gomory [12] (see also Schrijver [25] for an extension). For an exposition of cutting plane algorithms see e.g. [20, 8].

The expressive power of inequalities is at least as big as of clauses: the validity of a clause $\bigvee_{k \in I} p_k \vee \bigvee_{k \in J} \neg p_k$ is equivalent to $\sum_{k \in I} p_k + \sum_{k \in J} (1 - p_k) \geq 1$, (which can be rewritten into the form used in cutting plane proofs). Also the cut rule of the resolution system can be easily simulated by a constant number of applications of the above rules. For further results on cutting plane proofs see the papers listed as references.

The idea of the proof of the interpolation theorem for cutting planes proofs is the same: given an assignment for the common variables $\bar{p}$, we want to split the proof, so that we get a refutation either from q-clauses or from r-clauses. Now the rule which can mix variables $\bar{q}$ with variables $\bar{r}$ is the addition of two inequalities. Thus our strategy is not to perform this rule in such a case and keep two inequalities. In the original proof this would spoil the divisibility condition for the division rule, but as we replace variables $\bar{p}$ by an assignment, we can treat them as a part of the constant term and we do not need the divisibility condition for them.

Now we shall do it in more details.

**Theorem 3** *Let $P$ be a cutting plane proof of the contradiction $0 \geq 1$ from inequalities*

$$\sum_k c_{i,k} p_k + \sum_l b_{i,l} q_l \geq A_i, i \in I,$$

$$\sum_k c'_{j,k} p_k + \sum_m d_{j,m} r_m \geq B_j, j \in J,$$

*where $\bar{p}, \bar{q}, \bar{r}$ are disjoint sets of propositional variables. Then there exists a circuit $C(\bar{p})$ such that for every 0-1 assignment $\bar{a}$ for $\bar{p}$*

$C(\bar{a}) = 0 \quad \Rightarrow \quad \sum_k c_{i,k} a_k + \sum_l b_{i,l} q_l \geq A_i, i \in I$ *are unsatisfiable, and*

$C(\bar{a}) = 1 \quad \Rightarrow \quad \sum_k c'_{j,k} a_k + \sum_m d_{j,m} r_m \geq B_j, j \in J$ *are unsatisfiable.*

*The size of the circuit is polynomial in the binary length of the numbers $A_i, i \in I, B_j, j \in J$ and the number of inequalities in $P$.*

*Moreover, we can construct in polynomial time a cutting plane proof of the contradiction $0 \geq 1$ from inequalities $\sum_k c_{i,k} a_k + \sum_l b_{i,l} q_l \geq A_i, i \in I$ if $C(\bar{a}) = 0$, respectively $\sum_k c'_{j,k} a_k + \sum_m d_{j,m} r_m \geq B_j, j \in J$ if $C(\bar{a}) = 1$; the length of this proof is less than or equal to the length of $P$.*

*Proof.* Let $P$ and an assignment $\bar{p} \mapsto \bar{a}$ be given. We shall gradually replace each inequality in $P$

$$\sum_k e_k p_k + \sum_l f_l q_l + \sum_m g_m r_m \geq D$$

by a pair of inequalities

$$\sum_l f_l q_l \geq D_0, \quad \sum_m g_m r_m \geq D_1,$$

where $D_0, D_1$ are some integers. The sums may be empty, in which case we treat them as 0. We shall ensure that the pair of inequalities is at least as strong as the original inequality for the assignment $\bar{a}$, which means that

$$D_0 + D_1 \geq D - \sum_k e_k a_k.$$

An initial inequality $\sum_k c_{i,k} p_k + \sum_l b_{i,l} q_l \geq A_i$ will be replaced by the pair $\sum_l b_{i,l} q_l \geq A_i - \sum_k c_{i,k} a_k$ and $0 \geq 0$. Dually, an initial inequality $\sum_k c'_{j,k} p_k + \sum_m d_{j,m} r_m \geq B_j$ will be replaced by $0 \geq 0$ and $\sum_m d_{j,m} r_m \geq B_j - \sum_k c'_{j,k} a_k$.

The addition rule will be simulated by performing additions of the first inequalities from pairs and the second inequalities in the pairs in parallel. This clearly preserves the properties that we need. The multiplication rule is simulated in a similar way.

The division rule is also performed in parallel on the two inequalities in the pair. The divisibility condition is clearly satisfied, as we have the same coefficients at variables $\bar{q}$ and $\bar{r}$ as in the original proof. Let us check that the pair of inequalities is at least as strong as the original inequality after we perform a division rule. This means that for a positive integer $h$ which divides all $e_k$'s, we need

$$D_0 + D_1 \geq D - \sum_k e_k a_k \quad \Rightarrow \quad \left\lceil \frac{D_0}{h} \right\rceil + \left\lceil \frac{D_1}{h} \right\rceil \geq \left\lceil \frac{D}{h} \right\rceil - \sum_k \frac{e_k}{h} a_k,$$

which is obvious (first take the ceiling function on the l.h.s., then on the r.h.s.).

Consider the pair corresponding to the final inequality $0 \geq 1$. It is of the form $0 \geq D_0,\ 0 \geq D_1$ where $D_0 + D_1 \geq 1$. Hence either $D_0 \geq 1$ or $D_1 \geq 1$, thus we have a proof of contradiction either from $\sum_k c_{i,k} a_k + \sum_l b_{i,l} q_l \geq A_i, i \in I$ or from $\sum_k c'_{j,k} a_k + \sum_m d_{j,m} r_m \geq B_j, j \in J$.

In order to see that the above procedure can be done in polynomial time in the binary length of $A_i, i \in I, B_j, j \in J$ and the number of inequalities, we need a polynomial upper bound on the binary lengths of $D_i$'s. In general we cannot bound the coefficients in the proof in such a way, but Clote and Buss [5] proved that each proof $P$ can be transformed into another proof $P'$ such that $P'$ is at most polynomially longer than $P$ and all the coefficients in $P'$ have polynomially bounded binary length.

Finally use the well-known transformation of polynomial time algorithms into sequences of polynomial size circuits. □

To state a monotone version of the above theorem, we define a new class of circuits.

**Definition 1** *A* monotone real circuit *is a circuit which computes with real numbers and uses arbitrary nondecreasing real unary and binary functions as gates.*

*We say that a monotone real circuit computes a boolean function (uniquely determined by the circuit), if for all inputs of 0's and 1's the circuit outputs 0 or 1.*

Clearly such circuits can compute only monotone boolean functions. This class includes the class of *monotone boolean circuits*, since we can think of $\wedge$ resp. $\vee$ as min resp. max. (In fact, $\wedge$ resp. $\vee$ can be computed using only $+$ and division with rounding up, which are the operations that we will need in the simulation of cutting planes proofs.) Moreover monotone real circuit can use $+, \times$ etc. This is a much bigger class of circuits than is needed for the next theorem, but the lower bound method works equally well for such a class, thus we do not have to restrict it.

**Theorem 4** *Let $P$ be a cutting plane proof of the contradiction $0 \geq 1$ from inequalities*

$$\sum_k c_{i,k} p_k + \sum_l b_{i,l} q_l \geq A_i, i \in I,$$

$$\sum_k c'_{j,k} p_k + \sum_m d_{j,m} r_m \geq B_j, j \in J.$$

*Suppose that all the coefficients $c_{i,k}$ are nonnegative, or all the coefficients $c'_{i,k}$ are nonpositive, then one can construct a real monotone interpolating circuit $C$ (computing a boolean function) in the sense of Theorem 3 whose size is bounded by a linear function in the number of variables and the number of inequalities of the proof $P$.* □

*Proof.* Let us first realize that we only need to compute the constant $D_0$ (or only $D_1$) corresponding to the final inequality. We shall assume w.l.o.g. that coefficients $c_{i,k}$ are nonnegative. Then it will be more convenient to talk about $-D_0$. To this end we only need to compute successively $-D_0$ from each pair. As in the case of resolution proofs, we can use the graph of the proof for constructing a circuit. The gates will produce the new $-D_0$ from the previous ones. The circuit has 0 and 1 inputs, but computes arbitrary integers in the inner gates. Let us see what gates do we need. In general we need:

1. addition of an integer constant,
2. multiplication by an integer constant,
3. addition,
4. division by a positive integer constant with rounding down,
5. to get a 0-1 output we add a threshold gate as the output gate (the unary gate $t$ defined by $t(x) = 1$ if $x \geq 1$ and $t(x) = 0$ otherwise).

All the operations are nondecreasing except for multiplications by negative constants. In general they are needed in the initial inequalities, where for $\sum_k c_{i,k}p_k + \sum_l b_{i,l}q_l \geq A_i$ we need to compute $\sum_k c_{i,k}a_k - A_i$. As we assume that coefficients $c_{i,k}$ are nonnegative, we do not need multiplications by negative constants there. The remaining inequalities where we need multiplication by a negative constant are $-p_k \geq -1$. These can be treated as inequalities containing $r_m$'s, i.e., we put $D_0 = 0$ for them. Thus we get all gates nondecreasing. □

We do not know, if the real monotone circuits resulting from cutting plane proofs can be transformed into a *boolean* monotone circuit in such a way that the size increases only polynomially. The only transformation that we know of depends also on the size of the numbers that occur in the proof (in the integer valued circuit). The simulation is based on representing the computed integers in *unary notation.* First replace each multiplication by a constant by several additions. Then addition and division with rounding up can be computed by polynomial size monotone boolean circuits (e.g. addition can be computed by boolean convolution). In this way one can strengthen a theorem of Krajíček [18]:

**Theorem 5** *Under the same assumptions as in Theorem 3 and assuming moreover that all the coefficients $c_{i,k}$ are nonnegative, or all the coefficients $c'_{i,k}$ are nonpositive, one can construct a monotone boolean interpolating circuit $C$ such that its size is bounded by a polynomial in the number of variables, the number of inequalities and the maximal absolute value of a coefficient which occurs in the proof $P$.* □

# 5 An exponential lower bound on the size of monotone circuits over reals

In this section we prove an exponential lower bound on monotone real circuits computing the clique function. Before we start the proof a few remarks about the computational model are in order.

As we use arbitrary monotone gates, the arithmetical structure of the real numbers $\mathbf{R}$ will not be used for the lower bound, we use $\mathbf{R}$ as a "universal" linearly (=totally) ordered domain. Furthermore, the restriction to fan-in two gates is not essential, we could consider any constant ($\geq 2$) fan-in (actually also some slowly growing). Let us also note that there are infinitely many already unary nondecreasing real functions, so we do *not* assume the basis to be finite. Thus our lower bound is very strong in the sense that it covers a lot of natural models of computation.

The idea of the lower bound on the size of monotone circuits over reals is also very simple. Take the original proof for monotone boolean circuits of Razborov [22] and replace everywhere monotone boolean functions by monotone real functions (to get a better bound, we shall use an improved version of the original proof as presented in [26]). Surprisingly, this needs only a few inessential changes in the proof. A reader familiar with Razborov's proof would need only a few hints to do it himself, thus, perhaps, the main contribution of this paper is in realizing the importance of such circuits. Nevertheless, to make the paper selfcontained we shall give the whole proof.

For those unfamiliar with Razborov's proof we shall explain the basic idea of his *approximation method*. Let us consider $n$-variate real functions instead of boolean functions and nondecreasing binary real functions as gates, as this is what we shall need in our proof (let us forget about unary functions and constants for a moment). Suppose that we want to show that a function $F_0$ cannot be computed by a small circuit. We take a subclass of all $n$-variate functions, say $\mathcal{F}$ and define a *distance* between $n$-variate functions. If the distance between $F$ and $\tilde{F}$ is small, we say that $\tilde{F}$ approximates $F$ well. Now we need three properties

1. the projection functions $x_i$ are in $\mathcal{F}$,
2. if $F_1, F_2 \in \mathcal{F}$ and $g$ is a nondecreasing binary real function, then there is a very good approximation $\tilde{F}$ of $g(F_1, F_2)$ which is in $\mathcal{F}$,
3. every $\tilde{F} \in \mathcal{F}$ is a very bad approximation of $F_0$.

Then one argues that $F_0$ cannot be computed by a small circuit, since a big error cannot accumulate in a small circuit.

This is an oversimplification of what is really going on. Also the word "approximation" is misleading. It can happen that a good approximator of an $F$ never produces the same value as $F$ (nor a value which is close to it in the sense of the topology of the reals).[1] This is due to a special way of measuring the distance. There are two subsets of inputs, say $B$ and $C$, all inputs from $B$ are rejected (the function gives 0 on them), all inputs from $C$ are accepted (the function gives 1 on them). Let an $\tilde{F}$ be an intended approximator for an $F$. An input $a \in B$ is bad, if $\tilde{F}(a) > F(a)$, and an input $a \in C$ is bad, if $\tilde{F}(a) < F(a)$; we do not care for the remaining inputs. The distance is defined to be the number of bad inputs. A moment reflection suffices to realize that it is the right definition. If we want to get a function $F_0$ which gives 0 on $B$ and 1 on $C$, it is enough to get a function which is $\leq 0$ on $B$ and $\geq 1$ on $C$, since applying one more monotone gate (the threshold gate) we get such an $F_0$. Also if $\tilde{F}(a) < F(a)$ for an $a \in B$ or if $\tilde{F}(a) > F(a)$ for an $a \in C$, then on this input $\tilde{F}$ is "even better" than $F$.

Here is our lower bound.

**Theorem 6** *Suppose that the inputs for a monotone real circuit $C$ are 0-1 vectors of length $\binom{n}{2}$ encoding in the natural way graphs on an $n$-element set. Suppose that $C$ outputs 1 on all cliques of size $m$ and outputs 0 on all complete $m-1$-partite graphs where $m = \lfloor \frac{1}{8}(n/\log n)^{2/3} \rfloor$. Then the size of the circuit is at least*

$$2^{\Omega((n/\log n)^{1/3})}.$$

*Proof.* We shall follow the exposition of an improved version of Razborov's original proof for the clique function, due to Alon and Boppana [1], as presented in [26]. We shall modify the concepts so that we can talk about real valued functions instead of boolean functions and set systems. We shall talk about *monotone* real functions, meaning real functions which are *nondecreasing* w.r.t. the set inclusion ordering or natural ordering of boolean vectors or the ordering of the real numbers $\mathbf{R}$, depending on the domain on which they are defined.

Let $n$ be the size of the vertex set, let $V = \{1, \ldots, n\}$. We denote by $[X]^k$, resp $[X]^{\leq k}$ the set of subsets of $X$ of size $k$, resp. $\leq k$. Let, furthermore, integer parameters $l, r$ and $m$ be fixed so that $l \leq m$, $4rm \leq n$. We shall use the same relation $\vdash$ as in the boolean case

$$W_1, \ldots, W_r \vdash W \quad \equiv_{def} \quad \forall i \neq j \; W_i \cap W_j \subseteq W.$$

The basic concept is the following:

$$\mathcal{F}_{l,r} =_{def} \{f : [V]^{\leq l} \to \mathbf{R} \; ; \; f \text{ monotone } \wedge \; \forall \; W_1, \ldots, W_r \vdash W \quad f(W) \geq \min_i f(W_i)\}.$$

The functions in $\mathcal{F}_{l,r}$ will be called *r-closed.* For a general function $f : [V]^{\leq l} \to \mathbf{R}$ we shall denote by $f^*$ its *closure* – the minimal function $f^* \in \mathcal{F}_{l,r}$ such that $f \leq f^*$.

Our circuits will compute with inputs from $\{0,1\}^{[V]^2}$, which will be identified in the natural way with graphs on $V$. For a monotone $f : [V]^{\leq l} \to \mathbf{R}$ we denote by $\langle f \rangle$ the minimal monotone function $\langle f \rangle : \{0,1\}^{[V]^2} \to \mathbf{R}$ such that $f(W) \leq \langle f \rangle([W]^2)$ for all $W \in [V]^{\leq l}$, i.e.,

$$\langle f \rangle(G) =_{def} \max\{f(W) \; ; \; W \in [V]^{\leq l}, \; W \text{ a clique in } G\}.$$

[1]This is special for the monotone lower bounds and it is not true in some other applications of the approximation method.

Let us stress that we use graphs without loops, hence they do not contain one-element cliques. Thus the value $\langle f\rangle(G)$, for a nonempty graph $G$, is equal to $f(W)$ for some at least two-element clique of $G$. If $G$ is the empty graph, we define $\langle f\rangle(G) =_{def} f(\emptyset)$.[2]

Given two functions $F, \tilde{F} : \{0,1\}^{[V]^2} \to \mathbf{R}$ (think of $\tilde{F}$ as an approximation of $F$) we define the *sets of errors* by

$$\delta^+(\tilde{F}, F) =_{def} \{Z\ ;\ Z \text{ an } m \text{ clique and } \tilde{F}(Z) < F(Z)\},$$

$$\delta^-(\tilde{F}, F) =_{def} \{B\ ;\ \tilde{F}(B) > F(B), B \text{ complete } m-1\text{-partite graph }\}.$$

For a given real monotone circuit we define inductively approximations of the functions computed at the gates as follows. We assume that for each approximating functions $\tilde{F}$ there is an $f \in \mathcal{F}_{l,r}$ such that $\tilde{F} = \langle f\rangle$.

- If $\tilde{F}$ is an approximation of $F$ and a unary operation $\circ$ is applied to it, then we take $\circ(\tilde{F})$ as the approximation of $\circ(F)$.
- If $\tilde{F}_1, \tilde{F}_2$ are approximations of $F_1, F_2$ and a binary operation $\circ$ is applied to them, then we take the functions $f_1, f_2 \in \mathcal{F}_{l,r}$ such that $\tilde{F}_i = \langle f_i\rangle$ and take $\langle (f_1 \circ f_2)^*\rangle$ as the approximation of $F_1 \circ F_2$.

Note that the definition is the same for operation of any number of arguments, only accidentally for unary operations we do not have to take the closure. Thus unary operations do not introduce errors.

Let $f : [V]^{\leq l} \to \mathbf{R}$ be monotone. We shall say that $W \in [V]^{\leq l}$ is *$f$-minimal*, if for every $U \subset W$ (strict inclusion), $f(U) < f(W)$. Let $G$ be a nonempty graph. Then we know that $\langle f\rangle(G) = f(W)$ for some at least two-element clique of $G$. Furthermore we can take an $f$-minimal $U$ such that $U \subseteq W$ and $f(U) = f(W) = \langle f\rangle(G)$; this $U$ may have size 0 or 1.

**Lemma 1** *Let $f \in \mathcal{F}_{l,r}$, let $k \leq l$ be given. Then the number of $f$-minimal sets $W \in [V]^{\leq k}$ is at most $(k+1)(r-1)^k$.*

*Proof.* By induction on $r = 2, 3, \ldots$.

1. Let $r = 2$. Suppose that there are $> k+1$ minimal sets of size at most $k$. Then two, say $W_1, W_2$, must be incomparable. Clearly $W_1 \cap W_2$ is properly contained in $W_1$ and $W_2$ and $W_1, W_2 \vdash W_1 \cap W_2$. Hence $f(W_1 \cap W_2) = \min\{f(W_1), f(W_2)\}$, but this is a contradiction with the minimality of $W_1$ and $W_2$.

2. Let $r > 2$. Let $D$ be an $f$-minimal set with the maximal value $f(D)$. We shall apply the inductive assumption to the functions $f_C$ defined on $[V \setminus D]^{\leq (l-|C|)}$ by

$$f_C(W) = f(W \cup C),$$

for $C \subseteq D$.

[2]These special cases are not treated correctly in [26].

First we check that every $f_C$ is $r-1$ closed on the restricted domain. Clearly, if the $r-1$-closedness is violated by sets $W_1, \ldots, W_{r-1}, W$, i.e. $W_i \cap W_j \subseteq W$ for $i \neq j$ and $f_C(W) < \min_i f_C(W_i)$, then also the $r$-closedness of $f$ is violated by $D, W_1 \cup C, \ldots, W_{r-1} \cup C, W \cup C$, since $f(D)$ is the maximal value of $f$.

Hence we can estimate the number of $f$-minimal sets of size $\leq k$ by

$$\leq \sum_{i=0}^{k} \binom{k}{i} (k-i+1)(r-2)^{k-i} \leq (k+1) \sum_{i=0}^{k} \binom{k}{i} (r-2)^{k-i} = (k+1)(r-1)^k.$$

□

Next we shall estimate the number of elementary steps needed to produce the closure of a general function $f : [V]^{\leq l} \to \mathbf{R}$. We shall use the same (rough) estimate as in [26].

**Lemma 2** *The closure $f^*$ of an $f : [V]^{\leq l} \to \mathbf{R}$ can be obtained in at most $n^l$ steps where in each step we take some $W_1, \ldots, W_r \vdash W$ with $f(W) < \min_i f(W_i)$ and, for all $U \supseteq W$, we replace the values $f(U)$ by $\max\{f(U), \min_i f(W_i)\}$.*

*Proof.* Always choose $W_1, \ldots, W_r \vdash W$ with $f(W) < \min_i f(W_i)$ so that $\min_i f(W_i)$ is maximal. Then the values of $f(U)$, $U \supseteq W$, will not be increased afterwards. □

The next lemma need not be changed at all.

**Lemma 3** *Let $B$ be a random complete $m-1$ partite graph given by randomly coloring $V$ by $m-1$ colors. Then, for fixed $W_1, \ldots, W_r \vdash W$,*

$$Pr([W_1]^2 \not\subseteq B \wedge \ldots \wedge [W_r]^2 \not\subseteq B \wedge [W]^2 \subseteq B) \leq \left(1 - \frac{(m-1)\ldots(m-l)}{(m-1)^l}\right)^r \leq \left(\frac{l^2}{2(m-1)}\right)^r.$$

□

**Lemma 4** *Let $B$ be a random complete $m-1$ partite graph given by randomly coloring $V$ by $m-1$ colors, let $f : [V]^{\leq l} \to \mathbf{R}$ be monotone. Then*

$$Pr(\langle f^* \rangle(B) > \langle f \rangle(B)) \leq n^l \left(\frac{l^2}{2(m-1)}\right)^r.$$

*Proof.* In order to be forced to increase the value on $B$ we must have $W_1, \ldots, W_r \vdash W$ with $[W]^2 \subseteq B$ and $\langle f \rangle(B) < \min_i f(W_i)$, thus $[W_i]^2 \not\subseteq B$, for $i = 1, \ldots, r$. So we can apply Lemmas 2 and 3 (as in the original proof). □

This lemma gives us an estimate how much can the error $\delta^-$, counted using random colorings of $V$ by $m-1$ colors, increase on a monotone gate. Let us note, in passing, that for this error we do not need any bound on the fan-in of the gates. Next lemma estimates a possible increase of the error $\delta^+$. Again, the argument is isomorphic with the original one.

**Lemma 5** *For every monotone operation $\circ$ and every $f_1, f_2 \in \mathcal{F}_{l,r}$*

$$|\{Z\ ;\ Z \text{ an } m \text{ clique and } \langle (f_1 \circ f_2)^* \rangle (Z) < \langle f_1 \rangle (Z) \circ \langle f_2 \rangle (Z) \text{ for some monotone operation } \circ\ \}|$$

$$\leq 4(l+1)2^{-l}\binom{n}{m}.$$

*Proof.* We shall estimate the size of a possibly larger set which is defined in the same way, except that we omit the closure $^*$. Let $Z$ be in this set, $Z = [K]^2$, let $\circ$ be the corresponding monotone operation. By a remark above, there are $W_i \in [V]^{\leq l}$, $W_i \subseteq K$, such that $W_i$ is $f_i$-minimal and $\langle f_i \rangle (Z) = f_i(W_i)$ for $i = 1, 2$. For every $W \in [V]^{\leq l}$, $W \subseteq K$,

$$f_1(W) \circ f_2(W) \leq \langle (f_1 \circ f_2)^* \rangle (Z) < \langle f_1 \rangle (Z) \circ \langle f_2 \rangle (Z) = f_1(W_1) \circ f_2(W_2).$$

Since $f_1, f_2$ are monotone, $W_1 \cup W_2 \notin [V]^{\leq l}$, otherwise we could take it for $W$ and refute the inequality. Hence $|W_1 \cup W_2| \geq l$, thus at least one of the sets $W_1$ and $W_2$ has size $\geq l/2$. Consequently every such a $K$ contains either an $f_1$-minimal or an $f_2$-minimal set of size $\geq l/2$. Using Lemma 1 and $4rm \leq n$, we get the bound on the number of such $m$-cliques

$$\leq 2 \sum_{l/2 \leq k \leq l} (k+1)(r-1)^k \binom{n-k}{m-k} \leq 2(l+1)\binom{n}{m} \sum_{l/2 \leq k \leq l} (r-1)^k \frac{m \cdot \ldots \cdot (m-k+1)}{n \cdot \ldots \cdot (n-k+1)}$$

$$\leq 4(l+1)4^{-l/2}\binom{n}{m}.$$

□

It remains to estimate the minimal errors between a function $F$ that outputs 1 on all cliques of size $m$ and outputs 0 on all complete $m-1$ partite graphs and a function $\langle f \rangle$, for $f \in \mathcal{F}_{l,r}$.

**Lemma 6** *Let $f \in \mathcal{F}_{l,r}$ and suppose $\langle f \rangle (G) < 1$ for some nonempty $G$. Then $\langle f \rangle (Z) \geq 1$ for at most $\frac{7}{9}\binom{n}{m}$ $m$-cliques.*

*Proof.* The assumption $\langle f \rangle (G) < 1$, for at least one $G$, implies that $f(\emptyset) < 1$. So each $m$-clique $Z$ such that $\langle f \rangle (Z) \geq 1$ must contain some $f$-minimal set of size at least 1. The number of such cliques is thus estimated, using Lemma 1 and the bound $4rm \leq n$, by

$$\sum_{1 \leq k \leq l} (k+1)(r-1)^k \binom{n-k}{m-k} \leq \sum_{1 \leq k \leq l} (k+1)(r-1)^k \left(\frac{m}{n}\right)^k \binom{n}{m} \leq \binom{n}{m} \sum_{1 \leq k \leq l} \frac{k+1}{4^k}$$

$$\leq \binom{n}{m} \left( \sum_{k=1}^{\infty} \frac{1}{4^k} + \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} \frac{1}{4^k} \right) = \frac{7}{9}\binom{n}{m}.$$

□

Hence either

$$|\delta^+(F, \langle f \rangle)| \geq \frac{7}{9}\binom{n}{m},$$

or $\delta^-(F, \langle f \rangle)$ is the set of all complete $m-1$-partite graphs. Comparing these errors with the bounds on a possible increase of the errors on monotone gates, Lemmas 4 and 5, we get a lower bound on the number of gates in a monotone real circuit computing such an $F$:

$$\geq \min\left\{ \frac{\frac{7}{9}\binom{n}{m}}{4(l+1)2^{-l}\binom{n}{m}}, \frac{(m-1)^n}{n^l\left(\frac{l^2}{2(m-1)}\right)^r (m-1)^n} \right\}.$$

In order to get the bound $2^{\Omega((n/\log n)^{1/3})}$, we can take almost the same parameters as in [26], namely $m = \lfloor \frac{1}{8}(n/\log n)^{2/3} \rfloor$, $l = \lceil m^{1/2} \rceil$, $r = \lceil 4m^{1/2} \log n \rceil$, the slightly different bound in Lemma 1 does not influence the asymptotic bound. □

# 6 A lower bound on cutting plane proof systems

To get a lower bound on cutting plane proof systems we need a set of inequalities which satisfies Theorem 4 and for which the decision of which of the two subsets is not satisfiable is equivalent to the decision between a clique of size $m$ and a complete $m-1$-partite graph. We shall use essentially the same inequalities as in [18, 4]. To get it, we first write the statement *a graph either does not contain a clique of size* $m$*, or it is not* $m-1$*-colorable* as a propositional tautology. This can be done as a DNF, hence its negation is equivalent to a set of clauses. Finally, any clause can be easily expressed as an inequality. Now one has only to check that we get the right signs at the coefficients.

Let us write down the inequalities. Let $n$ and $m$ be given. We shall use variables $p_{i,j}$, $1 \leq i < j \leq n$ to code a graph on $n$ vertices; variables $q_{k,i}$ will code a one-to-one mapping from an $m$ element set into the vertices of the graph; variables $r_{i,l}$ will code an $m-1$-coloring of the graph. Hence the following inequalities express that a graph contains a clique of size $m$ and it is $m-1$-colorable.

$$\sum_i q_{k,i} \geq 1, \qquad \text{for } k = 1, \ldots, m; \tag{1}$$

$$\sum_k q_{k,i} \leq 1, \qquad \text{for } i = 1, \ldots, n; \tag{2}$$

$$p_{i,j} - q_{k,i} - q_{k',j} \geq -1 \qquad \text{for } 1 \leq i < j \leq n, 1 \leq k, k' \leq m, k \neq k'; \tag{3}$$

$$\sum_l r_{i,l} \geq 1, \qquad \text{for } i = 1, \ldots, n; \tag{4}$$

$$p_{i,j} + r_{i,l} + r_{j,l} \leq 2, \qquad \text{for } 1 \leq i < j \leq n, l = 1, \ldots, m-1. \tag{5}$$

The decision whether the first three are inconsistent or the last two are inconsistent is equivalent to the decision whether the graph coded by $p_{i,j}$'s does not contain an $m$-clique or it is not $m-1$-colorable.

Note that those $p_{i,j}$'s which occur with $q_{k,i}$'s have positive signs, and those $p_{i,j}$'s which occur with $r_{i,l}$'s have negative signs (if represented using only $\geq$). It is sufficient to have one of these properties, in order to get monotone circuits.

Hence by Theorems 4 and 6 we have:

**Corollary 7** *Any cutting plane proof of* $0 \geq 1$ *from the inequalities (1) – (5) has at least* $2^{\Omega((n/\log n)^{1/3})}$ *steps.*

# 7 Conclusions and open problems

Observe that the clique vs. coloring statement follows easily from the pigeonhole principle. This can be shown in bounded depth Frege systems where we can define the composition of the mapping from the $m$ element set into the graph and the coloring of the graph by $m-1$ colors, but not in resolution and cutting plane proof systems. Consider a modified clique vs. coloring statement where the clique has size $m$ and the number of colors is $m/2$. Using the same proof one can show that we still get an exponential lower bound on cutting planes proofs. On the other hand these tautologies have quasipolynomial (i.e. $2^{\log n^{O(1)}}$) proofs in a bounded depth Frege system, since the corresponding weak pigeonhole principle has such proofs. The latter is proved by translating a proof of Wilkie [21] from a fragment of arithmetic to propositional calculus (cf. [17] for an exposition). Thus our lower bound shows that bounded depth Frege systems are sometimes more powerful. The converse separation had been known before (namely, the (strong) pigeonhole principle has polynomial cutting plane proofs [9], while it does not have subexponential proofs in bounded depth Frege proofs [2, 3]).

For lower bounds on cutting plane proofs we only need two kinds of monotone gates, addition and division with rounding up. The lower bound on monotone real circuits gives us, however, much more. Can we use it to obtain exponential lower bounds for proofs in a proof system stronger than cutting planes, say, some extension of cutting plane proof system?

Let us note that there is a real polynomial which gives 1 on all cliques of size $m$ and 0 on all complete $m-1$ partite graphs (take the sum of monomials where each monomial is the product of variables corresponding to the edges of a clique). Thus such a function can be computed by arithmetic circuits with $+$ and $\times$. Our Theorem 6 implies that every arithmetic circuit with nondecreasing operations which computes such a function is exponentially large. Of course, this function does not occur often in algebraic computations, but, perhaps, some modifications of the proof may give good lower bounds for more "practical" algebraic functions.

The extension of the lower bound from monotone boolean circuits to monotone real circuits shows the power of Razborov's method on the one hand. On the other hand, it gives us another explanation why the method requires monotonicity. It is easy to construct a circuit with $n-1$ binary real monotone gates which gives different reals to different boolean inputs. Hence extending it with one *nonmonotone* unary gate we can compute any boolean function using $n$ gates. Thus the method fails even for circuits with a single nonmonotone gate.

The problem of proving exponential lower bounds for stronger proof systems, in particular for Frege systems and Extended Frege systems, remains open. There are strong indications that an approach based on interpolation theorems is not applicable for such systems, or must be essentially modified. Namely, if the encryption scheme $RSA$ is secure, then the effective interpolation theorem fails for Extended Frege systems, see [19]. It is likely that it fails also for Frege systems, but we do not know. In any case there are weaker versions that may be still true. In particular we do not know if the following is true:

**Problem.** *Suppose that $\varphi(\bar{p}) \vee \psi(\bar{q})$ has a Frege proof $P$, where $\varphi(\bar{p}), \psi(\bar{q})$ have disjoint sets of variables $\bar{p}, \bar{q}$. Does there exist a proof $P'$ either of $\varphi(\bar{p})$ or of $\psi(\bar{q})$ whose size is polynomial in the size of $P$ and $\varphi(\bar{p}) \vee \psi(\bar{q})$?*

# 8 Acknowledgments

I want to thank to S. Buss, V. Chvátal, J. Krajíček, A. Razborov and J. Sgall who read and commented an early draft of the paper and to S.A. Cook and Xudong Fu for pointing out an error there.

# References

[1] N. Alon, R.B. Boppana, *The monotone circuit complexity of Boolean functions,* Combinatorica 7/1, (1987), 1-22.

[2] M. Ajtai, *The complexity of the pigeonhole principle,* Proc. 29-th FOCS, 1988, 346-355.

[3] P. Beame, R. Impagliazzo, J. Krajíček, T. Pitassi, P. Pudlák, A. Woods, *Exponential lower bounds for the pigeonhole principle,* Proc. 24-th STOC, 1992, 200-221.

[4] Bonet, M., Pitassi, T. and Raz, R., *Lower bounds for Cutting Planes proofs with small coefficients,* Proc. 27-th STOC, 1995, 575-584.

[5] S.R. Buss and P. Clote, *Cutting planes, connectivity, and threshold logic,* Archive for Math. Logic, to appear

[6] V. Chvátal, *Edmonds polytopes and a hierarchy of combinatorial problems,* Discrete Math., 4 (1973), 305-337.

[7] V. Chvátal, *Some linear programing aspects of combinatorics,* in Proc. of the Conf. on Algebraic Aspects of Combinatorics, Toronto 1975, Utilitas Math. Publishing Inc., Winnipeg, 2-30.

[8] V. Chvátal, W. Cook, M. Martmann, *On cutting plane proofs in combinatorial optimization,* Linear Algebra and Its Applications 114/115 (1989), 455-499.

[9] W. Cook, C.R. Coullard, and Gy. Turán, *On the complexity of Cutting Plane proofs*, Discrete Applied Mathematics 18, (1987), 25-38

[10] W. Craig, *Linear reasoning: A new form of the Herbrand-Gentzen theorem,* Journ. of Symb. Logic 22(3), (1957), 250-287

[11] W. Craig, *Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory,* Journ. of Symb. Logic 22(3), (1957), 269-285

[12] R.E. Gomory, *An algorithm for integer solutions of linear programs,* in Recent Advances in Mathematical Programming, eds. R.L. Graves and P. Wolfe, McGraw-Hill, 1963, 269-302

[13] A. Haken, *The intractability of resolution,* Theoretical Computer Science, 39, (1985), 297-308

[14] A. Haken, *Counting bottlenecks to show monotone $P \neq NP$,* Proc. 36-th Annual IEEE Symp. on Foundations of Computer Science, Milwaukee, 1995, 36-40

[15] R. Impagliazzo, T. Pitassi, A. Urquhart, *Upper and lower bounds for tree-like cutting planes proofs,* Proc. 9-th Annual IEEE Symp. on Logic in Computer Science, 1994, 220-228

[16] J. Krajíček, *Lower bounds to the size of constant-depth propositional proofs,* Journ. of Symb. Logic 59,1 (1994), 73-86

[17] J. Krajíček, *Bounded Arithmetic, Propositional Logic and Complexity Theory,* Cambridge University Press 1995.

[18] J. Krajíček, *Interpolation theorems, lower bounds for proof systems, and independence results for bounded arithmetic,* Journ. of Symb. Logic, to appear

[19] J. Krajíček and P. Pudlák, *Some consequences of cryptographical conjectures for $S^1_2$ and $EF$,* in *Logic and Computational Complexity*, ed. D. Leivant, Springer-Verlag LNCS 960, 1995, 210-220.

[20] Ch.H. Papadimitriou, K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity,* Prentice-Hall 1982.

[21] J.B. Paris, A.J. Wilkie, A.R. Woods, *Provability of the pigeonhole principle and the existence of infinitely many primes,* Journ. of Symb. Logic 53 (1988), 1235-1244.

[22] A.A. Razborov, *Lower bounds on the monotone complexity of some boolean functions,* Doklady Akad. Nauk SSSR 282, (1985), 1033-1037

[23] 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.

[24] C.P. Schnorr, *A lower bound on the number of additions in monotone computations,* Theoretical Computer Science 2, (1976), 305-315.

[25] A. Schrijver, *On cutting planes,* in: Combinatorics 79, part II, Annals of Discrete Math. 9, eds. M. Deza and I.G. Rosenberg, 1980, North-Holland, 291-296.

[26] I. Wegener, *The Complexity of Boolean Functions,* Teubner and Wiley, 1987.