Exponential Separations between Restricted Resolution and Cutting Planes Proof Systems

Maria Luisa Bonet*, Juan Luis Esteban†, Nicola Galesi‡
Dept. Llenguatges i Sistemes Informatics
Universitat Politecnica de Catalunya
{bonet,esteban,galesi}@lsi.upc.es

Jan Johannsen§
Dept. of Mathematics
UCSD
johannsn@math.ucsd.edu

Abstract

We prove an exponential lower bound for tree-like Cutting Planes refutations of a set of clauses which has polynomial size resolution refutations. This implies an exponential separation between tree-like and dag-like proofs for both Cutting Planes and resolution; in both cases only superpolynomial separations were known before [30, 20, 10]. In order to prove this, we extend the lower bounds on the depth of monotone circuits of Raz and McKenzie [26] to monotone real circuits.

In the case of resolution, we further improve this result by giving an exponential separation of tree-like resolution from (dag-like) regular resolution proofs. In fact, the refutation provided to give the upper bound respects the stronger restriction of being a Davis-Putnam resolution proof. This extends the corresponding superpolynomial separation of [30].

Finally, we prove an exponential separation between Davis-Putnam resolution and unrestricted resolution proofs; only a superpolynomial separation was previously known [14].

1. Introduction

The motivation to work on the proof length of propositional proof systems comes from two sides. First, by the work of Cook and Reckhow [12], we know that the claim that for every propositional proof system there is a class of tautologies that requires superpolynomial proof size is equivalent to $NP \neq co\text{-}NP$. This connection explains the interest in developing combinatorial techniques to prove lower bounds for different proof systems. The second motivation comes from the interest in studying efficiency issues in Automated Theorem Proving. The question is which proof systems have efficient algorithms to find proofs. The most widely used proof system in implementations is resolution or restrictions of resolution. What we will show in this paper is that proving propositional proof complexity lower bounds has something to say about the non-efficiency of various strategies for finding proofs.

Haken [17] was the first who proved exponential lower bounds for unrestricted resolution. Later Urquhart [29] found another class of tautologies that require exponential size resolution proofs, and Chvátal and Szemerédi [8] showed that in some sense, almost all classes of tautologies require exponential size resolution proofs (see [3, 4] for simplified versions of these results). These exponential lower bounds are bad news for automated theorem provers, since they mean that often the time used in finding proofs will be exponentially long in the size of the tautology, given that the shortest proofs are. The next question is what about the classes of tautologies that have polynomial size proofs? Can we find these proofs efficiently? [3, 9, 4] give weakly exponential time ($2^{o(n)}$) algorithms for finding resolution proofs. But, can we do better? [19, 1] give weak evidence that the answer is negative.

Formally, we say that a propositional proof system S is *automatizable*, if there is an algorithm that for every tautology F finds a proof of F in S in time polynomial in the length of the shortest proof of F in S. The only propositional proof systems that we know are automatizable are algebraic proof systems like Hilbert's Nullstellensatz [2] and Polynomial Calculus [9]. On the other hand bounded-depth Frege proof systems are not automatizable, assuming factoring is hard [24, 7, 5]. Since Frege systems and Extended Frege systems polynomially simulates bounded-depth Frege systems, they are also not automatizable under the same assumptions.

A commonly used strategy for finding proofs is to reduce the search space by defining restricted versions of resolution

*Partially supported by projects SPRIT 20244 ALCOM-IT and TIC 97-1475-CE
†Partially supported by project KOALA:DGICYT:PB95-0787
‡Supported by an European Community grant under the TMR project
§Supported by DFG grant No. Jo 291/1-1

that are still complete. One possibility is to restrict to proofs that are tree-like, which would be a good strategy, given that [3, 9, 4] have quasipolynomial algorithms for finding tree-like proofs. Here we prove an exponential separation between tree-like resolution and resolution, showing that finding tree-like resolution proofs cannot be an efficient strategy for finding resolution proofs. Until now only superpolynomial separations were known [30, 10].

Many strategies for finding resolution proofs are described in [28], but very little theoretical work has been done until now. Goerdt [15, 14, 16] gave several superpolynomial separations between resolution and some restricted versions of it. In particular, he gave a separation between Davis-Putnam resolution and unrestricted resolution. We improve this result by giving an exponential separation between Davis-Putnam and unrestricted resolution, showing that using the Davis-Putnam restriction is not, in general, a good strategy for finding resolution proofs.

The Cutting Planes proof system (CP) is a refutation system based on manipulating integer linear inequalities for which the task of finding hard-to-prove tautologies is solved. [18] were the first to show such a result in the restricted case of CP proofs whose underlying graph is a tree. Pudlák [25] and Cook and Haken [11] give general circuit complexity results from which a exponential lower bounds for CP follow. Nothing is known about automatization of CP proofs. Since there is an exponential separation between CP and Resolution (CP is more efficient) it would be nice to find an efficient algorithm for finding CP proofs. A question to ask is if trying to find tree-like CP proofs would be an efficient strategy for finding Cutting Planes proofs.

One of the authors [20] gave a superpolynomial separation between tree-like CP and dag-like CP (this was previously known for a restricted form of CP from [6]). Here we improve that separation to exponential. This means again that trying to find tree-like proofs is not a good strategy.

This exponential separation is a consequence of extending the lower bounds of [26] to the case of real monotone circuits. As in [26] we prove an $\Omega(n^{\epsilon})$ lower bound on the depth of monotone real circuits computing a certain monotone function GEN_n in P. This also implies an $\Omega(2^{n^{\epsilon}})$ lower bound on the size of monotone real formulas computing GEN_n. This latter result allows us to obtain an exponential lower bound for the size of tree-like CP proofs for a formula associated to GEN_n, using the interpolation technique of [23, 25].

The paper is organized as follows: in Section 2 we give basic definitions of the proof systems we consider. Section 3 has the definitions of monotone real circuits, and the proof of the depth separation for them, extending the results of Raz and McKenzie. Section 4 gives the exponential separations between tree-like CP and CP, tree-like Resolution and Resolution and tree-like CP and bounded-depth Frege systems, and also the exponential separation between tree-like resolution and regular resolution. Finally section 5 has the exponential separation between Davis-Putnam resolution and unrestricted Resolution.

2. The Proof Systems

Resolution is a refutation proof system for formulas in CNF based on the following inference rule:

$$\frac{C \vee x \qquad D \vee \bar{x}}{C \vee D}.$$

A Resolution refutation for an inital set Σ of clauses is a derivation of the empty clause from Σ using the above inference rule. Several restrictions of the resolution proof system are known. Here we consider the following two: (1) the *regular* resolution system in which the proofs are restricted in such a way that any variable can be eliminated at most once in any path from an initial clause to the empty clause; (2) the *Davis Putnam* resolution system in which the proofs are restricted in such a way that there exists a sequence of the variables such that if a variable x is eliminated before a variable y on any path from an initial clause to the empty clause, then x is before y in the sequence.

Cutting Planes (CP) is a proof system operating with linear inequalities of the form $\sum_{i \in I} a_i x_i \geq k$, where the coefficients a_i and k are integers. The rules of CP are addition of two inequalities, multiplication of an inequality by a positive integer and the following division rule:

$$\frac{\sum_{i \in I} a_i x_i \geq k}{\sum_{i \in I} \frac{a_i}{b} x_i \geq \lceil \frac{k}{b} \rceil},$$

where b is a positive integer that evenly divides all a_i, $i \in I$.

A CP refutation of a set E of inequalities is a derivation of $0 \geq 1$ from the inequalities in E and the axioms $x \geq 0$ and $-x \geq -1$ for every variable x, using the rules of CP. It can be shown that a set of inequalities has a CP-refutation iff it has no $\{0, 1\}$-solution.

Cutting Planes can be used as a refutation system for propositional formulas in conjunctive normal form: note that a clause $\bigvee_{i \in P} x_i \vee \bigvee_{j \in N} \bar{x}_j$ is satisfiable iff the inequality $\sum_{i \in P} x_i - \sum_{j \in N} x_j \geq 1 - |N|$ has a $\{0, 1\}$-solution. It is also well-known that CP can simulate Resolution [13].

A proof system is *tree-like* if the proofs are restricted so that every line in a proof is used at most once as a premise of an inference. Otherwise we will call it *dag-like*.

3. Monotone Real Circuits

A *monotone real circuit* is a circuit of fan-in 2 computing with real numbers where every gate computes a nondecreasing real function. This class of circuits was introduced by

Pudlák [25]. We require that monotone real circuits output 0 or 1 on every input of zeroes and ones only, so that they are a generalization of monotone boolean circuits. Rosenbloom [27] shows that they are strictly more powerful than monotone boolean circuits.

The depth and size of a monotone real circuit are defined as usual, and we call it a *formula* if every gate has fan-out at most 1.

For a monotone boolean function f, we denote by $d_{\mathbb{R}}(f)$ the minimal depth of a monotone real circuit computing f, and by $s_{\mathbb{R}}(f)$ the minimal size of a monotone real formula computing f.

The method of proving lower bounds on the depth of monotone boolean circuits using communication complexity was used by Karchmer and Wigderson [21] to give an $\Omega(\log^2 n)$ lower bound on the monotone depth of *st*-connectivity. Using the notion of real communication complexity introduced by Krajíček [22], one of the authors [20] showed the same lower bound for monotone real circuits.

The monotone function GEN_n of n^3 inputs $t_{a,b,c}$, $1 \leq a, b, c \leq n$ is defined as follows: For $c \leq n$, we define the relation $\vdash c$ (c *is generated*) recursively by

$$\vdash c \text{ iff } c = 1 \text{ or there are } a, b \leq n \text{ with } \vdash a \ , \ \vdash b \text{ and } t_{a,b,c} = 1 \ .$$

Finally $\text{GEN}_n(\vec{t}) = 1$ iff $\vdash n$. From now on we will write $a, b \vdash c$ for $t_{a,b,c} = 1$.

Recently, Raz and McKenzie [26] gave a lower bound of $\Omega(n^\epsilon)$ for some $\epsilon > 0$ on the depth of monotone boolean circuits computing GEN_n. We show that their method applies to monotone real circuits:

Theorem 1 *For some $\epsilon > 0$ and sufficiently large n*

$$d_{\mathbb{R}}(\text{GEN}_n) \geq \Omega(n^\epsilon) \ \ and \ \ s_{\mathbb{R}}(\text{GEN}_n) \geq 2^{\Omega(n^\epsilon)} \ .$$

3.1. Real Communication Complexity

Let $R \subseteq X \times Y \times Z$ be a multifunction, i.e. for every pair $(x, y) \in X \times Y$, there is a $z \in Z$ with $(x, y, z) \in R$. A *real communication protocol* for R is executed by two players I and II, where I computes a function $f_I : X \times \{0,1\}^* \to \mathbb{R}$ and II computes a function $f_{II} : Y \times \{0,1\}^* \to \mathbb{R}$. Given inputs $x \in X$, $y \in Y$, the players generate a sequence w of bits as follows:

$$w_0 := \lambda$$

$$w_{k+1} := \begin{cases} w_k 0 & \text{if } f_I(x, w_k) > f_{II}(y, w_k) \\ w_k 1 & \text{else} \end{cases}$$

If there is a function $g : \{0,1\}^k \to Z$ such that

$$\forall x \in X \ \forall y \in Y \ (x, y, g(w_k)) \in R \ ,$$

then we say that the protocol solves R in k rounds. The *real communication complexity* $CC_{\mathbb{R}}(R)$ is the minimal number k such that there is a real communication protocol solving R in k rounds.

For a natural number n, let $[n]$ denote the set $\{1, \ldots, n\}$. Let $f : \{0,1\}^n \to \{0,1\}$ be a monotone boolean function, let $X := f^{-1}(1)$ and $Y := f^{-1}(0)$, and let the multifunction $R_f \subseteq X \times Y \times [n]$ be defined by

$$(x, y, i) \in R_f \quad \text{iff} \quad x_i = 1 \text{ and } y_i = 0$$

The Karchmer-Wigderson game for f is defined as follows: Player I receives an input $x \in X$ and Player II an input $y \in Y$. They have to agree on a position $i \in [n]$ such that $(x, y, i) \in R_f$. Sometimes we will say that R_f is the Karchmer-Wigderson game for the function f. There is a relation between the real communication complexity of R_f and the depth of a monotone real circuit or the size of a monotone real formula computing f, similar to the boolean case:

Lemma 2 (Krajíček [22]) *Let f be a monotone boolean function. Then*

$$CC_{\mathbb{R}}(R_f) \leq d_{\mathbb{R}}(f) \ and \ CC_{\mathbb{R}}(R_f) \leq \log_{3/2} s_{\mathbb{R}}(f) \ .$$

For a proof see [22] or [20]. Hence to establish Theorem 1, it suffices to prove:

Theorem 3 *For some $\epsilon > 0$ and sufficiently large n*

$$CC_{\mathbb{R}}(R_{\text{GEN}_n}) \geq \Omega(n^\epsilon).$$

3.2. DART games and structured protocols

Raz and McKenzie [26] introduced a special kind of communication games, called DART games, and a special class of communication protocols, the *structured protocols*, for solving them.

For $m, k \in \mathbb{N}$, the set of communication games $\text{DART}(m, k)$ is defined as follows:

- $X = [m]^k$. That is the inputs for the Player I are k-tuples of elements $x_i \in [m]$.
- $Y = (\{0,1\}^m)^k$. That is the inputs for the Player II are k-tuples of binary colorings y_i for $[m]$.
- For all $i = 1, \ldots, k$ let $e_i = y_i(x_i)$ (i.e. e_i is the x_i-th bit in y_i). The relation $R(x, y, z) \subseteq X \times Y \times Z$ defining the game, only depends on $e_1, \ldots, e_k$ and z. This means that we can describe $R(x, y, z)$ by $R((e_1, \ldots, e_k), z)$
- $R((e_1, \ldots, e_k), z)$ must be a DNF-Search-Problem. This means that always exists a tautology F_R defined over the variables $e_1, \ldots, e_k$ such that Z is the set of terms defining F_R and $R((e_1, \ldots, e_k), z)$ is true if and only if $z \in Z$ is the satisfied term of F_R.

A *structured protocol* for a DART game is a communication protocol for solving the relation R, where player I gets input $x \in X$, player II gets input $y \in Y$, and in each round, player I reveals the value x_i for some i, and II replies with $y_i(x_i)$. The structured communication complexity of $R \in \mathrm{DART}(m,k)$, denoted by $SC(R)$, is the minimal number of rounds in a structured protocol solving R.

The main theorem of [26] showed that for suitable m and k, the deterministic communication comlexity of a DART game cannot be much smaller than that of a structured protocol. We shall show the same for its real communication complexity. Obviously, a structured protocol solving R in r rounds can be simulated by a real communication protocol solving R in $r \cdot (\lceil \log m \rceil + 1)$ rounds. Conversely, the following holds:

Theorem 4 *For every relation $R \in \mathrm{DART}(m,k)$, where $m \geq k^{14}$,*

$$CC_{\mathbb{R}}(R) \geq SC(R) \cdot \Omega(\log m)$$

To prove this, first we need some combinatorial notions and results from [26]. Let $A \subseteq [m]^k$ and $1 \leq j \leq k$. For $x \in [m]^{k-1}$, let $\deg_j(x,A)$ be the number of $\xi \in [m]$ such that $(x_1, \dots, x_{j-1}, \xi, x_j, \dots, x_{k-1}) \in A$. Then we define

$$\begin{aligned} A[j] &:= \{\, x \in [m]^{k-1} \,;\, \deg_j(x,A) > 0 \,\} \\ AVDEG_j(A) &:= \frac{|A|}{|A[j]|} \\ MINDEG_j(A) &:= \min_{x \in A[j]} \deg_j(x,A) \\ Thickness(A) &:= \min_{1 \leq j \leq k} MINDEG_j(A) \;. \end{aligned}$$

The following lemmas about these notions were proved in [26]:

Lemma 5 *For every $A' \subseteq A$ and $1 \leq j \leq k$,*

$$AVDEG_j(A') \geq \frac{|A'|}{|A|} AVDEG_j(A) \qquad (1)$$

$$Thickness(A[j]) \geq Thickness(A) \qquad (2)$$

Lemma 6 *If for every $1 \leq j \leq k$, $AVDEG_j(A) \geq \delta m$ for some $0 < \delta < 1$, then for every $\alpha > 0$ there is $A' \subseteq A$ with $|A'| \geq (1-\alpha)|A|$ and*

$$Thickness(A') \geq \frac{(1-\alpha)\delta m}{k(1 + \alpha^{-1} \ln(\delta^{-1}))} \;.$$

In particular, setting $\alpha = \frac{1}{2}$ and $\delta = 4m^{-\frac{1}{14}}$, we get

Corollary 7 *If $m \geq k^{14}$ and for every $1 \leq j \leq k$, $AVDEG_j(A) \geq 4m^{\frac{13}{14}}$, then there is $A' \subseteq A$ with $|A'| \geq \frac{1}{2}|A|$ and $Thickness(A) \geq m^{\frac{11}{14}}$.*

For a relation $R \in \mathrm{DART}(m,k)$, $A \subseteq X$ and $B \subseteq Y$, let $CC_{\mathbb{R}}(R,A,B)$ be the real communication complexity of R restricted to $A \times B$.

Fix a large $m \in \mathbb{N}$. A triple (R,A,B) is called an (α,β,ℓ)-game if $R \in \mathrm{DART}(m,k)$ for some $k \leq m^{\frac{1}{14}}$ with $SC(R) \geq \ell$, $A \subseteq X$ with $|A| \geq 2^{-\alpha}|X|$ and $Thickness(A) \geq m^{\frac{11}{14}}$, and $B \subseteq Y$ with $|B| \geq 2^{-\beta}|Y|$.

Lemma 8 *For every $\alpha, \beta, \ell \geq 0$ with $\beta \leq m^{\frac{1}{7}}$ and every (α,β,ℓ)-game (R,A,B),*

1. *if for every $1 \leq j \leq k$, $AVDEG_j(A) \geq 8m^{\frac{13}{14}}$, then there is an $(\alpha+2, \beta+1, \ell)$-game (R',A',B') with*

$$CC_{\mathbb{R}}(R',A',B') \leq CC_{\mathbb{R}}(R,A,B) - 1 \;.$$

2. *if $\ell \geq 1$ and for some $1 \leq j \leq k$, $AVDEG_j(A) < 8m^{\frac{13}{14}}$, then there is an $(\alpha + 3 - \frac{\log m}{14}, \beta+1, \ell-1)$-game (R',A',B') with*

$$CC_{\mathbb{R}}(R',A',B') \leq CC_{\mathbb{R}}(R,A,B) \;.$$

To prove Theorem 3 from the lemma, we show that for every (α,β,ℓ)-game (R,A,B),

$$CC_{\mathbb{R}}(R,A,B) \geq \ell \cdot \Big(\frac{\log m}{42} - \frac{4}{3}\Big) - \frac{\alpha+\beta}{3} \;. \qquad (*)$$

The case $\alpha = \beta = 0$ gives the theorem.

For $\ell = 0$ and $\beta > m^{\frac{1}{7}}$, $(*)$ is trivial, since the right hand side gets negative for large m. We proceed inductively: Let (R,A,B) be an (α,β,ℓ)-game, and assume that $(*)$ holds for all (α',β',ℓ')-games with $\ell' \leq \ell$ and $\beta' > \beta$. For sake of contradiction, suppose that $CC_{\mathbb{R}}(R,A,B) < \ell \cdot \left(\frac{\log m}{42} - \frac{4}{3}\right) - \frac{\alpha+\beta}{3}$. Then either for every $1 \leq j \leq k$, $AVDEG_j(A) \geq 8m^{\frac{13}{14}}$, and Lemma 8 gives an $(\alpha+2, \beta+1, \ell)$-game (R',A',B') with

$$\begin{aligned} CC_{\mathbb{R}}(R',A',B') &\leq CC_{\mathbb{R}}(R,A,B) - 1 < \\ &< \ell \cdot \Big(\frac{\log m}{42} - \frac{4}{3}\Big) - \frac{(\alpha+2)+(\beta+1)}{3} \;, \end{aligned}$$

or for some $1 \leq j \leq k$, $AVDEG_j(A) < 8m^{\frac{13}{14}}$, then Lemma 8 gives an $(\alpha + 3 - \frac{\log m}{14}, \beta+1, \ell-1)$-game (R',A',B') with

$$\begin{aligned} CC_{\mathbb{R}}(R',A',B') &< \ell \cdot \Big(\frac{\log m}{42} - \frac{4}{3}\Big) - \frac{\alpha+\beta}{3} \\ &= (\ell-1) \cdot \Big(\frac{\log m}{42} - \frac{4}{3}\Big) - \frac{(\alpha + 3 - \frac{\log m}{14}) + (\beta+1)}{3} \;, \end{aligned}$$

both contradicting the assumption.

Proof of Lemma 8: For part 1, we first show that $CC_{\mathbb{R}}(R,A,B) > 0$. Assume otherwise, then there is a term C_z in the DNF tautology defining R that is satisfied for every $(x, y) \in A \times B$. Therefore $y_j(x_j)$ is constant for some $1 \leq j \leq k$. If γ denote the number of possible values of x_j in elements of A, then this implies that $|B| \leq 2^{mk-\gamma}$. On the other hand, $|B| \geq 2^{mk-\beta}$, hence it follows that $\beta \geq \gamma$, which is a contradiction since $\beta \leq m^{\frac{1}{7}}$, whereas $AVDEG_j(A) \geq 8m^{\frac{13}{14}}$ implies $\gamma \geq 8m^{\frac{13}{14}}$.

Now let an optimal real communication protocol solving R restricted to $A \times B$ be given. For $a \in A$ and $b \in B$, let ρ_a and σ_b be the real numbers played by I and II in the first round on input a and b, respectively. W.l.o.g. we can assume that these are $|A| + |B|$ distinct real numbers.

Now consider a $\{0, 1\}$-matrix of size $|A| \times |B|$ with columns indexed by the ρ_a and rows indexed by the σ_b, where the entry in position (ρ_a, σ_b) is the outcome of the first round when these numbers are played. Then it is obvious that either the upper right quadrant or the lower left quadrant must form a monochromatic rectangle.

Hence there are $A^\circ \subseteq A$ and $B' \subseteq B$ with $|A^\circ| \geq \frac{1}{2}|A|$ and $|B'| \geq \frac{1}{2}|B|$ such that R restricted to $A^\circ \times B'$ can be solved in one round fewer than the original protocol. By Lemma 5 (1), $AVDEG_j(A^\circ) \geq 4m^{\frac{13}{14}}$ for every $1 \leq j \leq k$, hence by Corollary 7 there is $A' \subseteq A^\circ$ with $|A'| \geq \frac{1}{2}|A^\circ| \geq \frac{1}{4}|A|$ and $Thickness(A') \geq m^{\frac{11}{14}}$. Thus (R, A', B') is an $(\alpha + 2, \beta + 1, \ell)$-game.

Part 2 is proved exactly like the corresponding lemma in [26], with the numbers slightly adjusted. □

3.3. A DART game related to GEN_n

The communication game $\text{PyrGen}(m, d)$ is defined as follows:

Let $Pyr_d := \{(i, j)\,;\, 1 \leq j \leq i \leq d\}$. We regard the indices as elements of Pyr_d, so that the inputs for the two players I and II are respectively sequences of elements $x_{i,j} \in [m]$ and $y_{i,j} \in \{0, 1\}^m$ with $(i, j) \in Pyr_d$, and we picture these as laid out in a pyramidal form with $(1, 1)$ at the top and (d, j), $1 \leq j \leq d$ and the bottom. The goal of the game is to find either an element colored 0 at the top of the pyramid, or an element colored 1 at the bottom of the pyramid, or an element colored 1 with the two elements below it colored 0, i.e. to find indices (i, j) such that one of the following holds:

1. $i = j = 1$ and $y_{1,1}(x_{1,1}) = 0$, or
2. $y_{i,j}(x_{i,j}) = 1$ and $y_{i+1,j}(x_{i+1,j}) = 0$ and $y_{i+1,j+1}(x_{i+1,j+1}) = 0$, or
3. $i = d$ and $y_{d,j}(x_{d,j}) = 1$.

Obviously, $\text{PyrGen}(m, d)$ is a game in DART$(m, \binom{d+1}{2})$. The following lower bound on the structured communication complexity of $\text{PyrGen}(m, d)$ was proved in [26]:

Lemma 9 *$SC(\text{PyrGen}(m, d)) \geq d$.*

Hence by Theorem 4, we get $CC_{\mathbb{R}}(\text{PyrGen}(m, d)) \geq \Omega(d \log m)$ for $m \geq d^{28}$.

The following lemma shows that the real communication complexity of $\text{PyrGen}(m, d)$ is bounded by the real communication complexity of the Karchmer-Wigderson game for Gen_n for a suitable n.

Lemma 10 *For $n := m \cdot \binom{d+1}{2} + 2$,*

$$CC_{\mathbb{R}}(\text{PyrGen}(m, d)) \leq CC_{\mathbb{R}}(\text{Gen}_n).$$

Proof: We interpret the elements between 2 and $n - 1$ as triples (i, j, k), where $(i, j) \in Pyr_d$ and $k \in [m]$.

Now player I computes from his input $x : Pyr_d \to [m]$ an input $\vec{t}_x$ to Gen_n with $\text{Gen}_n(\vec{t}_x) = 1$ by setting the following:

$$\begin{aligned} &1, 1 \vdash a_{d,j} && \text{for } 1 \leq j \leq d \\ &a_{1,1}, a_{1,1} \vdash n && \\ &a_{i+1,j}, a_{i+1,j+1} \vdash a_{i,j} && \text{for } (i, j) \in Pyr_{d-1} \end{aligned}$$

where $a_{i,j} := (i, j, x_{i,j})$. This completely determines $\vec{t}_x$.

Likewise Player II computes from his input $y : Pyr_d \to (2^{[m]})$ a coloring c of the elements from $[n]$ by setting $col(1) = 0$, $col(n) = 1$ and $col((i, j, k)) = y_{i,j}(k)$. From this, he computes an input $\vec{t}_y$ by setting $a, b \vdash c$ iff it is not the case that $col(c) = 1$ and $col(a) = col(b) = 0$. Obviously $\text{Gen}_n(\vec{t}_y) = 0$.

Playing the Karchmer-Wigderson game for Gen_n now yields a triple (a, b, c) such that $a, b \vdash c$ in $\vec{t}_x$ and $a, b \not\vdash c$ in $\vec{t}_y$. By definition of $\vec{t}_y$, this means that $col(a) = col(b) = 0$ and $col(c) = 1$, and by definition of $\vec{t}_x$ one of the following cases must hold:

- $a = b = 1$ and $c = a_{d,j}$ for some $j \leq d$. By definition of col, $y_{d,j}(x_{d,j}) = 1$.
- $c = n$ and $a = b = a_{1,1}$. In this case, $y_{1,1}(x_{1,1}) = 0$.
- $a = a_{i+1,j}$, $b = a_{i+1,j+1}$ and $c = a_{i,j}$. Then we have $y_{i,j}(x_{i,j}) = 1$, and $y_{i+1,j}(x_{i+1,j}) = y_{i+1,j+1}(x_{i+1,j+1}) = 0$.

In either case, the players have solved $\text{PyrGen}(m, d)$ without any additional communication. □

Now the lower bound on $CC_{\mathbb{R}}(\text{PyrGen}(m, d))$ obtained from Lemma 9 and Theorem 4, together with Lemma 10 immediately imply Theorem 3 with $\epsilon = \frac{1}{30}$ by taking $m = d^{28}$.

Let $\vec{t}$ be an input to Gen_n. We say that n is generated in a depth-d pyramidal fashion by $\vec{t}$ if there is a mapping $m : Pyr_d \to [n]$ such that $1, 1 \vdash m(d, j)$ for every $j \leq d$, $m(i+1, j), m(i + 1, j + 1) \vdash m(i, j)$ for every $(i, j) \in Pyr_{d-1}$

and $m(1,1), m(1,1) \vdash n$ (recall that $a, b \vdash c$ means $t_{a,b,c} = 1$).

As the reduction in Lemma 10 produces only inputs from $\mathrm{GEN}_n^{-1}(1)$ which have the additional property that n is generated in a depth-d pyramidal fashion, we can state the following strengthening of Theorem 1:

Corollary 11 *Let n, d be as above. Every monotone real formula that outputs 1 on every input to GEN_n for which n is generated in a depth-d pyramidal fashion, and outputs 0 on all inputs where GEN_n is 0, has to be of size $\Omega(2^{n^\epsilon})$.*

The other consequences drawn from Theorem 4 and Lemma 9 in [26] apply to monotone real circuits as well, e.g. we just state without proof the following result:

Theorem 12 *There are constants $\epsilon, c > 0$ such that for every function $d(n) \leq n^\epsilon$, there is a family of monotone functions $f_n : \{0,1\}^n \to \{0,1\}$ that can be computed by monotone boolean circuits of size $n^{O(1)}$ and depth $d(n)$, but cannot be computed by monotone real circuits of depth less than $c \cdot d(n)$.*

The method also gives a simpler proof of the lower bounds in [20], in the same way as [26] simplifies the lower bound of [21].

4. Separation between tree-like and dag-like versions of Resolution and Cutting Planes

Cutting Planes refutations are linked to monotone real circuits by the following interpolation theorem due to Pudlák:

Theorem 13 (Pudlák [25]) *Let $\vec{p}, \vec{q}, \vec{r}$ be disjoint vectors of variables, and let $A(\vec{p}, \vec{q})$ and $B(\vec{p}, \vec{r})$ be sets of inequalities in the indicated variables such that the variables $\vec{p}$ either have only nonnegative coefficients in $A(\vec{p}, \vec{q})$ or have only nonpositive coefficients in $B(\vec{p}, \vec{r})$.*

Suppose there is a CP refutation R of $A(\vec{p}, \vec{q}) \cup B(\vec{p}, \vec{r})$. Then there is a monotone real circuit $C(\vec{p})$ of size $O(|R|)$ such that for any vector $\vec{a} \in \{0,1\}^{|\vec{p}|}$

$$C(\vec{a}) = 0 \quad \to \quad A(\vec{a}, \vec{q}) \text{ is unsatisfiable}$$
$$C(\vec{a}) = 1 \quad \to \quad B(\vec{a}, \vec{r}) \text{ is unsatisfiable}$$

Furthermore, if R is tree-like, then $C(\vec{p})$ is a monotone real formula.

We now define an unsatisfiable set of clauses related to GEN_n. The variables $p_{a,b,c}$ for $a, b, c \in [n]$ represent the input to GEN_n. Variables $q_{i,j,a}$ for $(i,j) \in Pyr_d$ and $a \in [n]$ encode a pyramid where the element a is assigned to the position (i,j) by a certain mapping $m : Pyr_d \to [n]$ (cf. Corollary 7). Finally the variables r_a for $a \in [n]$ represent a coloring of the elements by $0, 1$ such that 1 is colored 0, n is colored 1 and the elements colored 0 are closed under generation.

The sets of clauses $Gen(\vec{p}, \vec{q})$ and $Col(\vec{p}, \vec{r})$ are defined in Table 1. Obviously, if $Gen(\vec{t}, \vec{q})$ is satisfiable for a fixed vector $\vec{t} \in \{0,1\}^{n^3}$, then n is generated in a depth-d pyramidal fashion, and if $Col(\vec{t}, \vec{r})$ is satisfiable, then $\mathrm{GEN}(\vec{t}) = 0$. Since the variables $\vec{p}$ occur only positively in $Gen(\vec{p}, \vec{q})$ and only negatively in $Col(\vec{p}, \vec{r})$, Theorem 13 is applicable, and the formula obtained from this application satisfies the conditions of Corollary 11. Hence we can conclude:

Theorem 14 *For some $\epsilon > 0$, tree-like CP refutations of the clauses $Gen(\vec{p}, \vec{q}) \cup Col(\vec{p}, \vec{r})$ have to be of size $2^{\Omega(n^\epsilon)}$.*

On the other hand, there are polynomial size dag-like resolution refutations of these clauses.

Theorem 15 *There are (dag-like) resolution refutations of size $n^{O(1)}$ of the clauses $Gen(\vec{p}, \vec{q}) \cup Col(\vec{p}, \vec{r})$.*

As the proof is very similar to that of Theorem 18 below, we omit it. The following corollary follows by the last two Theorems and well-known simulation results:

Corollary 16 *The clauses $Gen(\vec{p}, \vec{q}) \cup Col(\vec{p}, \vec{r})$ exponentially separate the following proof systems: Tree-like from dag-like Resolution, tree-like Cutting Planes from dag-like Cutting Planes and tree-like Cutting Planes from bounded-depth Frege systems.*

4.1. Separation of tree-like CP from regular resolution

We now modify the clauses $Col(\vec{p}, \vec{r})$, so that the modified clauses allow small regular resolutions, but in such a way that the lower bound proof still applies. We replace the variables r_a by $r_{a,i,D}$ for $a \in [n]$, $1 \leq i \leq d$ and $D \in \{L, R\}$, giving the coloring of element a, with auxiliary indices i being a row in the pyramid and D distinguishing whether an element is used as a left or right predecessor in the generation process.

The set $RCol(\vec{p}, \vec{r})$ is defined in Table 2. Due to the clauses (13) and (14), the variables $r_{a,i,D}$ are equivalent for all values of the auxiliary indices i, D. Hence a satisfying assignment for $RCol(\vec{p}, \vec{r})$ still codes a coloring of $[n]$ such that elements that can be generated from 1 are colored 0, the elements from which n can be generated are colored 1, and the 0-colored elements are closed under generation. Hence if $RCol(\vec{t}, \vec{r})$ is satisfiable, then $\mathrm{GEN}(\vec{t}) = 0$.

Hence any interpolant for the clauses $Gen(\vec{p}, \vec{q}) \cup RCol(\vec{p}, \vec{r})$ satisfies the assumptions of Corollary 11, and we can conclude

Theorem 17 *Tree-like CP refutations of the clauses $Gen(\vec{p}, \vec{q}) \cup RCol(\vec{p}, \vec{r})$ have to be of size $2^{\Omega(n^\epsilon)}$.*

$\bigvee_{1\leq a\leq n} q_{i,j,a}$	for $(i,j) \in Pyr_d$	(3)
$\bar{q}_{d,j,a} \vee p_{1,1,a}$	for $1 \leq j \leq d$ and $a \in [n]$	(4)
$\bar{q}_{1,1,a} \vee p_{a,a,n}$	for $a \in [n]$	(5)
$\bar{q}_{i+1,j,a} \vee \bar{q}_{i+1,j+1,b} \vee \bar{q}_{i,j,c} \vee p_{a,b,c}$	for $(i,j) \in Pyr_{d-1}$ and $a,b,c \in [n]$	(6)
$\bar{r}_1$		(7)
r_n		(8)
$r_a \vee r_b \vee \bar{p}_{a,b,c} \vee \bar{r}_c$	for $a,b,c \in [n]$	(9)

Table 1. The set $Gen(\vec{p},\vec{q})$ is given by (3) - (6), and $Col(\vec{p},\vec{r})$ by (7) - (9).

$\bar{p}_{1,1,a} \vee \bar{r}_{a,d,D}$	for $a \in [n]$ and $D \in \{L,R\}$	(10)
$\bar{p}_{a,a,n} \vee r_{a,1,D}$	for $a \in [n]$ and $D \in \{L,R\}$	(11)
$r_{a,i+1,L} \vee r_{b,i+1,R} \vee \bar{p}_{a,b,c} \vee \bar{r}_{c,i,D}$	for $(i,j) \in Pyr_{d-1},\ a,b,c \in [n]$ and $D \in \{L,R\}$	(12)
$\bar{r}_{a,i,D} \vee r_{a,i,\bar{D}}$	for $1 \leq i \leq d$ and $D \in \{L,R\}$	(13)
$\bar{r}_{a,i,D} \vee r_{a,j,D}$	for $1 \leq i,j \leq d$ and $D \in \{L,R\}$	(14)

Table 2. The set of clauses $RCol(\vec{p},\vec{r})$.

On the other hand, we have the following upper bound on (dag-like) regular resolution refutations of these clauses:

Theorem 18 *There are (dag-like) regular resolution refutations of the clauses $Gen(\vec{p},\vec{q}) \cup RCol(\vec{p},\vec{r})$ of size $n^{O(1)}$.*

Proof: First we resolve clauses (4) and (10) to get

$$\bar{q}_{d,j,a} \vee \bar{r}_{a,d,D} \tag{15}$$

for $1 \leq j \leq d$, $1 \leq a \leq n$ and $D \in \{L,R\}$. Next we resolve (5) and (11) to get

$$\bar{q}_{1,1,a} \vee r_{a,1,D} \tag{16}$$

for $1 \leq a \leq n$ and $D \in \{L,R\}$. Finally, from (6) and (12) we obtain

$$\bar{q}_{i+1,j,a} \vee \bar{q}_{i+1,j+1,b} \vee \bar{q}_{i,j,c} \vee r_{a,i+1,L} \vee r_{b,i+1,R} \vee \bar{r}_{c,i,D} \tag{17}$$

for $1 \leq j \leq i < d$, $1 \leq a,b,c \leq n$ and $D \in \{L,R\}$.

Now we want to derive $\bar{q}_{i,j,a} \vee \bar{r}_{a,i,D}$ for every $(i,j) \in Pyr_d$, $1 \leq a \leq n$ and $D \in \{L,R\}$, by induction on i downward from d to 1. The induction base is just (15).

For the inductive step, resolve (17) against the clauses

$$\bar{q}_{i+1,j,a} \vee \bar{r}_{a,i+1,L} \quad \text{and} \quad \bar{q}_{i+1,j+1,b} \vee \bar{r}_{b,i+1,R}\ ,$$

which we have by induction, to give

$$\bar{q}_{i+1,j,a} \vee \bar{q}_{i+1,j+1,b} \vee \bar{q}_{i,j,c} \vee \bar{r}_{c,i,D}$$

for every $1 \leq a,b \leq n$.

All of these are then resolved against two instances of (3), and we get the desired $\bar{q}_{i,j,c} \vee \bar{r}_{c,i,D}$.

Finally, we have in particular $\bar{q}_{1,1,a} \vee \bar{r}_{a,1,L}$, which we resolve against (16) to get $\bar{q}_{1,1,a}$ for every $a \leq n$. From these and an instance of (3) we get the empty clause. □

A proof of the upper bound in Theorem 15 can be obtained from this by simply omitting the auxiliary indices from the variables $r_{a,i,D}$. Note that the refutation given in the proof of Thm. 18 is actually a Davis-Putnam refutation: It respects the following elimination order

$$\begin{array}{l}
p_{1,1,1}\ \cdots\ p_{n,n,n} \\
r_{1,d,L}\ \ r_{1,d,R}\ \ \cdots\ \ r_{n,d,L}\ \ r_{n,d,R} \\
q_{1,d,1}\ \cdots\ q_{1,d,n}\ \cdots\ q_{d,d,1}\ \cdots\ q_{d,d,n} \\
r_{1,d-1,L}\ \cdots\ r_{n,d-1,R}\ \ q_{1,d-1,1}\ \cdots\ q_{d-1,d-1,n} \\
\vdots \\
r_{1,1,L}\ \ r_{1,1,R}\ \ q_{1,1,1}\ \cdots\ q_{1,1,n}\ .
\end{array}$$

5. Lower bound for Davis-Putnam resolutions

Goerdt [14] gives a superpolynomial separation of Davis-Putnam resolution from unrestricted resolution. The lower bound he gives is of the order $n^{\Omega(\log\log n)}$. By applying his method to a modification of the clauses $Gen(\vec{p}, \vec{q}) \cup Col(\vec{p}, \vec{r})$, we can improve the separation to exponential.

We modify the clauses $Gen(\vec{p}, \vec{q})$ in such a way as to make small Davis-Putnam resolution refutations impossible, while still allowing for small unrestricted resolutions. The lower bound is proved by a bottleneck counting argument similar to that used in [14], which is based on the original argument of [17].

Let $d \geq 8$ be divisible by 4 and let $n = d^3$, and choose a mapping $\mu : [d] \times [\frac{d}{2}] \to Pyr_d$ such that no element from column i is mapped to rows between $i-1$ between $i+1$, i.e. if $\mu(i,j) = (i', j')$, then $i' \notin \{i-1, i, i+1\}$, and such that no two elements from the same column are mapped to the same position, i.e. if $j_1 \neq j_2$, then $\mu(i, j_1) \neq \mu(i, j_2)$. Such mappings are easy to construct; note that we do not require μ to be injective.

The set of clauses $DPGen(\vec{p}, \vec{q})$ is built from $Gen(\vec{p}, \vec{q})$ by adding additional literals to some of the clauses (4) and (6). The clauses (4) for $1 \leq j \leq d$ and $a \leq \frac{d}{2}$ are replaced by

$$\bar{q}_{i',j',b} \vee \bar{q}_{d,j,a} \vee p_{1,1,a} \tag{18}$$

for every $b \in [n]$, where $(i', j') = \mu(d, a)$. The clauses (6) for $(i,j) \in Pyr_{d-1}$, $a, b \in [n]$ and $1 \leq c \leq \frac{d}{2}$ are replaced by

$$\bar{q}_{i',j',e} \vee \bar{q}_{i+1,j,a} \vee \bar{q}_{i+1,j+1,b} \vee \bar{q}_{i,j,c} \vee p_{a,b,c} \tag{19}$$

for every $e \in [n]$, where $(i', j') = \mu(i, c)$. All other clauses remain unchanged.

Proposition 19 *There are (dag-like) unrestricted resolution refutations of the clauses $DPGen(\vec{p}, \vec{q}) \cup Col(\vec{p}, \vec{r})$ of size $n^{O(1)}$.*

Proof: First, from the clauses (18) and (3) derive the original clauses (4), and from (19) and (3) derive (6). Then apply the refutations from the proof of Theorem 15, which of course work for any values of n and d. □

Definition: A critical assignment α is given by

- a coloring $col_\alpha \in 2^{[n]}$ such that $col_\alpha(1) = 0$ and $col_\alpha(n) = 1$. The values $\alpha(r_a)$ are assigned according to $col_\alpha(a)$.
- a set of triples $G_\alpha \subseteq [n]^3$ such that for no triple $(a, b, c) \in G_\alpha$, $col_\alpha(a) = col_\alpha(b) = 0$ and $col_\alpha(c) = 1$. Values $\alpha(p_{a,b,c})$ are assigned according to G_α.
- A position $(i_\alpha, j_\alpha) \in Pyr_d$ with $\alpha(q_{i_\alpha,j_\alpha,a}) = 0$ for every $a \in [n]$.
- A mapping $m_\alpha : Pyr_d \setminus \{i_\alpha, j_\alpha\} \to [n]$ such that
  - every triangle is consistent with G_α, i.e. for every $(i,j) \in Pyr_{d-1}$ such that $(i_\alpha, j_\alpha) \notin \{(i,j), (i+1, j), (i+1, j+1)\}$

    $$(m_\alpha(i+1, j), m_\alpha(i+1, j+1), m_\alpha(i,j))$$

    is in G_α.
  - if $(i_\alpha, j_\alpha) \neq (1,1)$, then $(m_\alpha(1,1), m_\alpha(1,1), n) \in G_\alpha$.
  - $(1, 1, m_\alpha(d, j)) \in G_\alpha$ for every j such that $(d, j) \neq (i_\alpha, j_\alpha)$.

  Then $\alpha(q_{i,j,m_\alpha(i,j)}) = 1$ and $\alpha(q_{i,j,b}) = 0$ for all $b \neq m_\alpha(i,j)$, for every $(i,j) \neq (i_\alpha, j_\alpha)$.

A critical assignment satisfies all clauses from $Col(\vec{p}, \vec{r})$, and all clauses from $DPGen(\vec{p}, \vec{q})$ except for $\bigvee_{a \in [n]} q_{i_\alpha, j_\alpha, a}$.

Theorem 20 *(Dag-like) Davis-Putnam resolution refutations of the clauses $DPGen(\vec{p}, \vec{q}) \cup Col(\vec{p}, \vec{r})$ have to be of size $\Omega(2^{\frac{1}{4} n^{\frac{1}{3}}})$.*

Proof: Let an elimination order $\langle x_1, \ldots, x_N \rangle$ be given, where $N = n^3 + \binom{d+1}{2} n + n$ is the number of variables, and a Davis-Putnam refutation R of $DPGen(\vec{p}, \vec{q}) \cup Col(\vec{p}, \vec{r})$ respecting this elimination order be given. For $(i,j) \in Pyr_d$ and $s \leq N$, let $S(i,j,s) := \left\{ a \leq \frac{d}{2} \, ; \, q_{i,j,a} \in \{x_1, \ldots, x_s\} \right\}$. Let (i_0, j_0) denote the unique position in Pyr_d such that there is an index $s_0 \leq N$ with $|S(i_0, j_0, s_0)| = \frac{d}{4}$, and for all $(i,j) \neq (i_0, j_0)$, $|S(i,j,s_0)| < \frac{d}{4}$. In other words, (i_0, j_0) is the first position in Pyr_d for which $\frac{d}{4}$ variables $q_{i_0,j_0,a}$ with $a \leq \frac{d}{2}$ are eliminated. Let $\{a_1, \ldots, a_{\frac{d}{4}}\}$ denote $S(i_0, j_0, s_0)$. For each $1 \leq k \leq \frac{d}{4}$, let (i_k, j_k) denote $\mu(i_0, a_k)$, and define $R_k := [\frac{d}{2}] \setminus S(i_k, j_k, s_0)$, i.e. R_k is the set of those $a \leq \frac{d}{2}$ for which $q_{i_k,j_k,a}$ is eliminated later than any q_{i_0,j_0,a_ℓ} for $1 \leq \ell \leq \frac{d}{4}$. Note that $|R_k| \geq \frac{d}{4}$ by definition of (i_0, j_0) and by the first requirement for μ.

A critical assignment α is 0-critical if $(i_\alpha, j_\alpha) = (i_0, j_0)$ and $m_\alpha(i_k, j_k) \in R_k$, and furthermore the following conditions hold

- $(m_\alpha(i_0+1, j_0), m_\alpha(i_0+1, j_0+1), a_k) \notin G_\alpha$ if $i_0 \neq d$ or $(1, 1, a_k) \notin G_\alpha$ if $i_0 = d$
- if $i_0, j_0 > 1$, then $(m_\alpha(i_0, j_0-1), a_k, m_\alpha(i_0-1, j_0-1)) \in G_\alpha$
- if $i_0 > 1$ and $j_0 < i_0$, then $(a_k, m_\alpha(i_0, j_0+1), m_\alpha(i_0-1, j_0)) \in G_\alpha$

for every $1 \leq k \leq \frac{d}{4}$.

The next lemma shows that there are many 0-critical assignments.

Lemma 21 *For every choice of pairwise distinct values $b_1, \ldots, b_{\frac{d}{4}}$ with $b_k \in R_k$, there is a 0-critical assignment α with $m_\alpha(i_k, j_k) = b_k$ for $1 \leq k \leq \frac{d}{4}$.*

Proof: The assignment α is constructed as follows:

1. If $i_0 < d$, then values $m_\alpha(i_0+1, j_0) = c_1$ and $m_\alpha(i_0+1, j_0+1) = c_2$ are assigned with $\frac{d}{2} < c_1, c_2 \leq d$.

2. For each $(i,j) \neq (i_0, j_0)$ for which no value $m_\alpha(i,j)$ has been assigned yet, i.e. $(i,j) \notin \{(i_1,j_1), \ldots, (i_{\frac{d}{4}}, j_{\frac{d}{4}}), (i_0+1, j_0), (i_0+1, j_0+1)\}$, assign a value $n - id \leq m_\alpha(i,j) < n-(i-1)d$, such that no value is assigned twice.

3. Put all triples occurring in the pyramid and those required by the definition of 0-critical into G_α, and no others, i.e. G_α contains the triple $(m_\alpha(1,1), m_\alpha(1,1), n)$, all triples $(1, 1, m_\alpha(d,j))$ for $(d,j) \in Pyr_d \setminus \{(i_\alpha, j_\alpha)\}$ and all triples $(m_\alpha(i+1,j), m_\alpha(i+1,j+1), m_\alpha(i,j))$ such that $\{(i,j),(i+1,j),(i+1,j+1)\} \subseteq Pyr_d \setminus \{(i_\alpha, j_\alpha)\}$, and for $i_0 > 1$, all triples $(m_\alpha(i_0, j_0-1), a_k, m_\alpha(i_0-1, j_0-1))$ if $j_0 > 1$ and $(a_k, m_\alpha(i_0, j_0+1), m_\alpha(i_0-1, j_0))$ if $j_0 < i_0$.

4. Color all elements in rows $i_\alpha, \ldots, d$ by 0, and also all elements that are thereby forced to have color 0 by the second clause in the definition of critical assignment, i.e. if $(a,b,c) \in G_\alpha$ and a, b have already been colored 0, then also c is colored 0. Color all remaining elements by 1.

To verify that α is 0-critical, observe that the only elements $\leq \frac{d}{2}$ appearing in the pyramid are the b_k, so this is the only way that the values a_k can occur in the pyramid.. If $i_0 < d$, then as $n = d^3 > d^2+d$, the elements c_1, c_2 do not appear in the pyramid anywhere else but at $(i_0+1, j_0), (i_0+1, j_0+1)$, hence no triple (c_1, c_2, a_k) gets put into G_α. If $i_0 = d$, then $i_k \neq d$ for every k, so no triple $(1, 1, a_k)$ gets put into G_α.

The elements $m_\alpha(i_0, j_0-1)$ and $m_\alpha(i_0, j_0+1)$, if defined, cannot occur adjacent to any a_k, and so the elements $m_\alpha(i_0-1, j_0-1)$ and $m_\alpha(i_0-1, j_0)$ are not forced to be colored 0, hence they get colored 1. Therefore everything that is above these positions in the pyramid gets colored 1 also, as indicated in Figure 1.

In particular, if $m_\alpha(1,1)$ is defined, it is colored 1, and thus n is colored 1. Hence α is critical, and by the remarks above, 0-critical. $\square$

Now we map 0-critical assignments to certain clauses in the proof. For a 0-critical assignment α, let C_α be the first clause in R such that $\{ a \leq \frac{d}{2} \; ; \; q_{i_0,j_0,a} \text{ occurs in } C_\alpha \} = [\frac{d}{2}] \setminus \{a_1, \ldots, a_{\frac{d}{4}}\}$ and α does not satisfy C_α. This clause exists because α determines a path through R from $\bigvee_{1 \leq a \leq n} q_{i_0,j_0,a}$ to the empty clause such that α does not satisfy any clause on that path. The variables $q_{i_0,j_0,a}$ with $a \leq \frac{d}{2}$ are eliminated along that path, and $q_{i_0,j_0,a_1}, \ldots q_{i_0,j_0,a_{d/4}}$ are the first among them in the elimination order. The following lemma shows that the clauses C_α have a certain complexity, which implies that the mapping $\alpha \mapsto C_\alpha$ does not map too many 0-critical assignments to the same clause.

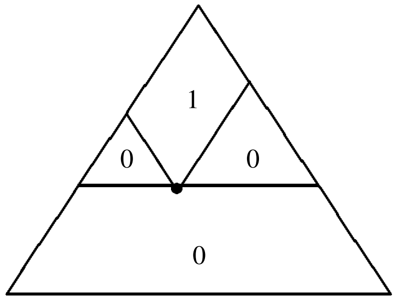

Figure 1. the black dot indicates (i_0, j_0).

Lemma 22 *Let α be a 0-critical assignment and $b_k := m_\alpha(i_k, j_k)$. Then for every $1 \leq k \leq \frac{d}{4}$, the literal $\bar{q}_{i_k,j_k,b_k}$ occurs in C_α.*

Proof: Let α' be the assignment defined by $\alpha'(q_{i_0,j_0,a_k}) := 1$ and $\alpha'(x) := \alpha(x)$ for all other variables x. As q_{i_0,j_0,a_k} does not occur in C_α, α' does not satisfy C_α either. If $i_0 < d$, the only clause from $DPGen(\vec{p}, \vec{q}) \cup Col(\vec{p}, \vec{r})$ that is not satisfied by α' is

$$\bar{q}_{i_k,j_k,b_k} \vee \bar{q}_{i_0+1,j_0,c_1} \vee \bar{q}_{i_0+1,j_0+1,c_2} \vee \bar{q}_{i_0,j_0,a_k} \vee p_{c_1,c_2,a_k}$$

where $c_1 := m_\alpha(i_0+1, j_0)$ and $c_2 := m_\alpha(i_0+1, j_0+1)$. If $i_0 = d$, then the only clause not satisfied by α' is

$$\bar{q}_{i_k,j_k,b_k} \vee \bar{q}_{i_0,j_0,a_k} \vee p_{1,1,a_k} \; .$$

The first item in the definition of 0-critical guarantees that these clauses are not satisfied, and the other two make sure that the other possible candidates, i.e. instances of (6) or (19) with (i_0, j_0) at the bottom of the triangle, are satisfied.

In both cases there is a path through R leading from the clause in question to C_α. The variable that is eliminated in the last inference on that path must be one of the q_{i_0,j_0,a_ℓ} for $1 \leq \ell \leq \frac{d}{4}$. Since $b_k \in R_k$, the variable q_{i_k,j_k,b_k} is later in the elimination order, so it cannot be eliminated on that path. Hence the literal $\bar{q}_{i_k,j_k,b_k}$ still occurs in C_α. $\square$

Now let α, β be two 0-critical assignments such that $b_k := m_\alpha(i_k, j_k) \neq m_\beta(i_k, j_k)$ for some $1 \leq k \leq \frac{d}{4}$, so that $\beta(q_{i_k,j_k,b_k}) = 0$. By Lemma 22, the literal $\bar{q}_{i_k,j_k,b_k}$ occurs in C_α, therefore β satisfies C_α and hence $C_\beta \neq C_\alpha$.

By Lemma 21, there are at least $\frac{d}{4}!$ distinct 0-critical assignments that differ in the values $m_\alpha(i_k, j_k)$. Thus R contains at least $\frac{d}{4}! \geq (\frac{d}{4e})^{\frac{d}{4}} = \Omega(2^{\frac{1}{4}n^{\frac{1}{3}}})$ different clauses of the form C_α, which proves the theorem. $\square$

Acknoweldgements

We would like to thank R. Raz for reading a previous version of this work and discovering an error, A. Goerdt for sending us copies of his papers, S. Buss for helpful discussions and finally P. Clote for suggestions about resolution separations.

References

[1] M. Alekhnovich, S. R. Buss, S. Moran, and T. Pitassi. Minimum propositional proof length is NP-hard to linearly approximate. In *MFCS '98*. Springer LNCS, 1998.

[2] P. Beame, R. Impagliazzo, J. Krajíček, T. Pitassi, and P. Pudlák. Lower bounds on Hilbert's Nullstellensatz and propositional proofs. *Proceedings of the London Mathematical Society*, 73:1–26, 1996.

[3] P. Beame and T. Pitassi. Simplified and improved resolution lower bounds. In *Proc. 28th STOC*, 1996.

[4] E. Ben-Sasson and A. Wigderson. Resolution refutation size vs. width. Manuscript, 1998.

[5] M. L. Bonet, C. Domingo, R. Gavalda, A. Maciel, and T. Pitassi. Non-automatizability of bounded-depth Frege proofs. Manuscript, 1998.

[6] M. L. Bonet, T. Pitassi, and R. Raz. Lower bounds for cutting planes proofs with small coefficients. *Journal of Symbolic Logic*, 62:708–728, 1997. Preliminary version in *Proc. 27th STOC*, 1995.

[7] M. L. Bonet, T. Pitassi, and R. Raz. No feasible interpolation for TC^0-Frege proofs. In *Proc. 38th FOCS*, pages 254–263, 1997.

[8] V. Chvátal and E. Szemerédi. Many hard examples for resolution. *Journal of the ACM*, 35:759–768, 1988.

[9] M. Clegg, J. Edmonds, and R. Impagliazzo. Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proc. 28th STOC*, pages 174–183, 1996.

[10] P. Clote and A. Setzer. On PHP, st-connectivity and odd charged graphs. In P. Beame and S. R. Buss, editors, *Proof Complexity and Feasible Arithmetics*, pages 93–117. AMS DIMACS Series Vol. 39, 1998.

[11] S. Cook and A. Haken. An exponential lower bound for the size of monotone real circuits. To appear in *J. Comp. System Sciences*, 1998.

[12] S. A. Cook and R. A. Reckhow. The relative efficiency of propositional proof systems. *Journal of Symbolic Logic*, 44:36–50, 1979.

[13] W. Cook, C. Coullard, and G. Turán. On the complexity of cutting plane proofs. *Discrete Applied Mathematics*, 18:25–38, 1987.

[14] A. Goerdt. Davis-Putnam resolution versus unrestricted resolution. *Annals of Mathematics and Artificial Intelligence*, 6:169–184, 1992.

[15] A. Goerdt. Unrestricted resolution versus N-resolution. *Theoretical Computer Science*, 93:159–167, 1992.

[16] A. Goerdt. Regular resolution versus unrestricted resolution. *SIAM Journal of Computing*, 22:661–683, 1993.

[17] A. Haken. The intractability of resolution. *Theoretical Computer Science*, 39:297–308, 1985.

[18] R. Impagliazzo, T. Pitassi, and A. Urquhart. Upper and lower bounds for tree-like cutting planes proofs. In *Proc. 9th LICS*, pages 220–228, 1994.

[19] K. Iwama. Complexity of finding short resolution proofs. In *MFCS '97*, pages 309–318. Springer LNCS 1295, 1997.

[20] J. Johannsen. Lower bounds for monotone real circuit depth and formula size and tree-like cutting planes. *Information Processing Letters*, 67:37–41, 1998.

[21] M. Karchmer and A. Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM Journal on Discrete Mathematics*, 3:255–265, 1990. Preliminary version in *Proc. 20th STOC*, 1988.

[22] J. Krajíček. Interpolation by a game. To appear in Math. Logic Quarterly, 1997.

[23] J. Krajíček. Interpolation theorems, lower bounds for proof systems and independence results for bounded arithmetic. *Journal of Symbolic Logic*, 62:457–486, 1997.

[24] J. Krajíček and P. Pudlák. Some consequences of cryptographical conjectures for S_2^1 and EF. *Information and Computation*, 140:82–94, 1998. Preliminary version in D. Leivant, ed., *LCC '94*, Springer LNCS 960, 1995.

[25] P. Pudlák. Lower bounds for resolution and cutting plane proofs and monotone computations. *Journal of Symbolic Logic*, 62:981–998, 1997.

[26] R. Raz and P. McKenzie. Separation of the monotone NC hierarchy. In *Proc. 38th FOCS*, pages 234–243, 1997.

[27] A. Rosenbloom. Monotone real circuits are more powerful than monotone boolean circuits. *Information Processing Letters*, 61:161–164, 1997.

[28] U. Schöning. *Logic for Computer Scientists*. Birkhäuser, 1989.

[29] A. Urquhart. Hard examples for resolution. *Journal of the ACM*, 34:209–219, 1987.

[30] A. Urquhart. The complexity of propositional proofs. *Bulletin of Symbolic Logic*, 1:425–467, 1995.