Electronic Colloquium on Computational Complexity, Report No. 138 (2014)

The space complexity of cutting planes refutations

Nicola Galesi* Pavel Pudlák† Neil Thapen†

October 29, 2014

Abstract

We study the space complexity of the *cutting planes* proof system, in which the lines in a proof are integral linear inequalities. We measure the space used by a refutation as the number of inequalities that need to be kept on a blackboard while verifying it. We show that any unsatisfiable set of inequalities has a cutting planes refutation in space five. This is in contrast to the weaker *resolution* proof system, for which the analogous space measure has been well-studied and many optimal lower bounds are known.

Motivated by this result we consider a natural restriction of cutting planes, in which all coefficients have size bounded by a constant. We show that there is a CNF which requires super-constant space to refute in this system. The system nevertheless already has an exponential speed-up over resolution with respect to size, and we additionally show that it is stronger than resolution with respect to space, by constructing constant-space cutting planes proofs of the pigeonhole principle with coefficients bounded by two.

We also consider *variable space* for cutting planes, where we count the number of instances of variables on the blackboard, and *total space*, where we count the total number of symbols.

1 Introduction

1.1 Background

The method of *cutting planes* for integer linear programming was introduced by Gomory [15] and Chvátal [10]. An initial polytope P, defined by a system of linear inequalities, can be transformed through a sequence of Gomory-Chvátal cuts into the integral hull of P, that is, into the smallest polytope containing the integral points of P. If the set of inequalities defining P has no integral solution,

*Computer Science Department, Sapienza University of Rome, via Salaria 113, 00198 Rome, Italy, galesi@di.uniroma.it. Part of this work was done while visiting the Institute of Mathematics of the Czech Academy of Sciences, partially supported by grant P202/12/G061 of GAČR.

†Institute of Mathematics, Czech Academy of Sciences, Žitná 25, 115 67 Prague 1, Czech Republic, {pudlak,thapen}@math.cas.cz. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement 339691. The Institute of Mathematics of the Czech Academy of Sciences is supported by RVO:67985840.

ISSN 1433-8092

then the integral hull of P is empty and the sequence of cuts can be used as a witness that there is no solution.

W. Cook et al. in [12] used this idea to define *cutting-plane* proofs. As we present it in this paper, *cutting planes*, or CP, is a system for refuting unsatisfiable systems of integral linear inequalities over Boolean (0/1-valued) variables. Each line in a CP refutation is an inequality, and there are rules for taking linear combinations and for a version of the Gomory-Chvátal cut (formal definitions follow in Section 1.4). In particular, CP can be used as a system for refuting unsatisfiable Boolean formulas in conjunctive normal form (CNFs), since these can be translated into sets of inequalities.

Cutting planes has been studied from the point of view of the *size* complexity of proofs, usually measured as the number of lines in a refutation. It has an exponential speed-up over the well-known *resolution* proof system [12]. Exponential lower bounds on size were shown in [17, 24].

By analogy with complexity theory, where we study the space needed by computations, as well as the time, we can also study the space requirements of proofs [14, 1]. In a refutational system based on successively deriving formulas, we imagine presenting a proof by writing formulas on a blackboard as we derive them. We can erase formulas and write down axioms at any time, but if we want to write a formula derived by a rule, all the premises of the rule must be present on the blackboard. How large a blackboard do we need? The most common measure of blackboard size is the number of formulas that will fit on it. This is called in general *formula space*, or *clause space* in resolution or *inequality space* in cutting planes. We also consider some other measures.

Space is by now fairly well-understood in resolution (see [23] for a survey) and increasingly also in the algebraic *polynomial calculus* proof system (see e.g. [6]). But little has been known about space in cutting planes. The basic space upper bounds known for resolution [14] carry over to CP, for example, that every unsatisfiable CNF has a refutation with linear space and quadratic total space. W. Cook in [11] showed that every unsatisfiable set of inequalities F has a refutation with total space polynomial in the space needed to write F (although his definitions are not quite the same as ours). A nontrivial lower bound for variable space in CP is mentioned as an open problem in [1]. Dantchev and Martin in [13] show lower bounds for a certain *width* measure. In a recent paper Göös and Pitassi [16], improving a result of Huynh and Nordström [20], give a family of CNFs of size m which cannot simultaneously be refuted with small space and small length — the space s and length ℓ of every CP refutation must satisfy $s \log \ell \geq m^{1/4-o(1)}$.

One motivation for studying cutting planes is that it has the potential to offer a more efficient foundation for SAT solving than resolution. From this point of view results about refutation size and refutation space are both interesting, as they may give information about respectively the time and the memory required for computations [22].

1.2 Results for cutting planes

Our main result, Theorem 5 in Section 2, is a general constant upper bound on the minimal inequality space of CP refutations: any unsatisfiable set of linear inequalities can be refuted in space five. This result, which holds in particular for unsatisfiable CNFs, is in contrast with resolution, where there are several families of CNFs, including random k-CNFs, which require refutations with linear clause space [14, 3] (the situation is similar with monomial space in polynomial calculus [6]). To prove the theorem we first prove that the *complete tree contradiction* CT_n has CP refutations in space five (Lemma 3), and then use these refutations to build small space refutations for any unsatisfiable set of inequalities.

Section 3 contains three small results that follow from the work in Section 2. First, we observe that the refutations in Lemma 3 use coefficients with absolute value at most 2^n. Hence the refutations have total space $O(n^2)$, where we measure *total space* by counting the total number of symbols that must be written simultaneously on the blackboard, not just the number of inequalities (we assume that the coefficients are written in binary, and do not consider variable names as taking space – see below). It follows that $O(n^2)$ total space is sufficient to refute any unsatisfiable set of linear inequalities, as long as the absolute values of the coefficients and the constant term are bounded by an exponential function (Corollary 6). Notice however that, restricted to CNFs, this upper bound already follows from the $O(n^2)$ upper bound for total space in resolution (see e.g. [14, 7]).

Second, we use our derivation of CT_n from any unsatisfiable set F of inequalities to observe, in Proposition 7, that F has a CP refutation in which the absolute values of the coefficients are relatively small — they are bounded by the maximum, over all inequalities I in F, of the sum of the absolute values of the coefficients and constant term of I. This gives smaller bounds than results in [12, 9]; however those are concerned with a different problem, of limiting the size of the coefficients while keeping the refutation short.

Lastly in Section 3 we consider *variable space* in CP. This measures the total number of instances of variables that appear simultaneously on the blackboard during a refutation. This is like total space, but ignores the size of the coefficients and constant terms. On the one hand, the minimal width of refuting an unsatisfiable CNF in resolution is a lower bound on the variable space in CP; on the other hand, Theorem 5 gives us a general linear upper bound on variable space. This allows us to use known width lower bounds in resolution to show tight linear bounds on variable space in CP (Theorem 9).

1.3 Results for cutting planes with small coefficients

The constant space refutations in Theorem 5 use coefficients as big as 2^n, and these seem to be necessary for our proof technique to work. In Sections 4 and 5 we study what can be said about space in CP if we rule out this kind of refutation, by putting an upper bound on the coefficients.

For $k \in \mathbb{N}$, we define CP^k as the restriction of cutting planes in which every inequality in a derivation must have coefficients with absolute value at most k. This is already quite a strong proof system for $k = 2$. It is exponentially stronger than resolution, since an inspection of the proofs in [12] shows that CP^2 efficiently simulates resolution and has polynomial size refutations of the pigeonhole principle PHP_m. Cutting planes with bounded coefficients has been considered before — the system *generalized resolution* studied in [19] is similar to CP^2, and size lower bounds for CP were initially shown for a restricted system CP^* with polynomially bounded coefficients [21, 8]. (Note that by a result of [18], if we bound the constant term[1] by k, rather than the coefficients, we get a system equivalent to resolution.)

In Section 4 we consider a natural candidate for proving inequality space lower bounds on CP^k refutations, the pigeonhole principle. We show in Theorem 10 that there is no such lower bound for PHP_m, and in fact that it has CP^2 refutations with inequality space five. Our refutation is broadly similar to the refutation in [12] (which uses space linear in the number of variables). It follows that CP^2 is strictly stronger than resolution with respect to space.

Finally in Section 5 we prove that small coefficients do not always suffice for constant space proofs, by showing in Theorem 17 that for any constant $k \in \mathbb{N}$, the contradiction CT_n requires inequality space $\Omega(\log^{(3)} n)$ to refute in CP^k. (In fact we prove something slightly stronger, that the refutation requires many *different* coefficients — our proof does not use the size of the coefficients directly.) Similarly, if we insist on constant inequality space then we get a barely super-constant lower bound on the coefficients. Our lower bounds are very small and surely not optimal. However, the proof is interesting because it is based on a counting argument, which is rare in proof complexity.

The contradiction CT_n is unusual in having exponential size in the number n of variables. However, using a padding argument one can easily show that there is contradiction F of linear size in n, and which even has linear size resolution refutations, but which still requires superconstant inequality space to refute in CP^k (Corollary 18). Nevertheless, it would be interesting to find a more natural example.

1.4 Technical preliminaries

The lines in a cutting planes (CP) proof are inequalities of the form $\sum \lambda_i x_i \geq t$ where the coefficients λ_i and the constant term t are integers, and the x_i are Boolean variables. A CP *derivation* of an inequality I from a set of inequalities F is a sequence of lines, ending with I, where each line is either (1) a member of F, or (2) a *Boolean axiom* $x \geq 0$ or $-x \geq -1$, or (3) follows from earlier lines by the *linear combination rule* or the *cut rule*. These are respectively

$$\frac{\sum \lambda_i^1 x_i \geq t_1 \qquad \cdots \qquad \sum \lambda_i^k x_i \geq t_k}{\sum \left(\sum_j s_j \lambda_i^j\right) x_i \geq \sum_j s_j t_j} \qquad \text{and} \qquad \frac{\sum s\lambda_i x_i \geq t}{\sum \lambda_i x_i \geq \lceil t/s \rceil}$$

where $s_1, \ldots, s_k$ and s must be strictly positive integers, and the linear combination rule can take any number of premises.[2]

To define our space measures we assume that our derivations come with some extra structure. We follow the model proposed by [14, 1] inspired by the definition of space for Turing machines. A *memory configuration* M is a set of linear inequalities. A CP *derivation of I from F* is then given by a sequence $M_0, \ldots, M_\ell$ of memory configurations, where M_i represents the contents of the

[1] More precisely, if we write all inequalities in the form $\sum_{i \in P} \lambda_i x_i + \sum_{i \in N} \lambda_i (1 - x_i) \geq t$ and put a constant upper bound on the term t.

[2] One can also define cutting planes using a binary addition rule, a unary multiplication rule and the cut rule. While the two systems polynomially simulate each other with respect to size, when considering questions about space or width they may differ substantially. We have chosen to use the linear combination rule since this captures better the geometric idea behind cutting planes. However our results, except for the discussion of width in Section 3, do not depend essentially on which definition one takes.

blackboard at the ith step in the derivation. The sequence must satisfy that M_0 is empty, that $I \in M_\ell$, and that for each $i < \ell$, M_{i+1} is obtained from M_i in one of three ways:

Axiom download: $M_{i+1} = M_i \cup \{J\}$ for some $J \in F$

Inference: $M_{i+1} = M_i \cup \{J\}$ where J follows from M_i by an inference rule, or is a Boolean axiom

Erasure: $M_{i+1} \subset M_i$.

A CP *refutation of* F is a CP derivation of $0 \geq 1$ from F.

We consider three measures of the space taken by a memory configuration M. The *inequality space* is the number of inequalities in M. The *variable space* is the sum, over all inequalities J in M, of the number of distinct variables appearing in J with a non-zero coefficient. We define the *total space* as the sum, over all inequalities J in M, of the length in binary of all non-zero coefficients in J and of the constant term of J (ignoring signs).[3]

For each measure, the corresponding space of a refutation Π is the *maximum* space of any configuration M_i in Π. The corresponding space needed to refute a set of inequalities F is the *minimum* space of any refutation of F.

If we refer just to the *space* of a refutation we mean the inequality space, just as in resolution the analogous measure, clause space, is often simply called space. The name *variable space* was introduced in [1] as a general measure of space complexity. In resolution there is no useful distinction between variable and total space, and the name *total space* has become standard (see e.g. [7]).

By an *assignment* to a set of inequalities or CNF F, we always mean a total assignment of 0/1 values to the variables appearing in F. We say that F is unsatisfiable if it is not satisfied by any such assignment.

The *complete tree contradiction* CT_n, which is central to this work, is a CNF in n variables $x_0, \ldots, x_{n-1}$, with 2^n clauses. For each assignment α, it contains the clause $\bigvee_{i \in Z} x_i \ \vee \bigvee_{i \in A} \neg x_i$ where $A = \{i : \alpha(x_i) = 1\}$ and $Z = \{i : \alpha(x_i) = 0\}$. This clause is falsified by α and by no other assignment.

We translate propositional clauses into inequalities, and thus CNFs into sets of inequalities, using the translation of [12]:

$$\bigvee_{i \in P} x_i \vee \bigvee_{i \in N} \neg x_i \quad \longmapsto \quad \sum_{i \in P} x_i + \sum_{i \in N} (1 - x_i) \geq 1.$$

When describing a CP refutation, we may freely rearrange the terms in an inequality and move the constant term around, for example treating $\sum \lambda_i x_i \geq t$ and $\sum \lambda_i x_i + s \geq t + s$ as the same inequality. Similarly, we will sometimes use the Boolean axiom $-x \geq -1$ in the form $1 - x \geq 0$.

When working in a fixed amount of inequality space, it is helpful to think of each unit of space as a "register" that can contain one inequality. We will frequently make use of the following observation, which we record as a lemma:

Lemma 1. *If we have one register free, we can treat addition, multiplication and rounding operations as if they happen "in place", with one of the assumptions overwritten by the conclusion. If we have two registers free, we can add any positive linear combination of axioms to any other register.* $\square$

[3]For simplicity, we do not count arithmetical symbols or variable names in total space. Counting these at most trebles the space, if we treat each variable name as a single symbol. It increases it by a factor of $O(\log n)$ if we include the symbols needed to write variable indices.

2 Inequality space upper bound

We show that any unsatisfiable set of inequalities F can be refuted in CP in constant inequality space. We do this by first showing that CT_n can be refuted in constant space, and then showing that each clause of CT_n can be derived from F in constant space. The overall form of the proof, and the idea of refuting CT_n by considering all assignments in lexicographic order, are inspired by the proof of a variable space upper bound on the Frege proof system in [1].

We first prove a useful lemma, then the upper bound for CT_n.

Lemma 2. *Suppose we have two registers free, and a third register that contains an inequality*

$$\sum_{i \in S} \lambda_i x_i + \sum_{i \in T} \lambda_i (1 - x_i) \geq b.$$

with $b \geq 1$. Then we can replace the inequality with

$$\sum_{i \in S} x_i + \sum_{i \in T} (1 - x_i) \geq 1.$$

Proof Choose an integer c greater than or equal to the maximum of b and all the coefficients λ_i. Using Lemma 1, add $(c - \lambda_i)x_i \geq 0$ to the inequality for each $i \in S$ and add $(c - \lambda_i)(1 - x_i) \geq 0$ for each $i \in T$. This gives

$$\sum_{i \in S} c x_i + \sum_{i \in T} c(1 - x_i) \geq b.$$

Then divide by c and round (by applying the cut rule). The constant term becomes $\lceil b/c \rceil = 1$. □

Lemma 3. *CT_n has a CP refutation with inequality space 5.*

Proof Given a number $a < 2^n$ we will write $(a)_0, \ldots, (a)_{n-1}$ for the bits of the binary expansion of a, so that $a = \sum 2^i (a)_i$. Throughout the proof sums $\sum$ are taken over $i < n$, or whichever subset of this is indicated.

For $a \in \mathbb{N}$, define the inequality T_a as

$$T_a : \sum 2^i x_i \geq a.$$

The assignments falsifying T_a are exactly those lexicographically strictly less than a. In other words, T_a is equivalent to the conjunction of the inequalities I_b over all $b < a$, where we write I_b for the clause of CT_n which is falsified exactly by the assignment $x_i \mapsto (b)_i$.

For $a < 2^n$, T_a and I_a together imply T_{a+1}. We will show that this implication can be proved in small space. In this way we can proceed by a kind of induction, first deriving T_0, then deriving in turn $T_1, T_2, \ldots, T_{2^n-1}$ and finally deriving a contradiction from T_{2^n-1} and I_{2^n-1}.

For the inductive step, fix $a < 2^n$. Let $A = \{i < n : (a)_i = 1\}$ and $Z = \{i < n : (a)_i = 0\}$. Define two inequalities

$$M_a : \sum_{i \in Z} x_i \geq 1 \qquad\qquad L_a^k : x_k + \sum_{\substack{i > k \\ i \in Z}} x_i \geq 1.$$

Notice that if β is an assignment such that $\beta \geq a$ lexicographically, then β satisfies L_a^k for each $k \in A$. If furthermore $\beta > a$, then β also satisfies M_a. We claim these implications are provable in small space:

Claim 1 We can derive M_a from T_a and I_a in space 3.

Claim 2 We can derive L_a^k from T_a in space 3, for any $k \in A$.

Using these two claims, we can then show

Claim 3 We can derive T_{a+1} from T_a and I_a in space 4.

This is enough to carry out the refutation sketched above, using five registers. The inequality T_0 is a linear combination of the axioms $x_i \geq 0$ so we may easily derive it in the first register. Then we derive T_1 using T_0, I_0 and the four free registers, then copy it to the first register. We repeat this for T_2, T_3 and so on. Once we have T_{2^n-1} we can derive M_{2^n-1}, which is exactly $0 \geq 1$.

It remains to prove the three claims.

Proof of Claim 1 We are given T_a, I_a and three free registers and want to derive M_a. We write I_a in the first register, that is,

$$\sum_{i \in Z} x_i + \sum_{i \in A} (1 - x_i) \geq 1.$$

We add to it the following two inequalities, both linear combinations of axioms:

$$\sum_{i \in Z} (2^i - 1) x_i \geq 0 \qquad \text{and} \qquad \sum_{i \in A} (2^i - 1)(1 - x_i) \geq 0.$$

The result is

$$\sum_{i \in Z} 2^i x_i - \sum_{i \in A} 2^i x_i \geq 1 - \sum_{i \in A} 2^i$$

whose right hand side equals $1 - a$. We add T_a to this, giving

$$2 \sum_{i \in Z} 2^i x_i \geq 1.$$

By Lemma 2 we can replace this with M_a.

Proof of Claim 2 We are given T_a and three free registers and want to derive L_a^k for a given $k \in A$. We copy T_a into the first register, rearranging it as

$$\sum_{i<k} 2^i x_i + 2^k x_k + \sum_{\substack{i>k \\ i \in A}} 2^i x_i + \sum_{\substack{i>k \\ i \in Z}} 2^i x_i \geq \sum_{\substack{i<k \\ i \in A}} 2^i + 2^k + \sum_{\substack{i>k \\ i \in A}} 2^i.$$

We add the following linear combination of axioms:

$$-\sum_{i<k} 2^i x_i - \sum_{\substack{i>k \\ i \in A}} 2^i x_i \geq -\sum_{i<k} 2^i - \sum_{\substack{i>k \\ i \in A}} 2^i.$$

The result is

$$2^k x_k + \sum_{\substack{i>k \\ i \in Z}} 2^i x_i \geq 2^k - \sum_{i<k} 2^i + \sum_{\substack{i<k \\ i \in A}} 2^i$$

whose right hand side is at least 1. Hence by Lemma 2 we can replace it with L_a^k.

Proof of Claim 3 We are given T_a, I_a and four free registers and want to derive T_{a+1}. By Claim 1, we can write M_a in the first register, that is,

$$\sum_{i \in Z} x_i \geq 1.$$

For each $k \in A$, we use Claim 2 to write L_a^k in the second register, and then multiply it by 2^k, giving

$$2^k x_k + 2^k \sum_{\substack{i>k \\ i \in Z}} x_i \geq 2^k.$$

We do this for each $k \in A$ in turn, each time adding the result to the first register. At the end of this process, the first register contains the inequality

$$\sum_{k \in A} 2^k x_k + \sum_{i \in Z} \Big(\sum_{\substack{k<i \\ k \in A}} 2^k \Big) x_i + \sum_{i \in Z} x_i \geq 1 + \sum_{k \in A} 2^k.$$

Here the right hand side equals $a+1$, and for $i \in Z$ the coefficient λ_i of x_i is less than or equal to 2^i. Hence for all $i \in Z$ we may add the inequality $(2^i - \lambda_i) x_i \geq 0$ to the first register, giving

$$\sum_{k \in A} 2^k x_k + \sum_{i \in Z} 2^i x_i \geq a+1$$

which is T_{a+1}. $\square$

Using the refutation constructed in Lemma 3 we first prove, in Theorem 4, a space upper bound for any unsatisfiable CNF. We then extend the argument to prove the more general result, an upper bound for any unsatisfiable set of inequalities, as Theorem 5.

Theorem 4. *Let F be any unsatisfiable CNF. Then F has a* CP *refutation with inequality space 5.*

Proof Suppose F has variables $x_0, \ldots, x_{n-1}$. It is enough to show that, for each assignment α, the inequality I_α of CT_n is derivable in space 4 from the translation of F. We can then imitate the refutation in the proof of Lemma 3.

Let α be any assignment and let $A = \{i : \alpha(x_i) = 1\}$ and $Z = \{i : \alpha(x_i) = 0\}$. Since F is unsatisfiable α falsifies some inequality from F, of the form

$$I : \sum_{i \in P} x_i + \sum_{i \in N} (1 - x_i) \geq 1.$$

Hence we must have $\alpha(x_i) = 0$ for each $i \in P$ and $\alpha(x_i) = 1$ for each $i \in N$. In other words, $P \subseteq Z$ and $N \subseteq A$. Hence we can derive I_α from F using space 3, by downloading I and adding

$$\sum_{i \in Z \setminus P} x_i + \sum_{i \in A \setminus N} (1 - x_i) \geq 0$$

which is a linear combination of axioms. $\square$

Theorem 5. *Let F be any set of unsatisfiable inequalities. Then F has a* CP *refutation with inequality space 5.*

Proof Suppose F has variables $x_0, \ldots, x_{n-1}$. As before, let α be any assignment and let $A = \{i : \alpha(x_i) = 1\}$ and $Z = \{i : \alpha(x_i) = 0\}$. The assignment α falsifies some inequality from F, of the form

$$I : \sum_{i \in P} \lambda_i x_i - \sum_{i \in N} \lambda_i x_i \geq t$$

where P and N are disjoint and all the coefficients λ_i are positive. We will derive I_α from I in space 3.

We first decompose I as

$$\sum_{i \in P \cap A} \lambda_i x_i + \sum_{i \in P \cap Z} \lambda_i x_i \; - \sum_{i \in N \cap A} \lambda_i x_i \; - \sum_{i \in N \cap Z} \lambda_i x_i \; \geq \; t. \tag{1}$$

Since I is falsified by α, if we evaluate the left-hand side of (1) under α we get

$$\sum_{i \in P \cap A} \lambda_i \; - \sum_{i \in N \cap A} \lambda_i \; < \; t.$$

Hence for some integer $\delta \geq 1$ we can rewrite (1) as

$$\sum_{i \in P \cap A} \lambda_i x_i + \sum_{i \in P \cap Z} \lambda_i x_i - \sum_{i \in N \cap A} \lambda_i x_i - \sum_{i \in N \cap Z} \lambda_i x_i \; \geq \; \sum_{i \in P \cap A} \lambda_i - \sum_{i \in N \cap A} \lambda_i + \delta. \tag{2}$$

We add to (2) the two inequalities

$$-\sum_{i \in P \cap A} \lambda_i x_i \; \geq \; -\sum_{i \in P \cap A} \lambda_i \qquad \text{and} \qquad \sum_{i \in N \cap Z} \lambda_i x_i \; \geq \; 0.$$

The result is

$$\sum_{i \in P \cap Z} \lambda_i x_i \; - \sum_{i \in N \cap A} \lambda_i x_i \; \geq \; -\sum_{i \in N \cap A} \lambda_i + \delta$$

which we rearrange as

$$\sum_{i \in P \cap Z} \lambda_i x_i \; + \sum_{i \in N \cap A} \lambda_i (1 - x_i) \; \geq \; \delta.$$

Since $\delta \geq 1$, we may use Lemma 2 to replace this with

$$\sum_{i \in P \cap Z} x_i \; + \sum_{i \in N \cap A} (1 - x_i) \; \geq \; 1$$

from which we can easily obtain I_α as in the previous theorem. □

3 Corollaries

Firstly, from the refutation constructed in Theorem 5, we immediately get a general upper bound on the total space needed for CP refutations. Note that there are threshold functions that require coefficients of size $n^{n/2}$ to write as a linear inequality, so the assumption about the coefficients in F is necessary.

Corollary 6. *Let F be any unsatisfiable set of linear inequalities over n variables in which the coefficients and the constant term are bounded by an exponential function $2^{O(n)}$. Then F has a* CP *refutation with total space $O(n)$ and with coefficients bounded by $2^{O(n)}$.*

Secondly, we observe that the reduction to CT_n at the end of Section 2 can be used directly to show an upper bound on the size of coefficients needed in a CP refutation.

Proposition 7. *Let F be any set of unsatisfiable inequalities. Let σ be the maximum, over all inequalities $\Sigma\lambda_i x_i \geq t$ in F, of $\Sigma|\lambda_i| + |t|$. Then there exists a* CP *refutation of F in which the absolute value of all coefficients is at most σ.*

Proof We use the constructions and notation from the proof of Theorem 5. We can derive from F all inequalities I_α of CT_n. Since these inequalities are translations of clauses, we can then simulate in CP the resolution refutation of CT_n. A simulation of resolution uses coefficients with absolute value at most 2. So it remains to check the size of the coefficients in the derivation of each I_α.

This is derived from a single inequality I in F, in two steps. First we obtain an inequality of the form

$$\sum_{i \in P \cap Z} \lambda_i x_i \; + \sum_{i \in N \cap A} \lambda_i (1 - x_i) \;\geq\; \delta \tag{3}$$

where all the λ_i are positive. The coefficients needed to derive this are just the coefficients from I. Furthermore $\delta = t - \sum_{i \in P \cap A} \lambda_i + \sum_{i \in N \cap A} \lambda_i$, so $|\delta| \leq \sigma$. We then reduce (3) to

$$\sum_{i \in P \cap Z} x_i \; + \sum_{i \in N \cap A} (1 - x_i) \;\geq\; 1 \tag{4}$$

as in Lemma 2, by letting $c = \max\{\lambda_1, \ldots, \lambda_n, \delta\}$, adding $(c - \lambda_i)x_i \geq 0$ to (3) for each $i \in P \cap Z$, adding $(c - \lambda_i)(1 - x_i) \geq 0$ to (3) for each $i \in N \cap A$, and then dividing by c and rounding. Since c and all the λ_i are positive, the largest coefficient that appears in this process is at most $\max\{|\lambda_1|, \ldots, |\lambda_n|, c\}$, which is bounded by σ.

From (4) we can get I_α using only coefficients ± 1. In fact, we do not even need this step, since (4) already is the translation of a clause, and the collection of all such clauses has a resolution refutation. □

Lastly we discuss bounds on variable space in CP. The *width* of a resolution refutation is the size of the largest clause in it. The next lemma is a simple special case of Lemma 8 of [2].

Lemma 8. *Let F be an unsatisfiable CNF. The minimal width of refuting F in resolution is at most the variable space of refuting F in* CP*.*

Proof In fact we will show that resolution width is at most the "variable space without repetitions" of refuting F in CP, where the space of a configuration is measured by counting the number of different variables that appear (this measure is called simply "variable space" in [4]).

Let Π be a CP refutation of F in which every configuration contains at most s many different variables with non-zero coefficients. We sketch how to simulate

Π by a resolution refutation ρ with width at most s. For any inequality I in Π, let X be the set of variables in I with non-zero coefficients, and let Φ_I be a CNF in variables X expressing the same Boolean function as I. Let $I_1, \ldots, I_m$ be the inequalities from which I was derived by a rule in Π. Then there is a resolution derivation of Φ_I from $\Phi_{I_1}, \ldots, \Phi_{I_m}$, since resolution is implicationally complete. The total number of different variables appearing in this derivation is at most s, since $I_1, \ldots, I_m$ and I must belong to the same configuration in Π, hence can mention no more than s variables in total. In particular, the width of the resolution derivation is at most s. □

The lemma allows us to use known lower bounds on width in resolution, together with the linear upper bound on variable space that follows immediately from Theorem 5, to derive tight bounds on variable space in CP. For example, using a result of [5], we get:

Theorem 9. *With high probability the variable space of refuting a random k-CNF in* CP *is $\Theta(n)$.* □

Note that if we had defined cutting planes using a binary addition rule and unary multiplication rule (rather than arbitrary linear combinations), the simulation in Lemma 8 would prove that resolution width is at most twice the CP width (if we define the width of an inequality as the number of variables appearing with non-zero coefficients). Clearly, in such a proof the particular form of the rules used is irrelevant; only their arity matters.

In the version of CP we use, it is not so easy to prove non-trivial width lower bounds. Dantchev and Martin in [13] show a width lower bound for an ordering principle in essentially this system, using a geometrical argument.

4 PHP_n with small coefficients

We consider the pigeonhole principle contradiction PHP_n. It is formalized, as usual, by the following set of inconsistent inequalities:

$$\begin{aligned} P_i: &\quad \sum_{j<n} x_{ij} \geq 1 \qquad \text{for } i < n+1 \\ H_{ii'j}: &\quad x_{ij} + x_{i'j} \leq 1 \quad \text{for } i < i' < n+1 \text{ and } j < n. \end{aligned}$$

To simplify our presentation we will be less strict about how we write inequalities in CP refutations, and allow the notation $\sum \lambda_i x_i \leq t$ (we do not change the formal rules of the system). With this notation the Boolean axioms look like $-x \leq 0$ and $x \leq 1$ and the cut rule looks like

$$\frac{\sum s\lambda_i x_i \leq t}{\sum \lambda_i x_i \leq \lfloor t/s \rfloor}$$

where we round the constant term down rather than up.

Theorem 10. PHP_n *has polynomial size* CP^2 *refutations with space 5.*

The non-trivial part of the proof is taken care of by the following lemma.

Lemma 11. *Given inequalities $y_i + y_j \leq 1$ for all $i < j < n$, we can derive $\sum y_i \leq 1$ in polynomial size and in space 4, using coefficients bounded by 2.*

Proof Let A_m be the inequality

$$A_m : \sum_{i<m} y_i \leq 1.$$

We claim that, for $m < n$, A_{m+1} can be derived from A_m in space 3. The lemma follows immediately.

So suppose we are given A_m, all inequalities $y_i + y_j \leq 1$, and three free registers. Our strategy is to derive the inequality

$$B_k : \sum_{i<k} y_i \ + \ y_m \leq 1 \tag{5}$$

in the first register, for $k = 1, \ldots, m-1$ in turn. For $k = 1$ this is an axiom, and for $k = m-1$ it is A_{m+1}, as required. Suppose we have derived B_k for some $1 \leq k < m-1$ and want to derive B_{k+1}. We add to B_k the inequalities

$$y_k + y_m \leq 1 \qquad \text{and} \qquad \sum_{i<k+1} y_i \leq 1. \tag{6}$$

The first of these is an axiom. The second is a weakening of A_m, which we could derive in three registers by downloading A_m and then adding the combination of Boolean axioms $-y_{k+1} - \cdots - y_{m-1} \leq 0$. However, since we only have two registers free, we achieve the same effect by adding A_m and $-y_{k+1} - \cdots - y_{m-1} \leq 0$ directly to the first register. The result is

$$\sum_{i<k} 2y_i \ + \ 2y_k \ + \ 2y_m \leq 3$$

since each index appears exactly twice in the three inequalities from (5) and (6). We derive B_{k+1} by dividing by two and rounding down the constant term. $\square$

Proof of Theorem 10 We are given the PHP_n axioms and five free registers. We use Lemma 11 and the first four registers to derive

$$\sum_{i<n+1} x_{ij} \leq 1$$

for each $j < n$ in turn, each time adding the result to the fifth register. The fifth register then contains the total

$$\sum_{j<n} \sum_{i<n+1} x_{ij} \leq n, \quad \text{or equivalently} \quad -\sum_{i<n+1} \sum_{j<n} x_{ij} \geq -n.$$

We obtain $0 \geq 1$ by adding to this the axioms P_i for all $i < n+1$. $\square$

5 Space lower bounds for small coefficients

We use a counting argument to show that any CP refutation of CT_n, in which there is a global constant bound on the number of different coefficients appearing in every configuration, must have superconstant inequality space. In particular, this implies superconstant lower bounds on inequality space for CT_n in the system CP^k.

Definition 12. *Call a set A of assignments* s-symmetric *if there is a partition of the variables into s or fewer blocks, such that A is closed under every permutation which preserves all blocks.*

Lemma 13. *Suppose I is a linear inequality in which no more than b different coefficients appear. Then the set of assignments falsifying I is b-symmetric.*

Suppose M is a CP *configuration in space c, such that no more than b different coefficients appear in any inequality in M. Then the set of assignments falsifying M is b^c-symmetric.*

Proof For the first part, the inequality I has the form

$$\lambda_1 \sum_{i \in B_1} x_i + \cdots + \lambda_b \sum_{i \in B_b} x_i \geq t.$$

The b-symmetry is witnessed by the blocks $B_1, \ldots, B_b$. For the second part, take the common refinement of the partitions for all of the inequalities in M. $\square$

Lemma 14. *Suppose that* CT_n *has a* CP *refutation in space c, in which no more than b different coefficients appear in any inequality. Then there is a sequence $A_1, \ldots, A_N$ of sets of b^c-symmetric assignments, beginning with the empty set and ending with the set of all assignments, such that for each $i < N$ either $A_{i+1} \subseteq A_i$ or $A_{i+1} = A_i \cup \{\alpha\}$ for some assignment α.*

Proof Let A_i be the set of assignments falsifying the ith configuration. $\square$

We define a *k-assignment* to be an assignment with exactly k variables set to 1 and all the rest set to 0.

Lemma 15. *Define $S(s,k) = \{|A| : A$ is an s-symmetric set of k-assignments$\}$. Then $|S(s,k)| < n^s 2^{k^s}$.*

This is proved after Theorem 16.

Theorem 16. *For $n \geq 2$, suppose that* CT_n *has a* CP *refutation in space c, in which no more than b different coefficients appear in any inequality. Then $b^c \geq \sqrt{\log \log n}$.*

Proof Let $s = b^c$ and $k = 2^s$. For trivial reasons $b, c \geq 2$ so $s \geq 4$.

Let $A_1, \ldots, A_N$ be the sequence of s-symmetric assignments from Lemma 14, and let $A'_i = \{\alpha \in A_i : \alpha \text{ is a } k\text{-assignment}\}$. Then A'_1 is empty, A'_N consists of all k-assignments, and for each $i < N$ either $A'_{i+1} \subseteq A'_i$ or $A'_{i+1} = A'_i \cup \{\alpha\}$ for some k-assignment α. It follows that the sequence $|A'_1|, \ldots, |A'_N|$ must contain every number between 0 and $\binom{n}{k}$. Since each A'_i is still s-symmetric, this in particular means that for every number m between 0 and $\binom{n}{k}$, there is at least one s-symmetric set A of k-assignments with $|A| = m$.

Hence, in the notation of Lemma 15, $S(s,k) = \binom{n}{k} + 1$. It follows that $\binom{n}{k} < n^s 2^{k^s}$. Using the bound $(n/k)^k \leq \binom{n}{k}$ and taking the logarithm of both sides, we get

$$k(\log n - \log k) < s \log n + k^s.$$

Substituting $k = 2^s$ gives

$$2^s(\log n - s) < s \log n + 2^{s^2}.$$

Now assume for a contradiction that $s < \sqrt{\log\log n}$. Then $\log n - s \geq \frac{1}{2}\log n$ (we may assume $n \geq 4$) and $2^{s^2} < \log n$. The inequality becomes

$$2^{s-1}\log n < (s+1)\log n$$

which is impossible. $\square$

Proof of Lemma 15 Let A be an s-symmetric set of k-assignments. Let $B_1, \ldots, B_s$ be a partition witnessing the s-symmetry (we allow some of the blocks to be empty). Then A is the union of orbits, where each orbit is parametrized by a distinct tuple $r_1, ..., r_s$ summing to k, and the orbit consists of every k-assignment which has exactly r_i many ones in each block B_i. Let $n_i = |B_i|$. Then

$$|A| = \sum_{j=1}^{m} \binom{n_1}{r_1^j} \cdot \ldots \cdot \binom{n_s}{r_s^j}$$

where there are m orbits and the jth orbit has parameters $\bar{r}^j = r_1^j, ..., r_s^j$. In particular, $|A|$ depends only on the sizes $n_1, \ldots, n_s$ and on the set of tuples $\{\bar{r}^1, \ldots, \bar{r}^m\}$ characterizing the set of orbits.

There are no more than n^s ways to choose $n_1, \ldots, n_s$. There are no more than k^s ways to choose the parameters $\bar{r}$ for an orbit, and therefore there are no more than 2^{k^s} possible sets of such parameters. Therefore there are at most $n^s 2^{k^s}$ possible values for $|A|$. $\square$

From Theorem 16 we immediately get:

Theorem 17. *For any constant $k \in \mathbb{N}$, the* complete tree contradiction CT_n *requires inequality space $\Omega(\log^{(3)} n)$ to refute in* CP^k.

Corollary 18. *There is a family of propositional CNFs F in n variables, with linear size and with linear sized resolution refutations, which require superconstant inequality space to refute in* CP^k *for any fixed $k \in \mathbb{N}$.*

Proof Let $m = \log n$, and let F be CT_m together with $2^m - m$ inequalities of the form $y_i \geq 1$ in variables $y_1, \ldots, y_{2^m - m}$ disjoint from the variables in CT_m. Then F has a resolution refutation of linear size, since CT_m has a refutation of size 2^m, and any constant-space CP^2 refutation of F can be made into a constant-space CP^2 refutation of CT_m by substituting 1 for all variables y_i. $\square$

We note that, as in Section 3, our lower bound relies only on the class of Boolean functions appearing as lines in the refutation, not on the particular rules used.

6 Open problems

There are many problems about cutting planes that are worth mentioning, but we confine ourselves to a small sample, directly connected with the results presented in this paper.

The first general problem is about the trade-off between inequality space and the size of coefficients. Our upper bound uses coefficients of exponential size, while we can only prove that if space is constant then coefficients can be lower-bounded by a very slowly growing function. In particular the following is open:

Problem 1. Can every unsatisfiable CNF be refuted in CP in constant space, if the coefficients are polynomially bounded?

A related open problem is:

Problem 2. Can every unsatisfiable CNF be refuted in CP in linear total space?

It seem plausible that some extension of the proof of Theorem 17 might work also for such a lower bound.

Among the restricted systems of CP, the system CP^2 stands out as already being strong enough to simulate resolution and to capture some of the counting available in CP, since it has efficient proofs of PHP_n. It would be interesting to improve our results at least for this system. In particular:

Problem 3. Prove a better space lower bound for CP^2.

References

[1] Michael Alekhnovich, Eli Ben-Sasson, Alexander A Razborov, and Avi Wigderson. Space complexity in propositional calculus. *SIAM Journal on Computing*, 31(4):1184–1211, 2002.

[2] Chris Beck, Jakob Nordstrom, and Bangsheng Tang. Some trade-off results for polynomial calculus. In *Proceedings of the 45th Annual Symposium on the Theory of Computing (STOC 2013)*, pages 813–822, 2013.

[3] Eli Ben-Sasson and Nicola Galesi. Space complexity of random formulae in resolution. *Random Structures and Algorithms*, 23(1):92–109, 2003.

[4] Eli Ben-Sasson and Jakob Nordström. Understanding space in proof complexity: separations and trade-offs via substitutions. In *Proceedings of the 2nd Symposium on Innovations in Computer Science (ICS '11)*, pages 401–416, 2011.

[5] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow – resolution made simple. *Journal of the ACM*, 48(2):149–169, 2001.

[6] Ilario Bonacina and Nicola Galesi. Pseudo-partitions, transversality and locality: a combinatorial characterization for the space measure in algebraic proof systems. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science (ITCS '13)*, pages 455–472, 2013.

[7] Ilario Bonacina, Nicola Galesi, and Neil Thapen. Total space in resolution. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:38, 2014. To appear in FOCS 2014.

[8] Maria Bonet, Toniann Pitassi, and Ran Raz. Lower bounds for cutting planes proofs with small coefficients. *Journal of Symbolic Logic*, 62(03):708–728, 1997.

[9] Samuel R Buss and Peter Clote. Cutting planes, connectivity, and threshold logic. *Archive for Mathematical Logic*, 35(1):33–62, 1996.

[10] Vašek Chvátal. Edmonds polytopes and a hierarchy of combinatorial problems. *Discrete Mathematics*, 4(4):305–337, 1973.

[11] William Cook. Cutting-plane proofs in polynomial space. *Mathematical Programming*, 47(1-3):11–18, 1990.

[12] William Cook, Collette R Coullard, and Gy Turán. On the complexity of cutting-plane proofs. *Discrete Applied Mathematics*, 18(1):25–38, 1987.

[13] Stefan Dantchev and Barnaby Martin. Cutting planes and the parameter cutwidth. *Theory of Computing Systems*, 51(1):50–64, 2012.

[14] Juan Luis Esteban and Jacobo Torán. Space bounds for resolution. *Information and Computation*, 171(1):84–97, 2001.

[15] Ralph E Gomory. An algorithm for integer solutions to linear programs. *Recent Advances in Mathematical Programming*, 64:260–302, 1963.

[16] Mika Göös and Toniann Pitassi. Communication lower bounds via critical block sensitivity. In *Proceedings of the 46th Annual Symposium on the Theory of Computing (STOC 2014)*, pages 847–856, 2014.

[17] Armin Haken and Stephen A Cook. An exponential lower bound for the size of monotone real circuits. *Journal of Computer and System Sciences*, 58(2):326–335, 1999.

[18] Edward A Hirsch and Sergey I Nikolenko. Simulating cutting plane proofs with restricted degree of falsity by resolution. In *Theory and Applications of Satisfiability Testing*, pages 135–142. Springer, 2005.

[19] John N Hooker. Generalized resolution for 0–1 linear inequalities. *Annals of Mathematics and Artificial Intelligence*, 6(1-3):271–286, 1992.

[20] Trinh Huynh and Jakob Nordström. On the virtue of succinct proofs: amplifying communication complexity hardness to time-space trade-offs in proof complexity. In *Proceedings of the 44th Annual Symposium on the Theory of Computing (STOC 2012)*, pages 233–248, 2012.

[21] Russell Impagliazzo, Toniann Pitassi, and Alasdair Urquhart. Upper and lower bounds for tree-like cutting planes proofs. In *Proceedings of the 9th Annual IEEE Symposium on Logic in Computer Science (LICS '94)*, pages 220–228, 1994.

[22] Jakob Nordström. A (biased) proof complexity survey for SAT practitioners. Talk at the 17th International Conference on Theory and Applications of Satisfiability Testing (SAT 2014), Vienna, July 2014.

[23] Jakob Nordström. Pebble games, proof complexity, and time-space tradeoffs. *Logical Methods in Computer Science*, 9(3):15, 2013.

[24] Pavel Pudlák. Lower bounds for resolution and cutting plane proofs and monotone computations. *Journal of Symbolic Logic*, 62(03):981–998, 1997.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de