On the Width of Semi-Algebraic Proofs and Algorithms

Alexander Razborov*

November 5, 2016

Abstract

In this paper we study width of semi-algebraic proof systems and various cut-based procedures in integer programming. We focus on two important systems: Gomory-Chvátal cutting planes and Lovász-Schrijver lift-and-project procedures. We develop general methods for proving width lower bounds and apply them to random k-CNFs and several popular combinatorial principles like the perfect matching principle and Tseitin tautologies. We also show how to apply our methods to various combinatorial optimization problems. We establish a "supercritical" tradeoff between width and rank, that is give an example in which small width proofs are possible but require *exponentially* many rounds to perform them.

1. Introduction

The duality between proof complexity of semi-algebraic proof systems and systematic approach to integer programming based on various LP and SDP hierarchies has been extremely fruitful and beneficial for both areas as it allows to bring together two a priori different communities working on the same kind of problems with slightly different perspectives. The interest in this subject is also re-kindled by the realization that systems like Sum-of-Squares,

*University of Chicago, razborov@math.uchicago.edu; part of this work was done at Steklov Mathematical Institute, Moscow, and at Toyota Technological Institute, Chicago. Supported by the Russian Science Foundation, grant # 14-50-00005.

previously known under the names Positivestellensatz [21] and Lasserre hierarchy [25], are in the midst of many exciting developments in combinatorial optimization and many other areas, see [5] and the references therein, and, more recently, [26].

Philosophically, the connection is very simple and is similar to the connection between SAT solving procedures and resolution proofs, only with logic replaced by geometry. The "cut phase" of a branch-and-cut algorithm for integer programming[1] consists of shrinking the polytope[2] of potential solutions using various cuts until it approaches the positive hull made by all integer solutions. In the dual, proof-complexity, world it can be alternately viewed as generating more valid constraints following a set of prescribed rules. The transcript of this execution makes a mathematical proof of the resulting bound on the goal function or, depending on the context, of infeasibility of the original problem. *Semi-algebraic proof systems* precisely capture the amount of geometric reasoning necessary to verify the validity of the transcript. As in many previous papers on the subject we confine ourselves to two most popular systems: cutting planes[3] proofs introduced by Cook et al. [14] as a way to capture Gomory-Chvátal mixed integer cuts [20, 12] and Lovász-Schrijver procedures [28] capturing various lift-and-project methods.

Given these parallel developments, it is no wonder that many fundamental concepts and ideas of theoretical proof complexity have their natural counterparts in combinatorial optimization, and vice versa. We highly recommend Jukna's detailed account [24, Chapter 19] of the story for the case of cutting planes. One notable exception (to the best of our knowledge), however, is the notion of *width* of resolution proofs; it measures how many literals a clause in a prospective resolution proof is allowed to contain. Its importance has been increasingly realized since the seminal paper [7] by Ben-Sasson and Wigderson, and by now width-based lower bound methods is among the most important tools for analyzing complexity of resolution proofs. More precisely,

[1] If we also bring into the picture the branching steps, we will arrive at more sophisticated proof systems mixing logic and geometry. Those are not considered in this paper; for more information on the subject see e.g. [15].

[2] In this paper we primarily consider algorithms/proof systems operating with *linear* constraints; see the concluding section 8 for a brief discussion of potential generalizations to higher-degree polynomials.

[3] Many authors use this term more generally, for collectively denoting cuts of all possible kind including those that are non-linear in their nature. In this paper we prefer to reserve this term for its original meaning.

the relation discovered in [7] says that

$$w(\tau_n \vdash 0) \leq O(n \log S(\tau_n \vdash 0))^{1/2} + w(\tau_n), \tag{1}$$

where $w(\tau_n \vdash 0)$ is the minimum width of a resolution refutation of an unsatisfiable CNF τ_n in n variables, $S(\tau_n \vdash 0)$ is the minimum size of such a refutation, and $w(\tau_n)$ is the maximum width of a clause C in τ_n itself. Hence strong width lower bounds imply exponential lower bounds on the *size* of resolution proofs of the same principle.

Width, however, is seldom mentioned for proof systems other than resolution. The reason for this is quite sound and simple: width is a fundamentally *semantical* measure and, as a consequence, it is extremely robust with respect to the choice of a particular proof system. More technically, once we care only about the number of variables in a constraint rather than its logical complexity, we can always expand arbitrarily fancy formulas as CNFs and simulate the original reasoning by a resolution proof of the same width, as long as the former is sound.

In this paper we attempt to argue that dynamic semi-algebraic proof systems should be exempted from this rule. Let us immediately come straight to the point and give a simple and somewhat extremal example illustrating why we think so.

Example 1 Let G be a bipartite graph of bounded degree with parts U and V such that $|U| > |V|$. Let $G - PHP$ be the principle asserting that G does not contain a matching from V to U that covers all vertices in U. Then if G has good expansion properties, this principle does not possess any sub-linear width proofs [7, Theorem 4.15]. But from the point of combinatorial optimization it is completely trivial: summing up the constraints expressing that all vertices in U are covered at least once, and deducting the sum of constraints expressing that all $v \in V$ are covered at most once, we see already that the initial polytope P is empty so there is nothing remaining to prove.

This example highlights the main difference between the two views. The proof complexity community highly prefers proof systems with *binary* or at least bounded fan-in rules, and often for good reasons. While in the combinatorial optimization community primarily working with dual objects in a closed form, the idea of convexity is so basic that splitting the convex combination rule into a binary tree of additions with two operands each looks

quite arbitrary. Moreover, Caratheodory's theorem makes this difference incremental or unimportant when we are interested in rank or size; we again refer to [24, Chapter 19] that very carefully elaborates on the issue. As we have just seen in Example 1, it makes all the difference if we are interested in width, and, as we hope to convince the reader, this leads to an interesting and nice model non-trivially extending the notion of width as it is known in logic-based proof complexity. To the best of our knowledge, even in the combinatorial optimization community this measure was systematically studied, under the name of "sparsity", only for rank one (i.e., non-iterative) cutting planes procedures [16].

As follows from the above discussion, in this paper we regard taking arbitrary convex combinations of the already deduced cuts as a primary and relatively inexpensive step. Then we define the width of a proof or an algorithm simply as the maximum number of variables involved in the cuts it makes. Besides natural curiosity and importance of width in the logic-based proof complexity, we may have yet another (unspoken) reason to be interested in this model.

Most interesting theoretical results about LP/SDP hierarchies pertain to *rank*, that is the minimum number of *iterations* one needs to shrink the original polytope so that it approaches the positive hull of its integer points. During one round we perform in parallel *all* possible cuts that immediately results in the exponential blow-up in the number of constraints and becomes prohibitive very soon. On the other side of the spectrum we have numerous examples of algorithms for specific problems based on linear or SDP relaxations that essentially succeed by finding a golden needle in this stack of hay (see e.g. the paper [1] that collected several prominent examples). Therefore, *length* of semi-algebraic proofs or algorithms defined, say, as the *number* of cuts used appears at least as important as rank, both theoretically and practically. Our current understanding of this measure, however, is quite miserable: the only existing methods are manifestly indirect and are based on the so-called "feasible interpolation theorem" [32, 33]. Developing combinatorial or geometric approaches for this task is one of the most prominent and difficult problems in modern proof complexity, and we would like to express a cautious hope that our methods and concepts might turn out useful here.

In another, more practical direction, what seems to be somewhat underdeveloped are *perhaps imperfect* but still *sufficiently general* heuristics for classifying available cuts into "useful" (to be kept) and "useless" (to be dis-

carded). This is in sharp contrast with the situation in the adjacent community of practical (logic-based) SAT solving that can rightly boast heuristics of such kind as their success story. The idea of identifying "useful" cuts with "local" or "myopic" cuts that involve only a few variables is clearly insufficient for most, if not all, applications. However, it is a very natural thing to try, and we (again, cautiously) hope that understanding in precise mathematical terms why and how exactly it fails might be helpful for promoting further research in this important direction.

Last, but not the least, it appears as if the width (= sparsity) already *is* an important measure of concern for real-world, commercial MILP solvers. The author, however, certainly does not feel qualified to further delve into the issue and refers instead the interested reader to [16] and the literature cited therein.

1.1. Results and organization of the paper

In Section 2 we recall some necessary preliminaries and present our main definitions and ideas. As a part of this exposition, we also prove that the hierarchy of bounded-width proofs converges within finitely many steps (Theorem 2.9), and formulate a strong tradeoff result stating that the crude exponential bound on the number of rounds in Theorem 2.9 can in fact sometimes be nearly optimal (Theorem 2.12). Then we introduce w-obstructing polytopes as our main (and universal) technical tool for proving lower bounds on width in semi-algebraic proof systems (Definition 2.15). In Section 3 we give a general recipe for actually constructing such polytopes (Theorem 3.4). These polytopes can be viewed as a hybrid of protection matrices widely used for rank lower bounds in semi-algebraic proof systems and formal complexity measures employed for width lower bounds in logic-based proof complexity.

In Section 4 we give several applications of this technique to a few prominent combinatorial principles; we are not aiming at a comprehensive list. We start with expansion-based lower bounds on semi-algebraic width for systems of $\mathbb{F}_2$-linear equations (Theorem 4.3). As usual, this almost immediately implies lower bounds for random 3-CNFs (Theorem 4.9), as well as for Tseitin tautologies (Theorem 4.12) and for random 3-XOR formulas (Theorem 4.8). Next, we look at perfect matching principles (Section 4.2). We show that for any fixed even $d \geq 14$ and $n \to \infty$, proving that a random d-regular graph does not contain a perfect matching requires linear width cutting planes proofs (Theorem 4.14). Besides the fact that the (perfect) matching poly-

tope is one of the most cherished polytopes in the combinatorial optimization literature, this result is also interesting since [23] showed that in *every* graph the absence of a perfect matching can be verified by a cutting plane proof (actually, even a tree-like proof) of polynomial size. Together, these two facts imply that no useful analogue of the width-size relation (1) may exist for semi-algebraic proof systems.

In the next section 5 we address the case of combinatorial optimization on feasible (that is, containing integer points) polytopes. As it turns out, the mere presence of integer points makes the task of constructing w-obstructing polytopes much easier even if for understandable reasons they cannot be used in them per se. As an illustration, we give relatively easy lower bounds on the integrality gaps for VERTEX COVER, MAX CUT and MAX SAT by essentially reducing them to known results about rank.

In Section 6 we prove our tradeoff result, Theorem 2.12. We use for the purpose the same CNFs that were used in [34] to prove a proportionally strong tradeoff between width and tree-like resolution proof size. The reasoning, though, is rather different: as an indicator, let us mention that we were not able to combine the two results into one unifying statement (see Section 8 for more details).

In Section 7 we prove technical lemmas left over from Section 4.

The paper is concluded with a brief discussion and some open problems in Section 8.

2. Preliminaries

Having discussed connections with propositional proof complexity at length, in the technical part of the paper we try to stick to the language of combinatorial optimization (that is, of polytopes and convex bodies) as much as possible. Also, we confine ourselves to the case of 0-1 integer programming: firstly, because this is the most interesting case, and, secondly, because most of our notions can be straightforwardly extended, if desired, to integer programs over $\mathbb{Z}$.

We let $[n] \stackrel{\text{def}}{=} \{1, 2, \ldots, n\}$. $[n]^w$ and $[n]^{\leq w}$ stand for the family of all subsets J of $[n]$ that have cardinality w and $\leq w$, respectively (many authors denote these by $\binom{[n]}{w}$ and $\binom{[n]}{\leq w}$).

Let $P \subseteq [0,1]^n$ be a convex body (most often, P will be a polytope

but we need this more general case to properly treat the Lovász-Schrijver hierarchy). We denote by $P_{0-1} \subseteq P$ the convex hull of all $\{0,1\}$-points in P. With a slight abuse of terminology, let us call P *integral* if $P_{0-1} = P$, *feasible* if $P_{0-1} \neq \emptyset$ (equivalently, $P \cap \{0,1\}^n \neq \emptyset$) and *infeasible* if $P_{0-1} = \emptyset$. For $x \in [0,1]^n$, $E(x)$ denotes the set $\{\, j \in [n] \mid 0 < x_j < 1 \,\}$. For $J \subseteq [n]$, x is *J-integral* if $\forall j \in J(x_j \in \{0,1\})$ or, equivalently, $E(x) \cap J = \emptyset$.

Definition 2.1 (Gomory-Chvátal cuts) If $c^T x \leq d$ is a linear inequality satisfied by all points in a convex body P, where $c \in \mathbb{Z}^n$ and $d \in \mathbb{R}$, then the inequality $c^T x \leq \lfloor d \rfloor$ is called a *Gomory-Chvatál cut* for P. The *Chvátal closure* of P is the convex body that consists of all points satisfying every Gomory-Chvatál cut for P; it is denoted by P'. This operation is clearly monotone: if $P \subseteq Q$ then $P' \subseteq Q'$.

Proposition 2.2 ([12, 37]) *If P is a polytope then P' is also a polytope. Defining the decreasing chain*

$$P \stackrel{\text{def}}{=} P^{(0)} \supseteq P^{(1)} \supseteq \ldots \supseteq P^{(r)} \supseteq \ldots$$

by letting $P^{(\ell+1)} \stackrel{\text{def}}{=} \left(P^{(\ell)}\right)'$, there exists an integer $r \geq 0$ such that $P^{(r)} = P_{0-1}$.

In the infeasible case, it is easy to see that one can actually take $r = n$, simply because cutting planes simulate resolution (see Fact 2.11 below). The feasible case is substantially more complicated; the best known polynomial upper bound

$$r \leq n^2(1 + \log_2 n) \tag{2}$$

was proved by Eisenbrand and Shulz [18].

Definition 2.3 ([28]) A linear inequality $c^T x \leq d$ is called an *N_+-cut* for $P \subseteq [0,1]^n$ if the linear form $d - c^T x$ allows a representation of the form

$$\left\{ \begin{array}{l} d - c^T x = \\ \quad \sum_{j=1}^n \left\{ \left(b_{j1} - a_{j1}^T x\right) x_j + \left(b_{j0} - a_{j0}^T x\right)(1 - x_j) + \lambda_j (x_j^2 - x_j) \right\} \\ \quad\quad + \sum_{k=1}^m f_k^2, \end{array} \right. \tag{3}$$

where the constraints $a_{j\epsilon}^T x \leq b_{j\epsilon}$ ($j \in [n], \epsilon \in \{0,1\}$) are satisfied by all points in P, λ_j are arbitrary reals and f_k are arbitrary linear functions. $N_+(P)$ is the convex body comprised of all points that satisfy every N_+-cut. This operation is also monotone in P.

The analogue of Proposition 2.2 for the Lovász-Schrijver hierarchy is much simpler even in the feasible case. Let the operator $N(P)$ be defined analogously to $N_+(P)$ except that in (3) we disallow the last term $\sum_{k=1}^m f_k^2$.

Proposition 2.4 ([28]) *Defining recursively* $N^0(P) \stackrel{\text{def}}{=} P$ *and* $N^{\ell+1}(P) \stackrel{\text{def}}{=} N(N^\ell(P))$*, for any convex body* $P \subseteq [0,1]^n$ *we have* $N^n[P] = P_{0-1}$.

The main technical (and, in fact, universal) tool for proving lower bounds for N_+-based hierarchies are so-called *protection lemmas*. We will only need the most basic one:

Proposition 2.5 ([19]) *Let* $P \subseteq [0,1]^n$ *be a convex body, and let* $x \in P$. *For* $j \in [n]$ *and* $\epsilon \in \{0,1\}$, *let* $x^{(j,\epsilon)}$ *be obtained from* x *by resetting* $x_j := \epsilon$.

Assume that for every $j \in E(x)$ *and every* $\epsilon \in \{0,1\}$, $x^{(j,\epsilon)} \in P$. *Then* $x \in N_+(P)$.

Let us now gradually proceed to our framework; as we noted in Introduction, in the non-iterative form it was already considered in [16] (and several other papers).

Definition 2.6 (width-restricted hierarchies) By a *cut for* P we will sometimes collectively mean either a Gomory-Chvátal cut or an N_+-cut. The *width* of a cut $c^T x \leq d$ is the number of non-zero entries in the vector c. For $w \leq n$, let $P'^{,w}$ and $N_{+,w}(P)$ be defined similarly to P' and $N_+(P)$, except that we restrict them to cuts of width $\leq w$. Similarly to Propositions 2.2, 2.4, let us define hierarchies

$$P = P^{(0,w)} \supseteq P^{(1,w)} \supseteq \ldots \supseteq P^{(r,w)} \supseteq \ldots \tag{4}$$

$$P = N^0_{+,w}(P) \supseteq N'_{+,w}(P) \supseteq \ldots \supseteq N^r_{+,w}(P) \supseteq \ldots \tag{5}$$

by letting $P^{(\ell+1,w)} \stackrel{\text{def}}{=} \left(P^{(\ell,w)}\right)'^{,w}$ and $N^{\ell+1}_{+,w} \stackrel{\text{def}}{=} N_{+,w}(N^\ell_{+,w}(P))$.

We might have also considered the mixed hierarchy allowing both kinds of cuts. But since our primary goal will be to study the minimum width w for which these hierarchies converge or "nearly converge", it will turn out soon (Corollary 2.18) that with respect to this "limit" measure, N_+-cuts are more powerful than Gomory-Chvatál cuts. Thus, this mixed hierarchy converges to the same body as (5). Note that this is in sharp contrast with rank complexity: in that model, CP can not in general be simulated by LS_+ [21,

Theorem 7.1], albeit it becomes possible after allowing degree 3 polynomials, at least for cutting planes with bounded coefficients [21, Section 5].

Speaking of convergence, it is not a priori clear that the hierarchies (4), (5) converge even within countably many steps. That is, denoting

$$P^{(\infty,w)} \stackrel{\text{def}}{=} \bigcap_{r\geq 0} P^{(r,w)} \quad N^{\infty}_{+,w}(P) \stackrel{\text{def}}{=} \bigcap_{r\geq 0} N^{r}_{+,w}(P), \tag{6}$$

it might not be immediately obvious that even $P'^{,w}\left(P^{(\infty,w)}\right) = P^{(\infty,w)}$. Therefore, our first order of business is to prove that these hierarchies actually converge within finitely many steps; the proof will also give a very good insight into what kind of techniques we need for lower bound results. We need to introduce some standard notation first.

Definition 2.7 (partial assignments) A *partial 0-1 assignment* ρ that, depending on the context, will be often called *a restriction*, is a mapping $\rho : [n] \longrightarrow \{0,1,*\}$. We let $\sup(\rho) \stackrel{\text{def}}{=} \rho^{-1}(\{0,1\})$ be the set of *assigned variables*. Two equivalent convenient representations of a partial assignment ρ are by the vector of its values $(\rho_1,\ldots,\rho_n)$, $\rho_i \in \{0,1,*\}$, and as a pair (J,a), where $J = \sup(\rho)$ and $a \in \{0,1\}^J$. We will use these representations interchangeably. Let $\mathcal{R}_n$ be the set of all partial assignments in n variables, and let

$$\mathcal{R}_{n,w} \stackrel{\text{def}}{=} \{\,(J,a) \in \mathcal{R}_n \mid |J| \leq w\,\}.$$

Note the trivial bound

$$|\mathcal{R}_{n,w}| \leq (2n)^w. \tag{7}$$

Definition 2.8 (geometric projections) For an n-dimensional vector x or a convex body $P \subseteq [0,1]^n$ and $J \subseteq [n]$, let x_J [P_J] be the projection of x [P, respectively] onto $[0,1]^J$. Note that

$$P \subseteq P_J \times [0,1]^{[n]\setminus J}. \tag{8}$$

We say that $(J,a) \in \mathcal{R}_n$ is *consistent* with P if $a \in P_J$. Equivalently, $(\rho_1,\ldots,\rho_n)$ is consistent with P if we can replace all stars in this vector with (possibly non-integer!) values in $[0,1]$ so that we obtain a point in P. Since both cut operations are monotone in P, (8) readily implies that they are well-behaved with respect to geometric projections:

$$(P')_J \subseteq (P_J)', \quad N_+(P)_J \subseteq N_+(P_J). \tag{9}$$

Theorem 2.9 *Let $P \subseteq [0,1]^n$ be a convex body, $w \leq n$, and let*

$$P = Q^{(0)} \supseteq Q^{(1)} \supseteq \ldots \supseteq Q^{(r)} \supseteq \ldots \tag{10}$$

be one of the two hierarchies (4), (5). *Then*

$$Q^{(r)} = Q^{(r+1)} = Q^{(r+2)} = \ldots,$$

where

$$r = (2n)^w \cdot w^2(1 + \log_2 w). \tag{11}$$

Proof. Before presenting the formal argument, let us briefly explain some intuition behind it (cf. the beginning of Section 8). We split a width $\leq w$ proof in either CP or LS_+ into "rounds", such that during every round we proceed locally within every J with $|J| \leq w$ and reduce Q_J to $(Q_J)_{0-1}$. At the end of each round all local constraints are put together, and we basically show that we either get stuck or make a progress by decreasing the number of restrictions in $\mathcal{R}_{n,w}$ consistent with our polytope.

Let us now formalize this simple idea. Let $\mathcal{R}^{(r)}$ be the set of all partial assignments in $\mathcal{R}_{n,w}$ that are consistent with $Q^{(r)}$. This sequence of sets is clearly non-increasing. Therefore, due to the bound (7), we only have to show that if this sequence stays constant for sufficiently long:

$$\mathcal{R}^{(r)} = \mathcal{R}^{(r+1)} = \ldots = \mathcal{R}^{(s)} = \mathcal{R}^{(s+1)}, \tag{12}$$

where $s = r + w^2(1 + \log_2 w)$, then $Q^{(s)} = Q^{(s+1)}$, that is, the hierarchy (10) collapses at that moment. For that we will show (under the assumption (12)) the following explicit characterization of $Q^{(s)}$ in terms of $Q^{(r)}$:

$$Q^{(s)} = Q^{(r)} \cap \bigcap_{|J| \leq w} \left(\left(Q_J^{(r)} \right)_{0-1} \times [0,1]^{[n] \setminus J} \right). \tag{13}$$

Denote the convex body in the right-hand side of (13) by R.

($\subseteq$ part). $Q^{(s)} \subseteq Q^{(r)}$ is obvious. Let $|J| \leq w$. Then $Q^{(s)} \subseteq \left(Q_J^{(r)} \right)_{0-1} \times [0,1]^{[n] \setminus J}$ is equivalent to

$$Q_J^{(s)} \subseteq \left(Q_J^{(r)} \right)_{0-1}. \tag{14}$$

Let $S \stackrel{\text{def}}{=} Q_J^{(r)}$, and build in $[0,1]^J$ the hierarchy $S = S^{(0)} \supseteq S^{(1)} \ldots \supseteq S^{(t)} \ldots$ either from Proposition 2.2 or 2.4, depending on the type of the hierarchy

(10). Then these propositions, along with the bound (2), imply that in either case $S^{(s-r)} = S_{0-1} = \left(Q_J^{(r)}\right)_{0-1}$. On the other hand, monotonicity properties (9) imply (by induction on $t = 0, 1, \ldots, s-r$) that $Q_J^{(r+t)} \subseteq S^{(t)}$. Plugging in $t := s - r$, we complete the proof of (14).

($\supseteq$ part). We prove by induction on $t = r, r+1, \ldots, s$ that $R \subseteq Q^{(t)}$. The base case $t = r$ is obvious. Assume that $r \leq t \leq s$ and $R \subseteq Q^{(t)}$. Let $c^T x \geq d$ be a cut for $Q^{(t)}$ (of an appropriate type) that has width $\leq w$, and let J be the set of non-zero positions in c. By definitions, $\left(Q_J^{(r)}\right)_{0-1}$ is the integer polytope in $[0,1]^J$ spanned by precisely those $a \in \{0,1\}^J$ for which $(J, a) \in \mathcal{R}^{(r)}$, which is the same as $\mathcal{R}^{(t+1)}$ since $t \leq s$. Since $c^T x \geq d$ is a cut for $Q^{(t)}$, this constraint holds for all points in $Q^{(t+1)}$. In particular, $c_J^T a \geq d$ as long as $(J, a) \in \mathcal{R}^{(r)} (= \mathcal{R}^{(t+1)})$ which means that the constraint $c_J^T x \geq d$ holds on $\left(Q_J^{(r)}\right)_{0-1}$. Since $R \subseteq \left(Q_J^{(r)}\right)_{0-1} \times [0,1]^{[n]\setminus J}$, $c^T x \geq d$ must hold on R. We have proved that R satisfies every cut of width $\leq w$ for $Q^{(t)}$, therefore $R \subseteq Q^{(t+1)}$.

Once we have established (13), the rest is easy. We have that $Q^{(s+1)} \supseteq Q^{(s)} = R$. On the other hand, as we have just proved, $R \subseteq Q^{(s+1)}$. Hence $Q^{(s)} = Q^{(s+1)} = R$.■

The bound (11) is discouragingly different from the neat polynomial bounds in Propositions 2.2 and 2.4. One might think that this is an artifact of our proof method. The following theorem, however, says that in certain cases the incremental reduction of the set of consistent partial assignments is (roughly) the only way of arriving at a contradiction when doing cuts of limited width. Given the general nature of this result, we formulate it now, and for that we need to introduce some logical notation, long overdue.

Definition 2.10 (clauses and resolution) A *literal* is either a Boolean variable x or its negation $\bar{x}$. A *clause* is either a disjunction of literals in which no variable appears along with its negation or 1. 0 is the empty clause. The *width* $w(C)$ of a clause C is defined as the number of literals appearing in it ($w(1) \stackrel{\text{def}}{=} 0$). For a clause C, we let $P_C \subseteq [0,1]^n$ be the polytope spanned by all $\{0,1\}$-points satisfying C. That is, if $C = \ell_1 \vee \ldots \vee \ell_w$, P_C is defined by the single linear constraint

$$f_C \stackrel{\text{def}}{=} \sum_{i=1}^{w} \ell_i \geq 1, \tag{15}$$

where the negated literal $\bar{x}$ is interpreted as $(1-x)$. In particular, $P_0 = \emptyset$, and we also let $P_1 \stackrel{\text{def}}{=} [0,1]^n$. We will sometimes use the uniform notation

$$x^\epsilon \stackrel{\text{def}}{=} \begin{cases} x & \text{if } \epsilon = 1 \\ 1-x & \text{if } \epsilon = 0. \end{cases} \tag{16}$$

A *CNF* τ is a conjunction of clauses. The *width* $w(\tau)$ of a CNF τ is the maximum width of a clause appearing in it. A *k-CNF* is a CNF of width $\leq k$. For $\tau = C_1 \wedge \ldots \wedge C_m$, we let $P_\tau \stackrel{\text{def}}{=} \bigcap_{i=1}^m P_{C_i}$. Note that, unlike P_C, the polytope P_τ is not necessarily integral.

The *resolution proof system* operates with clauses, and it allows just one *resolution rule*

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

The *width* $w(\Pi)$ of a resolution proof is the maximum width of a clause in it. For a CNF τ and a clause C, let $w_R(\tau \vdash C)$ denote the minimum possible value of $w(\Pi)$ taken over all resolution proofs Π of C from τ (we let $w_R(\tau \vdash C) \stackrel{\text{def}}{=} \infty$ if no such proof exists). Of particular interest is the quantity $w_R(\tau \vdash 0)$, that is the minimum possible width of a *refutation* of τ.

The following well-known fact states that resolution is weaker than both cutting planes and Lovász-Schrijver proof systems.

Fact 2.11 *For any two clauses C and D, $f_{C \vee D}$ is both a Gomory-Chvatál cut and an N_+-cut for the polytope $P_{C \vee x} \cap P_{D \vee \bar{x}}$. In particular:*

- **a)** $\left(P_{C \vee x} \cap P_{D \vee \bar{x}}\right)'^{,w} \subseteq P_{C \vee D}$ *for any* $w \geq w(C \vee D)$*;*
- **b)** $N_{+,w}(P_{C \vee x} \cap P_{D \vee \bar{x}}) \subseteq P_{C \vee D}$*, again for any* $w \geq w(C \vee D)$*;*
- **c)** *if τ is an unsatisfiable CNF and $w \geq w_R(\tau \vdash 0)$, then each of the two hierarchies* (4), (5) *converges to the empty polytope.*

Theorem 2.12 *Let $k = k(n) \geq 12$ be an integer parameter, and let $\epsilon > 0$ be an arbitrary constant. Then there exists a sequence of unsatisfiable k-CNF $\{\tau_n\}$, where n is the number of variables, such that $w_R(\tau_n \vdash 0) \leq O(k)$, but for any $w \leq n^{1-\epsilon}/k$, it requires $n^{\Omega(k)}$ steps for either of the two hierarchies* (4), (5) *to converge to the empty polytope.*

This theorem exhibits a very peculiar behavior called in an earlier version of [34] a *supercritical tradeoff*: in the width-restricted world even the most straightforward, exponential size proofs may cease to exist, and the complexity may jump up by an extra exponent. We refer the reader to [34] for an extended discussion of this phenomenon in general, and we prove Theorem 2.12 in Section 6.

Finally, we are ready to give our main definitions.

Definition 2.13 (width of semi-algebraic proof systems) For an infeasible polytope $P \subseteq [0,1]^n$, we let $w_{CP}(P \vdash \emptyset)$ $[w_{LS^+}(P \vdash \emptyset)]$ be the minimum w for which $P^{(\infty,w)} = \emptyset$ $[N_{+,w}(P) = \emptyset$, respectively]. In other words, this is the minimum w for which P can be "refuted" using cuts of the specified nature and of width $\leq w$. For a CNF τ, we abbreviate $w_{CP}(P_\tau \vdash 0)$ to $w_{CP}(\tau \vdash 0)$ and $w_{LS^+}(\tau \vdash 0)$, respectively. Note that Fact 2.11c) implies that $w_{CP}(\tau \vdash 0) \leq w_R(\tau \vdash 0)$ and $w_{LS^+}(\tau \vdash 0) \leq w_R(\tau \vdash 0)$. A conclusive separation between w_R and w_{CP}, w_{LS^+} is provided by Example 1.

Definition 2.14 (integrality gaps) Let $P \subseteq [0,1]^n$ be a feasible polytope and $g : [0,1]^n \longrightarrow \mathbb{R}$ is a continuous *goal function* viewed as a minimization problem. We define the *integrality gap with respect to cuts of width* $\leq w$ as

$$\mathrm{IGap}_{LS^+}(P, g, w) \stackrel{\text{def}}{=} \frac{\min\left\{ g(x) \,\middle|\, x \in N_{+,w}^{(\infty)}(P) \right\}}{\min\{ g(x) \mid x \in P_{0-1} \}}.$$

The definition of $\mathrm{IGap}^{\max}_{LS^+}(P, g, w)$ for maximization problems g or for CP in place of LS^+ is analogous.

An universal method for proving lower bounds on the width of semi-algebraic proofs/algorithms is almost immediate from definitions and the proof of Theorem 2.9.

Definition 2.15 A non-empty polytope $P \subseteq [0,1]^n$ is *w-obstructing* for CP $[LS^+]$ if any Gomory-Chvatál cut [any N_+-cut, respectively] for P of width $\leq w$ is satisfied by P itself. In other words, we require that $P = P^{\prime,w}$ $[P = N_{+,w}(P)$, respectively].

Lemma 2.16 *Let $\mathcal{P}$ be either CP or LS^+.*

a) *For an infeasible polytope P, $w_{\mathcal{P}}(P \vdash \emptyset) \leq w$ if and only if P does not contain any w-obstructing polytope for $\mathcal{P}$.*

b) *For a feasible polytope P,* $\mathrm{IGap}_{\mathcal{P}}(P,g,w) < \alpha$ *if and only if P does not contain any w-obstructing polytope Q for $\mathcal{P}$ with* $\min_{x\in Q} g(x) \geq \alpha \min_{x\in P_{0-1}} g(x)$.

Proof. By Theorem 2.9, the hierarchies (4) and (5) converge to w-obstructing bodies with required properties within finitely many steps, and we only have to prove that $N^{\infty}_{+,w}(P)$ is actually a polytope (for $P^{(\infty,w)}$ it is obvious, and equally obvious is the opposite direction in Lemma 2.16). Denote $Q \stackrel{\mathrm{def}}{=} N^{\infty}_{+,w}(P)$. Then the fact that Q is a polytope readily follows from the following representation:

$$Q = \bigcap_{|J|\leq w} (Q_J)_{0-1} \times [0,1]^{[n]\setminus J}.$$

To prove this, note that $Q \subseteq (Q_J)_{0-1} \times [0,1]^{[n]\setminus J}$ for $|J| \leq w$ simply because Q_J is integral (otherwise we would have had non-trivial N_+-cuts supported on J). For the $\supseteq$-part we once more utilize the identity (13), where this time we choose r so large that $Q^{(r)} = Q^{(s)} = Q$.■

In the case of cutting planes w-obstructing polytopes have a very clean geometric meaning:

Fact 2.17 *$P \subseteq [0,1]^n$ is w-obstructing for CP if and only if P_J is integral for any J with $|J| \leq w$.*

Proof. By Proposition 2.4, any non-integral polytope has at least one non-trivial Gomory-Chvatál cut.■

This fact reflects the local nature of Gomory-Chvatál cuts, and for that reason w-obstructing polytopes for CP will be sometimes called *w-integral polytopes*.

For the purposes of the following corollary, $w_{CP+LS^+}(P \vdash \emptyset)$ and $\mathrm{IGap}_{CP+LS^+}(P,g,w)$ are defined naturally, with respect to the mixed hierarchy combining both kinds of cuts.

Corollary 2.18 **a)** $w_{LS^+}(P \vdash \emptyset) = w_{CP+LS^+}(P \vdash \emptyset) \leq w_{CP}(P \vdash \emptyset)$.

b) *For integrality gaps, we similarly have* $\mathrm{IGap}_{LS^+}(P,g,w) = \mathrm{IGap}_{CP+LS^+}(P,g,w) \leq \mathrm{IGap}_{CP}(P,g,w)$.

Proof. By Proposition 2.4, every w-obstructing polytope P for LS^+ must be also w-integral, that is P_J must be integral for any J with $|J| \leq w$. Hence, by Fact 2.17, P is also a w-obstructing polytope for CP. Now we only have to apply Lemma 2.16.■

Remark 1 As in any lower bound game, Lemma 2.16 in itself is totally useless due to its universal nature. What we need are *constructive* methods for establishing that concrete polytopes are w-obstructing.

Let us note for comparison that Dey, Molinaro and Wang studied in [16] the "analytical" version of w-integral polytopes. Namely, they considered a natural distance $d(P, P^{'},^w)$ as a way to measure progress the sparse cuts can make during one round. [16] proved strong lower and upper bounds on this quantity applicable in rather general situations. In their terminology, we are looking for recipes of constructing reasonably general examples with the ultimate upper bound $d(P, P^{'},^w) = 0$ (or $d(P, N_{+,w}(P)) = 0$). This makes for most of the content of the next three sections.

3. A general construction of w-obstructing polytopes

The general recipe for obtaining w-obstructing polytopes is suggested by the proof of Theorem 2.9, and in particular by (13).

Definition 3.1 For a polytope $P \subseteq [0,1]^n$ and $J \subseteq [n]$, let us define

$$P[J] \stackrel{\text{def}}{=} (P_J)_{0-1} \times [0,1]^{n \setminus J}. \tag{17}$$

Let $\Delta \subseteq \mathcal{P}([n])$ be a non-empty family of subsets of $[n]$ which is *downward closed*, that is $\forall J, J' \subseteq [n](J \subseteq J' \wedge J' \in \Delta \Longrightarrow J \in \Delta)$. Then we define

$$P[\Delta] \stackrel{\text{def}}{=} P \cap \bigcap_{J \in \Delta} P[J].$$

Example 2 (downward closed sets) The most natural downward closed sets is $[n]^{\leq \ell}$ for an integer $\ell \leq n$. It will be used in Section 4.1. For a graph G and an integer ℓ, let $\Delta_{G,\ell} \subseteq \mathcal{P}(E(G))$ be the family of all sets of edges $E \subseteq E(G)$ that can be covered by at most ℓ vertices. This family will be used in Section 4.2.

We are going to prove that under certain conditions on Δ, the polytope $P[\Delta]$ is w-obstructing. It might be tempting to assume that $P[[n]^{\leq w}]$ itself has this property, but unfortunately this is not the case in all interesting cases: the obstruction property is rather subtle and unstable. The remedy is traditional for proof complexity and consists in replacing $[n]^{\leq w}$ with a larger family Δ that has nice "closure properties".

Before proceeding to the corresponding statement (Theorem 3.4), we need one more simple observation essentially stating that the definitions given in the previous section behave well with respect to restrictions. We formulate this property in a slightly more general context known in complexity theory as (propositional) projections that we will need in Section 7.3.

Definition 3.2 (propositional projections) A (propositional) *projection* is a mapping π from the set $\{x_1, \ldots, x_n\}$ of original propositional variables to a set consisting of Boolean constants 0,1 and literals. Projections act naturally on clauses, CNFs etc.; let in particular $\tau|_\pi$ be the CNF obtained from τ by applying the projection π.

Any projection $\pi : \{x_1, \ldots, x_n\} \longrightarrow \{0, 1, x_1, \bar{x}_1, \ldots, x_{n'}, \bar{x}_{n'}\}$ defines the dual mapping $\pi^\sharp : [0,1]^{n'} \longrightarrow [0,1]^n$ by

$$\pi^\sharp(a_1, \ldots, a_{n'})(i) \stackrel{\text{def}}{=} \begin{cases} \pi(x_i) & \text{if } \pi(x_i) \in \{0,1\} \\ a_j & \text{if } \pi(x_i) = x_j \\ \bar{a}_j & \text{if } \pi(x_i) = \bar{x}_j. \end{cases}$$

For a polytope $P \subseteq [0,1]^n$, let $P|_\pi \stackrel{\text{def}}{=} (\pi^\sharp)^{-1}(P)$. This operation is well-behaved:

$$P|_{\tau|_\pi} = (P_\tau)|_\pi \text{ for any CNF } \tau, \quad (P|_\pi)' \subseteq (P')|_\pi, \quad N_+(P|_\pi) \subseteq N_+(P)|_\pi \quad (18)$$

(the two latter containments readily follow from the fact that propositional projections also naturally act on cuts) etc.

For $J \subseteq [n]$ and a projection π as above, let

$$\pi(J) \stackrel{\text{def}}{=} \{\, j' \in [n'] \mid \exists j \in J(\pi(x_j) \in \{x_{j'}, \bar{x}_{j'}\})\,\}.$$

For a downward closed family $\Delta \subseteq \mathcal{P}([n])$, let

$$\pi(\Delta) \stackrel{\text{def}}{=} \{\, \pi(J) \mid J \in \Delta \,\}.$$

Restrictions can be viewed as a special kind of projections assigning every variable to either a Boolean constant or to itself. For a restriction $\rho = (J, a)$, the general definitions given above simplify to:

$$P|_\rho = \left\{ y \in [0,1]^{[n]\setminus J} \mid (a, y) \in P \right\}$$

and

$$\rho(\Delta) = \{ J \setminus \sup(\rho) \mid J \in \Delta \}.$$

Lemma 3.3 *For any non-empty Δ and a propositional projection π, we have*

$$P|_\pi[\pi(\Delta)] \subseteq P[\Delta]|_\pi.$$

Proof. Assume that $y \in P|_\pi[\pi(\Delta)]$; we need to show that $\pi^\sharp(y) \in P[\Delta]$. First, $\pi^\sharp(y) \in P$ simply because $y \in P|_\pi$, hence we only need to prove that $\pi^\sharp(y) \in P[J]$ for every $J \in \Delta$. Fix any such J. Then, since $\pi(J) \in \pi(\Delta)$, we know that $y \in P|_\pi[\pi(J)]$. This means that there exist $\pi(J)$-integral points $z_1, \ldots, z_\ell \in P|_\pi$ such that $y|_{\pi(J)}$ is in the convex hull of $z_1|_{\pi(J)}, \ldots, z_\ell|_{\pi(J)}$. But for every $j \in J$, $\pi(x_j)$ is either a constant 0,1 or a literal of a variable $x_{j'}$ such that $j' \in \pi(J)$. Hence $\pi^\sharp(y)_J$ is in the convex hull of $\pi^\sharp(z_1)_J, \ldots, \pi^\sharp(z_\ell)_J$, and these latter points are integral points in P_J. This implies $\pi^\sharp(y) \in P[J]$.■

In the following theorem we could in principle let Δ be equal to the download closure of $\mathcal{J}$. But that would be contrary to the spirit of its applications in which Δ is a natural and clean object, and $\mathcal{J}$ is constructed ad hoc with the only purpose to make this Δ work.

Theorem 3.4 *Let $P \subseteq [0,1]^n$ be a non-empty polytope, $w \leq n$, and let Δ be a downward closed family of subsets of $[n]$. Assume that there exists $\mathcal{J} \subseteq \Delta$ such that $\emptyset \in \mathcal{J}$ and the following holds.*

- **a)** *For every $J \in [n]^{\leq w}$ there exists $\widehat{J} \in \mathcal{J}$ such that $\widehat{J} \supseteq J$ (or, in other words, the downward closure of $\mathcal{J}$ contains $[n]^{\leq w}$).*
- **b)** *For every restriction $\rho = (J, a)$ with $J \in \mathcal{J}$ that is consistent with P, the polytope $P|_\rho[\rho(\Delta)]$ is non-empty.*

Then $P[\Delta]$ is w-integral, which in particular implies $w_{CP}(P \vdash \emptyset) > w$.

If, moreover, in b) *the conclusion can be strengthened to $N_+\left(P|_\rho[\rho(\Delta)]\right) \neq \emptyset$, then $P[\Delta]$ is w-obstructing for LS_+.*

Proof. We first note that $P[\Delta] \neq \emptyset$. Indeed, since $P \neq \emptyset$, the empty restriction is consistent with it, and since $\emptyset \in \mathcal{J}$, we can apply b).

As for the rest, let us start with the case of cutting planes as it is a bit simpler. Due to assumption a) and the fact that the integrality of a polytope is preserved under projections, it is sufficient to prove that for every $J \in \mathcal{J}$, the polytope $P[\Delta]_J$ is integral. Fix $J \in \mathcal{J}$; we claim that in fact $P[\Delta]_J = (P_J)_{0-1}$.

The inclusion $P[\Delta]_J \subseteq (P_J)_{0-1}$ immediately follows from (17) since $P[\Delta] \subseteq P[J]$. For the opposite direction, let $a \in \{0,1\}^J$ be a vertex of the polytope $(P_J)_{0-1}$. This means that $\rho \stackrel{\text{def}}{=} (J, a)$ is consistent with P. By assumption b), $P|_\rho[\rho(\Delta)] \neq \emptyset$. Lemma 3.3 then implies that $P[\Delta]|_\rho \neq \emptyset$, that is precisely $a \in P[\Delta]_J$.

Assume now that in b) we have the stronger assumption $N_+\left(P|_\rho[\rho(\Delta)]\right) \neq \emptyset$, and let $c^T x \geq d$ be an N_+-cut for $P[\Delta]$ of width $\leq w$. Let $J \in \mathcal{J}$ be a set containing all non-zero coordinates in c, which exists due to the assumption a). We need to show that $c^T x \geq d$ holds on $P[\Delta]$, and since, as we have just proved, $P[\Delta]_J = (P_J)_{0-1}$, it suffices to show that $c_J^T a \geq d$ for every $a \in \{0,1\}^J$ such that $\rho = (J, a)$ is consistent with P. From $N_+(P|_\rho[\rho(\Delta)]) \neq \emptyset$ we conclude, by Lemma 3.3 and monotonicity of the operator N_+ that $N_+(P[\Delta]|_\rho) \neq \emptyset$. Since $\{a\} \times P[\Delta]|_\rho \subseteq P[\Delta]$, we also have $\{a\} \times N_+(P[\Delta]|_\rho) \subseteq N_+(P[\Delta])$. Hence $c^T x \geq d$ holds on $\{a\} \times N_+(P[\Delta]|_\rho)$, and since this polytope is non-empty, we conclude $c_J^T a \geq d$.■

4. Concrete lower bounds: infeasible case

In this section we give several applications of Theorem 3.4 in the context of proof complexity, i.e., lower bounds on $w_{CP}(\tau_n \vdash 0)$ or $w_{LS^+}(\tau_n \vdash 0)$ for unsatisfiable CNFs τ_n. In all our examples, τ_n will have constant width: even if it is formally not stipulated by our definitions, we find the case when the original polytope cannot be even expressed in our proof system to be much less natural and interesting.

In order to preserve momentum and highlight the main ideas, we present our arguments in a distinctly modular way. As a part of this effort, we defer to Section 7 the proofs of most context-specific lemmas, particularly since some of them are rather tedious.

4.1. Systems of linear equations and random CNFs

Let A be a $m \times n$ 0-1 matrix, and let $b \in \{0,1\}^m$. For $i \in [m]$ we let

$$J_i(A) \stackrel{\text{def}}{=} \{\, j \in [n] \mid a_{ij} = 1 \,\}.$$

For $J \subseteq [n]$, X_J is the set of the corresponding propositional variables: $X_J \stackrel{\text{def}}{=} \{\, x_j \mid j \in J \,\}$; let also $X_i(A) \stackrel{\text{def}}{=} X_{J_i(A)}$ and $\bigoplus X_i(A) \stackrel{\text{def}}{=} \bigoplus_{j \in J_i(A)} x_j$, where $\oplus$ is addition in $\mathbb{F}_2$. For any row $i \in [m]$ of the matrix A we introduce the corresponding linear equation:

$$\bigoplus X_i(A) = b_i, \tag{19}$$

and we let $\tau(A,b)$ be the CNF obtained from this $\mathbb{F}_2$-linear system $AX = b$ by expanding all equations as CNFs in the straightforward way. Our goal is to prove a lower bound on $w_{LS^+}(\tau(A,b) \vdash 0)$ when A has relatively mild expansion properties. Let us recall some necessary definitions.

Definition 4.1 For a set $I \subseteq [m]$ of rows in the matrix A, we define its *boundary* $\partial_A(I)$ as the set of all $j \in [n]$ for which there exists a *unique* $i \in I$ with $a_{ij} = 1$. For $r \geq 1$ we say that A is an (r,s,c)-*boundary expander*[4] if $|J_i(A)| \leq s$ for all $i \in [m]$ and

$$\forall I \subseteq [m] (|I| \leq r \Longrightarrow |\partial_A(I)| \geq c \cdot |I|).$$

Whenever the upper bound on the size of $J_i(A)$ is unimportant, (r,∞,c)-boundary expanders will be simply called (r,c)-boundary expanders.

For a set of columns $J \subseteq [n]$ let

$$\mathrm{Ker}(J) \stackrel{\text{def}}{=} \{\, i \in [m] \mid J_i(A) \subseteq J \,\}$$

be the set of rows completely contained in J. Let $A \setminus J$ be the sub-matrix of A obtained by removing all columns in J and all rows in $\mathrm{Ker}(J)$.

Random matrices have very good expansion properties, and statements to that effect tailored to different ranges of parameters are abundant in the literature. We will utilize [3, Lemma 4.1] that, in turn, was with minor modifications adopted from [13].

[4] In [2] such matrices were called simply expanders

Proposition 4.2 *Assume that $k \geq 3$ is a fixed integer constant, $0 < c < k-2$ is a real constant and $d = d(n)$ is an integer parameter satisfying $d \leq o\left(n^{(k-c-2)/2}\right)$. Then a random $(dn \times n)$ matrix $\boldsymbol{A}$ in which every row $J_i(\boldsymbol{A})$ is chosen from $[n]^k$ independently and uniform at random is an $\left(\Omega\left(\frac{n}{d^{2/(k-c-2)}}\right), k, c\right)$-expander with probability $1 - o(1)$.*

We can now state the main result of this section.

Theorem 4.3 *Assume that $c > 0$ is an arbitrary fixed constant, and that A is an (r, s, c)-boundary expander. Then for any $b \in \{0,1\}^m$ we have $w_{LS^+}(\tau(A, b) \vdash 0) \geq \Omega(r/s)$.*

Before proving this theorem we need two auxiliary facts. The first one is also omnipresent (in different forms) in the proof complexity literature.

Lemma 4.4 *Let A be an $m \times n$ (r, s, c)-boundary expander, and let $c' < c$. Then for every $J \subseteq [n]$ with $|J| \leq \frac{r}{2}(c - c')$ there exists $\widehat{J} \supseteq J$ such that $A \setminus \widehat{J}$ is an $(r/2, c')$-boundary expander, $\left|Ker\left(\widehat{J}\right)\right| \leq \frac{|J|}{c-c'}$ and*

$$\left|\widehat{J}\right| \leq |J|\left(1 + \frac{s}{c - c'}\right). \tag{20}$$

The proof is deferred to Section 7.1.

Lemma 4.5 *Let A be an $m \times n$ (r, s, c)-boundary expander, where $r \geq 2$ and $rc \geq 2(s+1)$, and let*

$$\ell \stackrel{\text{def}}{=} \lfloor \frac{rc}{2} \rfloor - s - 1.$$

Then for any $b \in \{0,1\}^m$ we have $N_+\left(P_{\tau(A,b)}[[n]^{\leq \ell}]\right) \neq \emptyset$.

For the proof see Section 7.3.

Proof of Theorem 4.3. We abbreviate $P_{\tau(A,b)}$ to $P_{A,b}$. Thus, $P_{A,b} = \bigcap_{i=1}^m P_i$, where P_i is the polytope determined by the ith equation in (19) (the polytopes P_i happen to be integral, but we will not need this fact in what follows). These polytopes behave extremely well with respect to restrictions. Namely, let us call a restriction ρ *weakly consistent* with the system (19) if it does not reduce any of its equations to 1=0 (it is allowed to reduce them to 0=0). Otherwise, it is *strongly inconsistent*.

Fact 4.6 *If ρ is strongly inconsistent with the system $Ax = b$, then $(P_{A,b})|_\rho = \emptyset$. Otherwise, if, say, $\rho = (J, a)$, then $(P_{A,b})|_\rho = P_{A\setminus J,b'}$ (that still may turn out to be empty), where $b' \in \{0,1\}^{[m]\setminus Ker(J)}$ is a vector naturally determined by ρ: $b'_i \stackrel{\text{def}}{=} b_i \oplus \bigoplus \{\, a_j \mid j \in J \cap J_i(A) \,\}$.*

Let us now proceed to the actual proof. We can assume w.l.o.g. that

$$r \geq \max\left(4, \frac{20s}{c}\right)$$

as otherwise the bound becomes trivial. Let

$$w \stackrel{\text{def}}{=} \min\left\{\frac{rc}{4}, \left(\frac{rc}{8} - s - 1\right)\left(1 + \frac{2s}{c}\right)^{-1}\right\},$$

and note that $w \geq \Omega(r/s)$. We will prove that $w_{LS^+}(P_{A,b} \vdash 0) \geq w$. For that purpose we are going to apply Theorem 3.4 with $\Delta := [n]^{\leq \ell}$, where

$$\ell \stackrel{\text{def}}{=} w\left(1 + \frac{2s}{c}\right) \leq O(ws).$$

We have to design the family $\mathcal{J}$. Let us call a set of columns J *closed* if $A \setminus J$ is an $(r/2, c/2)$-boundary expander. Since

$$\ell \leq \frac{rc}{8} - s - 1 \tag{21}$$

due to our choice of w, Lemma 4.4 with $c' := c/2$ implies (note that $w \leq rc/4$) that every J with $|J| \leq w$ is contained in a closed set $\widehat{J}$ of cardinality $\leq \ell$. We let $\mathcal{J}$ be the family of all such closed $\widehat{J}$. It remains to check the assumption b) in the statement of Theorem 3.4 in its stronger form, i.e., for N_+-cuts.

Let $J \in \mathcal{J}$, i.e., J is a closed set with $|J| \leq \ell$, and let $\rho = (J, a)$ be consistent with $P_{A,b}$. We need to prove that

$$N_+\left((P_{A,b})|_\rho\left[\rho\left([n]^{\leq \ell}\right)\right]\right) \neq \emptyset. \tag{22}$$

But by Fact 4.6, $(P_{A,b})|_\rho = P_{A',b'}$, where $A' \stackrel{\text{def}}{=} A \setminus J$ is an $(r/2, c/2)$-boundary expander since J is closed. Note also that $\rho\left([n]^{\leq \ell}\right) = \{\, J' \subseteq [n] \setminus J \mid |J'| \leq \ell \,\}$. Hence (22) immediately follows from Lemma 4.5 applied to $A := A'$, $r := r/2$, $c := c/2$ using the bound (21). This completes the proof of Theorem 4.3■

From Theorem 4.3 we can now derive the "traditional" set of corollaries.

Definition 4.7 A *random 3-CNF with n variables and m clauses* [a *random 3-XOR formula with n variables and m equations*] is obtained by picking independently and uniformly at random, with repetitions, m clauses [m affine constraints over $\mathbb{F}_2$] from the set of all clauses of width 3 [all $\mathbb{F}_2$-affine constraints with 3 variables, respectively].

Theorem 4.8 *Let $h, \epsilon > 0$ be arbitrary constants, and let $\boldsymbol{A}X = \boldsymbol{b}$ be a random 3-XOR formula with n variables and n^{1+h} equations. Then*

$$w_{LS^+}(\tau(A, b) \vdash 0) \geq \Omega(n^{1-2h-\epsilon})$$

with probability $1 - o(1)$.

Proof. Set $k = 3$, $d = n^h$ and $c = \frac{\epsilon}{2h+\epsilon}$ in Proposition 4.2. We can assume w.l.o.g. that $2h + \epsilon < 1$ since otherwise the bound is trivial. Then A is an $\left(\Omega\left(n^{1-2h-\epsilon}\right), 3, c\right)$-boundary expander with probability $1 - o(1)$. Now we apply Theorem 4.3.■

Theorem 4.9 *Let $h, \epsilon > 0$ be arbitrary constants, and let τ be a random 3-CNF with n variables and n^{1+h} equations. Then*

$$w_{LS^+}(\tau \vdash 0) \geq \Omega(n^{1-2h-\epsilon})$$

with probability $1 - o(1)$.

Proof. Let A be the incidence matrix between clauses and variables of τ. Let $b_i \in \{0, 1\}$ be the parity of the number of occurrences of negative literals in the ith clause C_i of τ, plus one. Then C_i appears in the CNF expansion of (19), and hence τ is a sub-CNF of $\tau(A, b)$. Now Theorem 4.9 immediately follows from Theorem 4.8.■

Definition 4.10 (Tseitin tautologies) Let G be a connected undirected graph, and let $\sigma : V(G) \longrightarrow \{0, 1\}$ be a function such that $\bigoplus_{v \in V(G)} \sigma(v) = 1$. The *Tseitin tautology* $T(G, \sigma)$ is defined as the unsatisfiable CNF in the variables $(x_e \mid e \in E(G))$ resulting from the system

$$\left\{ \bigoplus_{e \ni v} x_e = \sigma(v) \mid v \in V(G) \right\} \tag{23}$$

of $\mathbb{F}_2$-linear equations.

Definition 4.11 (edge cuts and edge expansion) For a graph G and a set of vertices $S \subseteq V(G)$, $\delta(S)$ is its *edge cut* defined as the set of edges between S and its complement $V(G) \setminus S$. The graph G has *edge expansion* $\geq c$ if for every set S of vertices with $|S| \leq \frac{|V(G)|}{2}$, it holds that $|\delta(S)| \geq c \cdot |S|$.

It is well known since [31] that for $d \geq 3$ random d-regular graphs have edge expansion $\geq c$ with probability $1 - o(1)$, where $c > 0$ is an absolute constant.

Theorem 4.12 *Let $c > 0$ be an arbitrary constant. Then for every constant-degree graph G with n vertices and edge expansion $\geq c$ we have*

$$w_{LS^+}(T(G, \sigma) \vdash 0) \geq \Omega(n).$$

Proof. Let $m = |E(G)|$ and A be the $n \times m$ (!) matrix corresponding to the system (23), that is the transpose of the incidence matrix of the graph G. Then the assumption on the edge expansion of G translates to the fact that A is an $(n/2, \Delta(G), c)$-boundary expander, where $\Delta(G)$ is the maximum degree of G. The result now immediately follows from Theorem 4.3.■

4.2. Perfect matching principle

Matching and perfect matching polytopes, as well as their variations, are among the most widely studied polytopes in combinatorial optimization, beginning with the pioneering work by Edmonds [17]. Remarkably (but not surprisingly) these also were the first polytopes on which the cutting plane method was demonstrated [12]. In this section we apply Theorem 3.4 to obtain lower bounds on the width of cutting planes refutations of perfect matching principles for bounded-degree graphs. On the high level, the argument is similar to the one used in Section 4.1 and relies on a suitable notion of a closed set of vertices (Definition 4.18) below. However, proofs of essential lemmas in this section use ideas that are rather different from the proof of Lemma 4.5; we defer further discussion to Section 7.3 in which both will be presented, and to the concluding section 8.

Throughout the section $G = (V(G), E(G))$ will stand for a simple undirected graph. We will work in the real space $\mathbb{R}^{E(G)}$ (most of the time in the cube $[0, 1]^{E(G)}$); the corresponding variables will be denoted by x_e ($e \in E(G)$). For a set of edges $E \subseteq E(G)$, $x(E) \stackrel{\text{def}}{=} \sum_{e \in E} x_e$. Recall (Definition 4.11) that

for a set of vertices S, $\delta(S)$ is the edge cut defined by S; we abbreviate $\delta(\{v\})$ to $\delta(v)$. $E(S)$ is the set of edges that have *both* endpoints in S; let also $N(S) \stackrel{\text{def}}{=} E(S) \dot\cup \delta(S)$ be the (edge) neighbourhood of S.

The *fractional perfect matching polytope* $FPM(G) \subseteq [0,1]^{E(G)}$ is determined by the set of constraints $x_e \geq 0$ ($e \in E(G)$) and $x(\delta(v)) = 1$ ($v \in V(G)$) [27]. 0-1 points in $FPM(G)$ are precisely [characteristic functions of] perfect matchings in G; thus $FPM(G)_{0-1} = PM(G)$, where $PM(G)$ is the *perfect imatching polytope*. By the celebrated results due to Edmonds, $PM(G)$ is determined within $FPM(G)$ by the set of additional constraints

$$x(E(S)) \leq \lfloor \frac{|E|}{2} \rfloor \ (S \subseteq V(G)).$$

These polytopes coincide when G is bipartite but in general they can be different. In this section we essentially show that they can be *quite* different as long as sublinear-width cutting planes procedures are concerned. More exactly, our goal is to prove a linear lower bound on $w_{CP}(FPM(G) \vdash 0)$ for constant-degree random graphs G. For the sake of completeness, let us remind the *uniform model* that is a very handy way of generating such graphs (see e.g. [9, 39]).

Definition 4.13 (random graphs) Let dn be even. Consider a set V of dn vertices partitioned into n *cells* $V_1, \ldots, V_n$, of d vertices each. Pick uniformly at random a perfect matching $\boldsymbol{P_{n,d}}$ on $V_1 \dot\cup \ldots \dot\cup V_n$ and make an undirected graph on $[n]$ out of it by mapping every $(v,w) \in \boldsymbol{P_{n,d}}$ with $v \in V_i$ and $w \in V_j$ to (i,j). The result $G(\boldsymbol{P_{n,d}})$ of this operation is not necessarily simple, but for any fixed d and $n \to \infty$ it is simple with probability $\Omega(1)$ [39, Section 2.2]. $\mathcal{G}_{n,d}$ is the distribution resulted from conditioning by the event "$G(\boldsymbol{P_{n,d}})$ is simple". The importance of this model stems from the fact that this distribution is actually *uniform* on the set of all labelled d-regular graphs on n vertices, and we let $\boldsymbol{G_{n,d}}$ be the random variable sampling from it.

Our main result in this section is the following theorem.

Theorem 4.14 *Let $d \geq 14$ be any fixed constant, and let $n \to \infty$. Then with probability $1 - o(1)$ we have*

$$w_{CP}(FPM(\boldsymbol{G_{n,d}}) \vdash \emptyset) \geq \Omega(n).$$

Note that if $n \to \infty$ is even, then for any fixed d $\boldsymbol{G_{n,d}}$ contains a perfect matching with probability $1 - o(1)$ [9], and Theorem 4.14 holds for trivial reasons. Thus we may assume w.l.o.g. that n is odd (and hence d must be even). Also, the combinatorics can be simplified a bit if we compromise on the exact value of d. But since one of our primary goals is to develop techniques that can be potentially used elsewhere, we prefer to (reasonably) optimize on it instead.

We begin the proof of Theorem 4.14 with distilling combinatorial properties of $\boldsymbol{G_{n,d}}$ that are needed for it.

Let us say that a graph G is (ℓ, ϵ)-*sparse* if for every set of vertices U with $|U| \leq \ell$ we have $|E(U)| \leq (1+\epsilon)|U|$. Clearly, this property is *hereditary*, that is any (not necessarily induced) subgraph of an (ℓ, ϵ)-sparse graph is also (ℓ, ϵ)-sparse.

Lemma 4.15 *For every integer constant $d > 0$ and any $\epsilon > 0$ there exists*[5] *$b > 0$ such that $\boldsymbol{G_{n,d}}$ is (bn, ϵ)-sparse with probability $1 - o(1)$.*

Lemma 4.15 implies that $\boldsymbol{G_{n,d}}$ may have only a few short cycles and that it is an extremely good edge expander for $r \leq bn$, where b is a tiny constant. For our argument, however, we need to eliminate all such cycles whatsoever and be sure that the graph has extremely good *vertex* expansion properties that hold up to much larger values of r. We will combine all three goals into one lemma, but before stating it we need to recall a few more key concepts from graph theory.

Definition 4.16 (graph trivia) The *girth* $\mathrm{girth}(G)$ of a graph G is the length of the shortest cycle in it. $\Delta(G)$ is the maximum degree of a vertex $v \in V(G)$. A subgraph G' of G is *spanning* if $V(G') = V(G)$. $G - U$ is the subgraph induced by the set of vertices $V(G) \setminus U$; its edge set is $E(V(G) \setminus U)$.

The (vertex) *neighborhood* $\partial_G(S)$ of a vertex set $S \subseteq V(G)$ (that will be denoted simply by $\partial(S)$ whenever G is clear from the context) is the set of vertices in $V(G) \setminus S$ connected to at least one vertex in S. Equivalently, $\partial_G(S)$ is the set of all $v \notin S$ that are endpoints of edges in $\delta(S)$. G is an (r, c)-(*vertex*)*expander* if for any set of vertices S with $|S| \leq r$ we have $|\partial(S)| \geq c \cdot |S|$. Thus, the case $r = n/2$ corresponds to ordinary vertex expansion.

[5] One can actually take $b \sim d^{-O(1/\epsilon)}$

Lemma 4.17 *Let $d \geq 14$ and $g \geq 0$ be arbitrary integer constants. Then with probability $1 - o(1)$ as $n \to \infty$, $\boldsymbol{G_{n,d}}$ contains a spanning subgraph G with $\mathrm{girth}(G) \geq g$ that is an $(0.203n, 5/2)$-vertex expander.*

The most tedious part here is the bound on expansion. The reader interested in a stand-alone statement that in particular implies this bound should consult Theorem 7.2.

Lemmas 4.15 and 4.17 will be proved in Section 7.2.

Definition 4.18 A subset $S \subseteq V(G)$ is *(r, c)-closed* if for any $D \subseteq \partial(S)$, $G - (S \cup D)$ is an (r, c)-expander.

The following is somewhat analogous to Lemma 4.4.

Lemma 4.19 *Let G be an (r, c)-vertex expander, and let $c' < c$, $r' < r$. Then for every $U \subseteq V(G)$ with*

$$|U| \leq \frac{(r - r')(c - c')}{\Delta(G)} \tag{24}$$

there exists an (r', c')-closed set $S \supseteq U$ with

$$|S| \leq \frac{\Delta(G)}{c - c'}|U|.$$

The proof of Lemma 4.19 will be given in Section 7.1.

Recall from Example 2 that $\Delta_{G,\ell} \subseteq \mathcal{P}(E(G))$ is the downward closed family consisting of all edge sets $E \subseteq E(G)$ such that $E \subseteq N(U)$ for some U with $|U| \leq \ell$. We need an analogue of Lemma 4.5 to prove that under certain conditions the polytopes $FPM(G)[\Delta_{G,\ell}]$ are non-empty. We again present this proof (which is crucial for the entire argument) in the modular fashion and find a point in this polytope via several more-or-less independent steps.

Lemma 4.20 *Let a, c be constants such that $c > 2$ and $ac > 1/2$. Then there exists $\varepsilon > 0$ such that for sufficiently large n and for any (an, c)-vertex expander G on n vertices the polytope $FPM(G)$ contains a point x with $||x||_\infty \leq \frac{1-\varepsilon}{2}$.*

In particular, $FPM(G) \neq \emptyset$.

The proof of Lemma 4.20 will be given in Section 7.3.

Next, we are going to show that for a "nice" graph G, *every* point $x \in FPM(G)$ with $||x||_\infty \leq \frac{1-\varepsilon}{2}$ actually belongs to $FPM(G)[\Delta_{G,\ell}]$ for an appropriately chosen ℓ. As an intermediate step in this proof we need to introduce an auxiliary polytope nicely interpolating between matching and perfect matching polytopes.

Definition 4.21 (S-matching polytopes) For $S \subseteq V(G)$, a matching in G is an *S-matching* if it covers all vertices in S (and possibly some other vertices). Let $M_S(G)$ be the integral polytope spanned by S-matchings (here and in what follows we identify matchings with their characteristic functions).

Thus, the matching polytope $M(G)$ is simply $M_\emptyset(G)$, and the perfect matching polytope $PM(G)$ is $M_{V(G)}(G)$. The dual description of the polytope $M_S(G)$ is almost immediate from Edmond's theorem [17]:

Lemma 4.22 *The polytope $M_S(G)$ is determined by the following set of constraints:*

a) $x_e \geq 0$ $(e \in E(G))$;

b) $x(\delta(v)) \leq 1$ $(v \in V(G))$ *and, moreover,* $x(\delta(v)) = 1$ *if* $v \in S$;

c) $x(E(U)) \leq \lfloor \frac{|U|}{2} \rfloor$ $(U \subseteq V(G))$.

Proof. Every S-matching clearly satisfies all these constraints, hence one direction is obvious. In the opposite direction, assume that $x \in [0,1]^{E(G)}$ satisfies constraints a)-c). As this set contains Edmond's original constraints, $x \in M(G)$ $(= M_\emptyset(G))$. Hence $x = \sum_i c_i y_i$ where $c_i \geq 0$, $\sum_i c_i = 1$ and y_i are matchings. But $\sum_{v \in S} x(\delta(v)) = |S|$, and for every i we have $\sum_{v \in S} y_i(\delta(v)) \leq |S|$, with the equality taking place if and only if y_i is an S-matching. Hence all y_is are actually S-matchings and $x \in M_S(G)$ follows.■

For our purposes we are more interested in the projection of this polytope onto $N(S)$.

Lemma 4.23 *The polytope $M_S(G)_{N(S)}$ is determined by the following set of constraints:*

a) $x_e \geq 0$ $(e \in N(S))$;

b) $x(\delta(v)) = 1$ $(v \in S)$ *and* $x(\delta(v) \cap N(S)) \leq 1$ $(v \in \partial(S))$*;*

c) $x(E(U) \cap N(S)) \leq \lfloor \frac{|U|}{2} \rfloor$ $(U \subseteq S \cup \partial(S))$.

Proof. In one direction it is again obvious. In the opposite direction, let $x \in [0,1]^{N(S)}$ satisfies all these constraints. Extend x to $\widehat{x} \in [0,1]^{E(G)}$ by zeros. Then this vector satisfies all constraints in Lemma 4.22 (for the constraint c) note that $\widehat{x}(E(U)) = x(E(U \cap (S \cup \partial(S))) \cap N(S))$ for any set U). Hence $\widehat{x} \in M_S(G)$ and thus $x \in M_S(G)_{N(S)}$.■

Now, Lemma 4.23 will allow us to place the (projection of the) point x guaranteed by Lemma 4.20 into the (projection of the) polytope $M_S(G)$.

Lemma 4.24 *Assume that G is an (r, ε)-sparse graph with girth$(G) \geq \varepsilon^{-2}$. Then for every $S \subseteq V(G)$ with $|S| \leq \frac{r}{\Delta(G)+1}$ and $x \in FPM(G)$ with $||x||_\infty \leq \frac{1-\varepsilon}{2}$ we have*

$$x_{N(S)} \in M_S(G)_{N(S)}.$$

Proof. $x_{N(S)}$ obviously satisfies constraints a), b) in Lemma 4.23. In order to check c), fix $U \subseteq S \cup \partial(S)$. Note first that $|U| \leq |S|(\Delta(G)+1) \leq r$. Also, $x_{N(S)}(E(U) \cap N(S)) \leq x(E(U))$. Now, if $|U| < \varepsilon^{-2}$ then $G|_U$ is a forest (since girth$(G) \geq \varepsilon^{-2}$), hence it is bipartite and $x(E(U)) \leq \lfloor \frac{|U|}{2} \rfloor$ follows simply because $\lfloor \frac{|U|}{2} \rfloor$ is an upper bound on the size of at least one of its two parts. On the other hand, if $|U| \geq \varepsilon^{-2}$, we apply the sparsity condition and conclude that

$$x(E(U)) \leq \frac{1-\varepsilon}{2}|E(U)| \leq \frac{1-\varepsilon^2}{2} \cdot |U| \leq \frac{|U|-1}{2} \leq \lfloor \frac{|U|}{2} \rfloor.$$

■

The polytope $M_S(G)_{N(S)}$ may in general be larger than the polytope $(FPM(G)_{N(S)})_{0-1}$ we are interested in (see Definition 3.1). Roughly speaking, the former is generated by all S-matchings while the latter is generated only by those S-matchings that are extendable to a point in $FPM(G)$. However, we will show that they happen to coincide when S is closed in the sense of Definition 4.18, and it turns out to be good enough for our purposes. Moreover, we will also be able to re-use Lemma 4.20 for this concluding step (even if it is a bit of an overkill here) rather than prove a separate statement to that effect.

Having thus postponed to Section 7 all real work that, for the record, consists of Lemmas 4.15, 4.17, 4.19, 4.20, we can comfortably finish the proof.

Proof of Theorem 4.3. Choose some constants $a' < a < 0.203$ and $c' < c < 5/2$ such that $a'c' > 1/2$ (which in particular implies $c' > 2$). Let $\varepsilon > 0$ be the constant guaranteed by Lemma 4.20 (note that $ac > a'c' > 1/2$ and $c > c' > 2$). Pick $b > 0$ according to Lemma 4.15 so that $\boldsymbol{G_{n,d}}$ is (bn, ε)-sparse with probability $1 - o(1)$. Let

$$\begin{aligned} L &\stackrel{\text{def}}{=} \frac{bn}{d+1} \\ \ell &\stackrel{\text{def}}{=} \min\left\{\frac{(a-a')(c-c')}{d}n,\ \frac{c-c'}{d}L\right\} \\ w &\stackrel{\text{def}}{=} \min\left\{\frac{(0.203-a)(5/2-c)}{d}n, \frac{5/2-c}{d}\ell\right\}. \end{aligned} \tag{25}$$

Note that $L, \ell, w \geq \Omega(n)$. We are going to prove that with probability $1 - o(1)$ we have

$$w_{CP}(FPM(\boldsymbol{G_{n,d}}) \vdash \emptyset) \geq w. \tag{26}$$

By Lemma 4.17, it is sufficient to prove (26) for any graph G that has the following set of properties:

1. $\Delta(G) \leq d$;
2. $\text{girth}(G) \geq \varepsilon^{-2}$;
3. G is (bn, ε)-sparse;
4. G is an $(0.203n, 5/2)$-vertex expander.

Fix any such graph G and note for the record that the first three properties here are hereditary and hence holds for an arbitrary subgraph of G.

To prove the bound (26) for the graph G, we will show that the polytope $FPM(G)[\Delta_{G,\ell}]$ is w-integral, and as in the previous section we are going to apply Theorem 3.4. This time $\mathcal{J} \subseteq \Delta_{G,\ell}$ will consist of all sets of edges that are of the form $N(S)$, where S is (an, c)-closed (see Definition 4.18) and $|S| \leq \ell$. Then property a) in the statement of Theorem 3.4 follows from Lemma 4.19. Namely, given a set of edges E with $|E| \leq w$, pick an arbitrary set of vertices U with $|U| \leq w$ such that $E \subseteq N(U)$. Apply Lemma 4.19 with

$c' = c,\ r = 0.203n,\ c = 5/2,\ r' = an$; (24) is guaranteed by the first term in (25). This gives us the desired set $S \supseteq U$ of cardinality at most $\frac{d}{5/2-c}w$ which is $\leq \ell$ by the second term in (25).

Let us now check the second condition b) in Theorem 3.4. Fix an (an, c)-closed S and a restriction ρ with $\sup(\rho) = N(S)$ that is consistent with $FPM(G)$. The latter condition in particular implies that $\rho^{-1}(1)$ is an S-matching, and the polytope $FPM(G)|_\rho$ is isomorphic to $FPM(G')$ for $G' = G-(S\cup D)$, where D is the set of vertices in $\partial(S)$ covered by this S-matching. Moreover, $\rho(\Delta_{G,\ell}) \subseteq \Delta_{G',\ell}$. By Definition 4.18, G' is an (an, c)-expander, and G' inherits from G the three hereditary properties above. We are left to show that

$$FPM(G')[\Delta_{G',\ell}] \neq \emptyset. \tag{27}$$

By Lemma 4.20, $FPM(G')$ contains a point x with $||x||_\infty \leq \frac{1-\varepsilon}{2}$. We claim that in fact $x \in FPM(G')[\Delta_{G',\ell}]$. Indeed, by Definition 3.1 we have to prove that $x \in FPM(G')[N(U)]$ for any $U \subseteq V(G')$ with $|U| \leq \ell$. Applying Lemma 4.19 again, this time with $r = an,\ r' = a'n$, we find an $(a'n, c')$-closed set $V(G') \supseteq S' \supseteq U$ (where c' is as above) such that $|S'| \leq L$, and it suffices to show that $x \in FPM(G')[N(S')]$ or, equivalently,

$$x_{N(S')} \in \left(FPM(G')_{N(S')}\right)_{0-1}. \tag{28}$$

We note that all assumptions of Lemma 4.24 are satisfied for $G = G'$ and $r = bn$, hence

$$x_{N(S')} \in M_{S'}(G)_{N(S')}. \tag{29}$$

So, in order to finish the proof we only have to show that

$$M_{S'}(G)_{N(S')} = \left(FPM(G')_{N(S')}\right)_{0-1}. \tag{30}$$

In the $\supseteq$ direction it is obvious, and in the opposite direction (which is actually the one we need) it amounts, as we already noticed above, to showing that any S'-matching, viewed as a restriction ρ' with $\sup(\rho') = N(S')$, is consistent with $FPM(G')$. This is done simply by re-using Lemma 4.20. First, $FPM(G')|_{\rho'} = FPM(G'')$, where $G'' \stackrel{\text{def}}{=} G' - (S' \cup D')$ for some $D' \subseteq \partial(S')$. Since S' is $(a'n, c')$-closed, G'' is an $(a'n, c')$-expander and hence, by Lemma 4.20, $FPM(G'') \neq \emptyset$ (the bound on $||x||_\infty$ is irrelevant this time).

To re-cap the argument, we have proved (30) which, along with (29), gives (28). This proves $FPM(G')[\Delta_{G',\ell}] \neq \emptyset$ and, in particular, gives us (27).

Thus, we have verified the condition b) in Theorem 3.4, and this completes the proof of Theorem 4.14.■

5. Concrete lower bounds: feasible case

In this section we present applications of Theorem 3.4 in the context of combinatorial optimization, that is to proving integrality gaps (see Definition 2.14) for feasible problems. As we will see in Section 7.3, in the infeasible case proving that $P[\Delta] \neq \emptyset$ can be a bit challenging. In the feasible case, however, since $P_{0-1} \subseteq P[\Delta]$ for any P and any Δ, all we have to do is to exhibit an integral point in P. This is not the whole story since, unlike before, we still have to show an integrality gap for the w-obstructing body we have constructed. But we can often re-use for this purpose rank lower bounds that are abundant in the literature using the following

Fact 5.1 *For any polytope $P \subseteq [0,1]^n$ and $\ell \leq n$, we have $N^\ell(P) \subseteq P[[n]^{\leq \ell}]$.*

Proof. Immediate from Definition 3.1 and Proposition 2.4.■

These two simple observations allow us to routinely transfer integrality gaps for bounded rank procedures into integrality gaps in the bounded width context. An interesting feature is *hardness amplification*, when integrality gaps for the LS-hierarchy (and potentially for any system that is sufficiently strong to satisfy Proposition 2.4, like LS_0) lead to integrality gaps for the width hierarchy based on much stronger N_+-cuts.

Let us now see a few applications of this "transference principle".

Example 3 (vertex cover and independent set) For a graph $G = (V, E)$, its *vertex cover polytope* $VC(G)$ is determined within $[0,1]^V$ by the set of additional constraints $x_u + x_v \geq 1$ ($(u,v) \in E$). MINIMUM VERTEX COVER is the minimization problem specified by the goal function $g(G) \stackrel{\text{def}}{=} \sum_{v \in V} x_v$.

We apply Theorem 3.4 with $\Delta = \mathcal{J} = [n]^{\leq w}$, where $n \stackrel{\text{def}}{=} |V|$ and w is arbitrary. Given $U \subseteq V$ and a restriction $\rho = (U, a)$ that is consistent with $VC(G)$, ρ must satisfy the constraints $x_u + x_v \geq 1$ for $(u,v) \in E(U)$. Now, $P|_\rho$ is feasible since setting $x_v = 1$ for $v \notin U$ satisfies all the remaining constraints. Hence the polytope $P[[n]^{\leq w}]$ is w-obstructing for NS_+, and therefore the result from [36], along with Fact 5.1 implies the bound

$$\mathrm{IGap}_{LS^+}(VC(G_n), g(G_n), \delta n) \geq 2 - \varepsilon, \tag{31}$$

where $\varepsilon > 0$ is an arbitrary constant, $\delta = \delta(\varepsilon) > 0$ and G_n are the graphs from [36].

Another prominent problem in combinatorial optimization (dual to the MINIMUM VERTEX COVER) is the MAXIMUM INDEPENDENT SET $IS(G)$ defined in $[0,1]^{V(G)}$ by additional constraints $x_u + x_v \leq 1$ $((u,v) \in E)$ and the same goal function $g(G) = \sum_{v \in V} x_v$, except that now it is viewed as a maximization problem. It is analyzed in exactly the same way as the MINIMUM VERTEX COVER, the only difference is that the variables x_u $(u \notin U)$ are now set to 0 rather than to 1. This again allows us to apply to it the transference principle and convert known rank bounds for this problem (see e.g. [38]) to respective lower bounds on $\mathrm{IGap}^{\max}_{LS^+}(MC(G_n), g(G_n), w)$.

Example 4 (Max Cut etc.) Let again $G = (V, E)$ be an arbitrary graph, but this time we introduce two groups of variables: x_v $(v \in V)$ and y_e $(e \in E)$. The *MAX-CUT polytope* $MC(G)$ is the polytope $P_{A,0^{|V|}}$ (in the notation of Section 4.1) corresponding to the following system A of $\mathbb{F}_2$-linear equations:

$$y_{(u,v)} = x_u \oplus x_v \ ((u,v) \in E). \tag{32}$$

This is the maximization problem with the goal function $g_{MC}(G_n) \stackrel{\text{def}}{=} \sum_{e \in E} y_e$.

Let us call a set of variables Z *closed* if along with every y_{uv} $((u,v) \in E)$ it also contains x_u, x_v. Let $\Delta \stackrel{\text{def}}{=} [n]^{\leq 3w}$ and $\mathcal{J}$ consist of all closed sets of cardinality $\leq 3w$. Then any $J \in [n]^{\leq w}$ can be extended to a closed $\widehat{J} \in \mathcal{J}$ simply by adding relevant x-variables. Moreover, every restriction ρ of a closed set of variables that is consistent with $MC(G)$ can be extended to an assignment satisfying all constraints (32). Namely, we first assign all unassigned x-variables arbitrarily, and then assign the remaining y-variables according to (32). Hence $MC(G)[[n]^{\leq 3w}]$ is w-obstructing for NS_+, and from the same paper [36] we have the bound

$$\mathrm{IGap}^{\max}_{LS^+}(MC(G_n), g_{MC}(G_n), \delta n) \leq \frac{1}{2} + \varepsilon,$$

where ε, δ and G_n have the same meaning as in (31).

An analogous argument can be applied to a host of optimization problems that share the ground polytope $MC(G_n)$ with MAX-CUT. The list includes many prominent problems in combinatorial optimization, like SPARSEST CUT or c-BALANCED SEPARATOR.

Example 5 (MAX-SAT) Given a CNF τ with variables $x_1, \ldots, x_n$ and clauses $C_1, \ldots, C_m$, we consider the CNF $\widehat{\tau} \stackrel{\text{def}}{=} (C_1 \vee \bar{y}_1) \wedge (C_2 \vee \bar{y}_2) \wedge \ldots \wedge (C_m \vee \bar{y}_m)$ in the variables $x_1, \ldots, x_n, y_1, \ldots, y_m$; let $MSAT(\tau) \stackrel{\text{def}}{=} P_{\widehat{\tau}}$ be its associated polytope. The MAX-SAT problem is the maximization problem with the goal function $\sum_{i=1}^m y_i$, and whenever τ is a k-CNF, it is called kMAX-SAT problem.

This problem (and its numerous variants obtained by varying the set of admissible constraints) is treated similarly to MAX-CUT. Namely, we call a set of variables Z *closed* if along with every y_i it also contains all x-variables appearing in C_i. In order to extend a $MSAT(\tau)$-consistent restriction ρ with closed support, we set all remaining y-variables to 0, and all remaining x-variables arbitrarily. This shows that for a k-CNF τ, $MSAT(\tau)[[n]^{\leq (k+1)w}]$ is w-obstructing for NS_+. Hence, by utilizing e.g. the result from [35] we get

$$\mathrm{IGap}^{\max}_{LS^+}(MSAT(\tau_n), g_{MS}(\tau_n), \delta(\varepsilon)n) \leq \frac{7}{8} + \varepsilon,$$

for some sequence τ_n of 3-CNFs.

6. A tradeoff between width and rank

In this section we prove Theorem 2.12. Our construction is identical to [34], although the reasoning is somewhat different: as we already noted, we have not been able to prove a common generalization of [34, Theorem 3.1] and Theorem 2.12 even if it is clear how it should look, see Section 8 for more details.

Definition 6.1 ([34]) For a clause C or a CNF τ in the variables $y_1, \ldots, y_m$ and a 0-1 $m \times n$ matrix A, we let $C[A]$ [$\tau[A]$] be the CNF in new variables $x_1, \ldots, x_n$ obtained from C [τ, respectively] by the $\mathbb{F}_2$-linear substitution $y_i \mapsto \bigoplus X_i(A)$ (recall from Section 4.1 that $X_i(A) = X_{J_i(A)}$) followed by expanding the resulting functions as CNFs in such a way that for every clause F appearing in $C[A]$ we have

$$Vars(F) = \bigcup \{ X_i(A) \mid y_i \in Vars(C) \}. \tag{33}$$

The *depth* $D(\Pi)$ of a resolution derivation Π is the height (the number of edges in the longest path) of its underlying tree. For a CNF τ and a clause C, we let $D(\tau \vdash C)$ be the minimum possible depth of a resolution

derivation of C from τ ($D(\tau \vdash C) \stackrel{\text{def}}{=} \infty$ if C is not derivable). This is the exact (and, surprisingly, not well-studied) propositional analogue of rank in semi-algebraic proof systems we are interested in this section. More precisely, Fact 2.11 (in which we set $w = n$) readily implies that

$$D(\tau \vdash C) \leq r \Longrightarrow (P_\tau)^{(r)} \subseteq P_C \text{ and } N_+^r(P) \subseteq P_C.$$

Theorem 6.2 *Let τ_m be an arbitrary unsatisfiable CNF in the variables $y_1 \ldots, y_m$, and let A be an $m \times n$ $(r, 4)$-boundary expander for some r. Then for any $\ell \geq 0$, any one of the two facts $P^{(\ell, r/4)}_{\tau_m[A]} = \emptyset$ or $N^\ell_{+,r/4}(P_{\tau_m[A]}) = \emptyset$ (see Definition 2.6) implies*

$$D(\tau_m \vdash 0) \leq \frac{r}{2}\ell.$$

Remark. Theorem 6.2 also holds for the mixed hierarchy combining Gomory-Chvátal and N_+-cuts.

Proof of Theorem 6.2. As in [34], we formulate a more general claim so that we can reason by induction. Call a set J of columns *closed* (cf. the proof of Theorem 4.3) if $A \setminus J$ is an $(r/2, 5/2)$-boundary expander. A restriction ρ is *closed* if $\sup(\rho)$ is closed. Let $x(\rho) \in [0, 1]^n$ be the extension of a restriction ρ obtained by letting $x(\rho)_j = 1/2$ whenever $\rho(x_j) = *$.

Claim 6.3 *In the set-up of Theorem 6.2, assume that $\rho = (J, a)$ is a closed restriction such that*

$$x(\rho) \notin P^{(\ell, r/4)}_{\tau_m[A]} \cap N^\ell_{+,r/4}(P_{\tau_m[A]}).$$

Define $C = C(\rho, A)$ as the (uniquely determined) clause with $Vars(C) = \{ y_i \mid i \in Ker(J) \}$ such that $C[A]|_\rho \equiv 0$. Then

$$D(\tau_m \vdash C) \leq \frac{r}{2}\ell. \tag{34}$$

Note for the record that Claim 6.3 immediately implies Theorem 6.2 if we let ρ be the empty restriction (in which case $C = 0$).

Proof of Claim 6.3. By induction on ℓ.

Base $\ell = 0$, that is $x(\rho) \notin P_{\tau_m[A]}$. Then $x(\rho) \notin P_{\widetilde{C}[A]}$ for some clause $\widetilde{C}$ appearing in τ_m. This in turn means that $x(\rho) \notin P_F$ for some clause F appearing in the CNF $\widetilde{C}[A]$. We claim that in fact $\widetilde{C}$ is a subclause of C which makes the bound (34) vacuous. Indeed, since J is closed and $A \setminus J$

has boundary expansion $5/2 > 2$, for any $i \notin \mathrm{Ker}(J)$ at least two variables in $X_i(A) \setminus J$ are set by $x(\rho)$ to $1/2$. Since $x(\rho) \notin P_F$, we conclude that $\widetilde{C}$ may not contain literals of such variables, that is $Vars(\widetilde{C}) \subseteq \{\, y_i \mid i \in \mathrm{Ker}(J) \,\} = Vars(C)$. Also, both $C[A]$ and $\widetilde{C}[A]$ are set by ρ to 0. Hence C and $\widetilde{C}$ must be consistent, which precisely means that $\widetilde{C}$ is a sub-clause of C.

Inductive step $\ell > 0$. Our assumption implies that $x(\rho)$ is violated by a cut $e^T x \leq d$ for the polytope $P^{(\ell-1,r/4)}_{\tau_m[A]} \cap N^{\ell-1}_{+,r/4}(P_{\tau_m[A]})$ (of either type) that has width $\leq r/4$. Let J' be the set of non-zero positions in e. Since $|J'| \leq r/4$, we can apply Lemma 4.4 with $c = 4$, $c' = 7/2$ and $J = J'$ and conclude that there exists $J'' \supseteq J'$ with $|\mathrm{Ker}(J'')| \leq r/2$ (the bound (20) is irrelevant for our current purposes) and such that $A \setminus J''$ is an $(r/2, 7/2)$-boundary expander. Let ρ' be the restriction obtained from ρ by un-assigning all values $\rho(x_j)$ for $j \in J \setminus J''$. Since ρ and ρ' agree on $X_{J'}$, $x(\rho')$ is also violated by the chosen cut $e^T x \leq d$. As an aside, let us note that it is this un-assigning step that makes the main novelty compared to traditional rank lower bounds proofs: in the latter, ρ can only increase in the course of the argument.

By (the contrapositive of) Proposition 2.5, there exist $j_0 \notin J \cap J''(= \mathrm{sup}(\rho'))$ and $\epsilon \in \{0, 1\}$ such that

$$x(\rho')^{(j_0,\epsilon)} \notin P^{(\ell-1,r/4)}_{\tau_m[A]} \cap N^{\ell-1}_{+,r/4}(P_{\tau_m[A]}).$$

Set

$$\widehat{J} \stackrel{\mathrm{def}}{=} J'' \cup \{j_0\}, \tag{35}$$

and let ρ^+ be the restriction that extends ρ' by additionally assigning x_{j_0} to ϵ so that $x(\rho')^{(j_0,\epsilon)} = x(\rho^+)$. We thus have

$$x(\rho^+) \notin P^{(\ell-1,r/4)}_{\tau_m[A]} \cap N^{\ell-1}_{+,r/4}(P_{\tau_m[A]}). \tag{36}$$

Removing one extra column j_0 from an $(r/2, 7/2)$-expander $A \setminus J''$ results in an $(r/2, 5/2)$-expander. Hence $\widehat{J}$ is closed. Also, $\mathrm{Ker}\left(\widehat{J}\right) = \mathrm{Ker}\,(J'')$ since for any $i \notin \mathrm{Ker}\,(J'')$ the size of the set $J_i(A) \setminus J''$ must be at least $\lceil 7/2 \rceil = 4$.

For the time being let us fix an arbitrary clause $\widehat{C}$ in the variables $\left\{ y_i \,\middle|\, i \in \mathrm{Ker}\left(\widehat{J}\right) \right\}$ that agrees with C on their common variables $\left\{ y_i \,\middle|\, i \in \mathrm{Ker}\left(\widehat{J}\right) \cap \mathrm{Ker}(J) \right\}$. Our goal is to find an extension $\widehat{\rho}$ of the restriction ρ^+ to all the remaining variables $X_{J'' \setminus (J \cup \{j_0\})}$ in $X_{\widehat{J}}$ in such a way that the assumptions of Claim 6.3 will be fulfilled with $J = \widehat{J}$, $\rho = \widehat{\rho}$, $\ell = \ell - 1$, $C = \widehat{C}$. Let us first note that this would suffice to finish the inductive step. Indeed, by the inductive

assumption, we would have that $D(\tau_m \vdash \widehat{C}) \leq \frac{r}{2}(\ell-1)$. But since $\widehat{C}$ is an *arbitrary* clause in the variables $\left\{ y_i \,\middle|\, i \in \mathrm{Ker}(\widehat{J}) \right\}$ consistent with C, we can glue together all these derivations to derive C in additional depth

$$\left|\mathrm{Ker}\left(\widehat{J}\right) \setminus \mathrm{Ker}(J)\right| \leq \left|\mathrm{Ker}\left(\widehat{J}\right)\right| = |\mathrm{Ker}(J'')| \leq r/2. \tag{37}$$

Thus, it remains to prove the existence of an extension $\widehat{\rho}$ of ρ^+ that has these two properties:

$$x\left(\widehat{\rho}\right) \;\notin\; P^{(\ell-1,r/4)}_{\tau_m[A]} \cap N^{\ell-1}_{+,r/4}(P_{\tau_m[A]}); \tag{38}$$

$$\widehat{C}[A]|_{\widehat{\rho}} \;=\; 0. \tag{39}$$

Let us first examine (39). We can view $\widehat{\rho}$ as an element of the $\mathbb{F}_2$-linear vector space $\{0,1\}^{\widetilde{J}}$, where $\widetilde{J} \stackrel{\mathrm{def}}{=} J'' \setminus (J \cup \{j_0\})$ is the set of (indices of) the variables to be assigned. The condition (39) then means that for every $i \in \mathrm{Ker}\left(\widehat{J}\right)$ we have

$$\bigoplus \left\{ \widehat{\rho}(x_j) \,\middle|\, j \in J_i(A) \cap \widetilde{J} \right\} = b_i, \tag{40}$$

where b_i are some constants fully determined by the sign with which y_i appears in $\widehat{C}$, as well as by the already known part ρ^+. Note that if also $i \in \mathrm{Ker}(J)$ then $X_i(A) \subseteq J \cap J''$ and hence $x_{j_0} \notin X_i(A)$. Thus, ρ and ρ^+ are consistent on $X_i(A)$ and since we also know that y_i occurs with the same sign in C and $\widehat{C}$, $C[A]|_\rho = 0$ implies that (40) has the form 0=0, whenever $i \in \mathrm{Ker}(J)$. We conclude that the set of all extensions $\widehat{\rho}$ satisfying (39) is an $\mathbb{F}_2$-linear subspace in $\{0,1\}^{\widetilde{J}}$ determined by those constraints (40) for which $i \in \mathrm{Ker}\left(\widehat{J}\right) \setminus \mathrm{Ker}(J)$.

Claim 6.4 *Every non-trivial affine form in its dual space, i.e., in the affine subspace* **generated** *by the constraints* (40)*, has width* ≥ 2*.*

Proof of Claim 6.4. For any non-empty $I \subseteq \mathrm{Ker}(\widehat{J}) \setminus \mathrm{Ker}(J)$, the sum (over $\mathbb{F}_2$) of the corresponding constraints (40) contains all variables x_j with $j \in \partial_A(I) \cap \widetilde{J} = \partial_{A \setminus J}(I) \setminus \{j_0\}$ as they do not cancel. But since by our original assumption $A \setminus J$ is an $(r/2, 5/2)$-boundary expander, $|\partial_{A\setminus J}(I)| \geq 3$, and this proves our claim.■

Now, Claim 6.4, along with elementary duality, immediately implies that if we take the uniform distribution $\widehat{\boldsymbol{\rho}}$ on the set of all solutions of the system

(40), then for any particular $j \in \widetilde{J}$, $\widehat{\boldsymbol{\rho}}$ takes values 0 and 1 with probability 1/2 each. Hence $\mathbf{E}[x(\widehat{\boldsymbol{\rho}})] = x(\rho^+)$ and now (36) implies that (38) is also true for at least one vertex $\widehat{\rho}$ of this polytope.

This gives us the desired $\widehat{\rho}$, and, as we already noted, we can now complete the inductive step by applying to these restrictions the inductive assumption. This completes the proof of Claim 6.3.■

As we also noted above, Theorem 6.2 is a special case of Claim 6.3.■

Proof of Theorem 2.12. As in [34], our starting point is a slightly weaker[6] form of the following result by Ben-Sasson, Impagliazzo and Wigderson:

Proposition 6.5 *There exists an increasing sequence $\{\tau_m\}$ of 4-CNF contradictions such that $w(\tau_m \vdash 0) \leq 6$, but $D(\tau_m \vdash 0) \geq \Omega(m/\log m)$.*

Let us first do the case $k \leq O(1)$ in Theorem 2.12. In that case let A_m be the $m \times 3m$ matrix in which all sets $J_i(A)$ are disjoint and have cardinality 3 each.

It is easy to see that the proof of Theorem 6.2 can be adapted to this matrix even if it is only a $(m, 3)$-boundary expander. Namely, a set of columns turns out to be closed iff it is a (disjoint) union of sets of the form $J_i(A)$. Finding a closed J'' is trivial, with the better bound

$$|\mathrm{Ker}(J'')| \leq |J'| \leq r/4. \tag{41}$$

One minor difference is that in (35) we should let $\widehat{J}$ be the *closure* of the set in the right-hand side, that is along with j_0 we also append to J' its two twin columns; that may increase the size of $\mathrm{Ker}\left(\widehat{J}\right)$ by 1. But the estimate (37) will still hold as now we have the better bound (41).

Remark. For this particular matrix we do not need the "shrinking" step $\rho \mapsto \rho'$ and we could simply let instead $J'' \stackrel{\mathrm{def}}{=} J$. That would have resulted in the stronger conclusion $D(\tau_m \vdash C) \leq r$ *regardless* of the width of the original $CP+LS^+$-proof. This kind of techniques is quite well-known in proof complexity under the name "substitution formulas" (see e.g. [30, Section 2.4]) but we preferred to stick to the letter of the proof of Theorem 2.12 in order to not complicate things any further.

Thus, when $k \leq O(1)$, Theorem 2.12 is witnessed by the CNF $\tau_m[A_m]$, where τ_m are the 4-CNFs from Proposition 6.5. Hence for the rest of the proof we assume w.l.o.g. that k is arbitrarily large.

[6]The original result was formulated for tree-like resolution size.

We need the following slight variation of Proposition 4.2 (the differences are as follows: k need not be a constant any longer, but we impose a stronger condition on the expansion rate c).

Proposition 6.6 ([34, Lemma 2.2]) *Let $n \to \infty$ and m, s, c, r be arbitrary integer parameters possibly depending on n such that $c \le \frac{3}{4}s$ and*

$$r \le o(n/s) \cdot m^{-\frac{2}{s-c}}.$$

Then for sufficiently large n there exist $m \times n$ (r, s, c)-boundary expanders.

With this proposition, the proof of Theorem 2.12 is completed almost literally as in [34]. Namely, we set $w = n^{1-\varepsilon}/k$, $r = 4w$ and $s = \lfloor k/4 \rfloor$. Since k is arbitrarily large, we have $s \ge 6$, and we can apply Proposition 6.6 with $c = 4$. This gives us an $m \times n$ $(r, s, 4)$-boundary expander A with $m \ge (n/kw)^{\Omega(k)} \ge n^{\Omega(k)}$. Recalling once more that k is arbitrarily large, we can assume that $m \ge n^2$. The required CNF is $\tau_m[A]$, where τ_m is again provided by Proposition 6.5; $D(\tau_m \vdash 0) \ge \Omega(m/\log m)$. The width 6 refutation of τ_m can be converted into a width $O(k)$ refutation of $\tau_m[A]$ simply by applying the operator $C \mapsto C[A]$ to its lines. On the other hand, if either of the two hierarchies (4), (5) converges within ℓ steps, Theorem 6.2 implies that $\ell \ge \Omega\left(\frac{D(\tau_m \vdash 0)}{r}\right)$. The proof of Theorem 2.12 is now completed by the calculation

$$\frac{D(\tau_m \vdash 0)}{r} \ge \Omega\left(\frac{m}{r \log m}\right) \ge \Omega\left(\frac{m}{n \log m}\right) \overset{\text{since } m \ge n^2}{\ge} m^{\Omega(1)} \ge n^{\Omega(k)}.$$

■

7. Lemmas

In this section we prove auxiliary statements deferred from Section 4. We group them by topic rather than by the order of appearance in that section.

7.1. Closure properties of expanders

Here we prove Lemmas 4.4, 4.19. The former is a very minor modification of [34, Lemma 2.3], and the proof of the latter to a considerable extent goes

along similar lines. We present a self-contained proof of the easier Lemma 4.4, and then indicate how to adapt it to get Lemma 4.19.

Lemma 4.4 *Let A be an $m \times n$ (r, s, c)-boundary expander, and let $c' < c$. Then for every $J \subseteq [n]$ with $|J| \leq \frac{r}{2}(c - c')$ there exists $\widehat{J} \supseteq J$ such that $A \setminus \widehat{J}$ is an $(r/2, c')$-boundary expander, $\left|Ker\left(\widehat{J}\right)\right| \leq \frac{|J|}{c-c'}$ and*

$$\left|\widehat{J}\right| \leq |J| \left(1 + \frac{s}{c - c'}\right).$$

Proof. Define a strictly increasing sequence of sets of columns $J_0 \subset J_1 \subset \ldots \subset J_t \subset \ldots$ as follows. Let $J_0 \stackrel{\text{def}}{=} J$. For $t > 0$, we first let S_t be an arbitrary set of rows in $A \setminus J_{t-1}$ violating the $(r/2, c')$-boundary expansion condition if such a set exists; otherwise, the construction terminates. Then we let

$$J_t \stackrel{\text{def}}{=} J_{t-1} \cup \bigcup_{i \in S_t} J_i(A).$$

Note that since the chain $J_0 \subset J_1 \subset \ldots \subset J_t \ldots$ is strictly increasing, the process does terminate at some point, and let J_T be the set at which it happens. We will prove that $\widehat{J} = J_T$ has all the required properties.

Claim 7.1 *For every $t = 0, 1, \ldots, T$ we have $|Ker(J_t)| \leq \frac{|J|}{c-c'}$.*

Proof of Claim 7.1. By induction on t. Assume that $|\mathrm{Ker}(J_{t-1})| \leq \frac{|J|}{c-c'}$ where for the uniformity of notation we let $J_{-1} \stackrel{\text{def}}{=} \emptyset$ and $S_0 = \emptyset$. Since $|S_t| \leq r/2$, $|\mathrm{Ker}(J_{t-1})| \leq \frac{|J|}{c-c'} \leq r/2$ and $\mathrm{Ker}(J_{t-1}) \cup S_t \subseteq \mathrm{Ker}(J_t)$, we can choose a set of rows I such that $\mathrm{Ker}(J_{t-1}) \cup S_t \subseteq I \subseteq \mathrm{Ker}(J_t)$ and

$$|I| = \min(r, |\mathrm{Ker}(J_t)|). \tag{42}$$

Applying to I the expansion condition, we get

$$|\partial_A(I)| \geq c|I|. \tag{43}$$

On the other hand, $\mathrm{Ker}(J_t) \supseteq I \supseteq \mathrm{Ker}(J_{t-1}) \cup S_t$ implies that

$$\partial_A(I) \subseteq J \cup \bigcup_{u=1}^{t} \partial_{A \setminus J_{u-1}}(S_u).$$

Since S_u's violate the $(r/2, c')$-boundary expansion conditions in their respective matrices, we conclude that

$$|\partial_A(I)| \leq |J| + c' \sum_{u=1}^{t} |S_u| \leq |J| + c'|I|. \tag{44}$$

Comparing (43) and (44), we see that $|I| \leq \frac{|J|}{c-c'}$. Since $\frac{|J|}{c-c'} \leq r/2$, the minimum in (42) must be realized by the second term, and hence $|\mathrm{Ker}(J_t)| = |I| \leq \frac{|J|}{c-c'}$. This completes the proof of Claim 7.1.■

Now, the bound $\left|\mathrm{Ker}\left(\widehat{J}\right)\right| \leq \frac{|J|}{c-c'}$ is simply Claim 7.1 with $t = T$. The required bound on $\left|\widehat{J}\right|$ is also immediate from the construction: since $\widehat{J}$ is obtained from J by adding *entire* rows $J_i(A)$, we conclude that

$$\left|\widehat{J}\right| \leq |J| + s \cdot \left|\mathrm{Ker}\left(\widehat{J}\right)\right| \leq |J| \left(1 + \frac{s}{c-c'}\right).$$

■

Recall that (r, c)-closed sets in simple graphs G were introduced in Section 4.2 (see Definition 4.18).

Lemma 4.19. *Let G be an (r, c)-vertex expander, and let $c' < c$, $r' < r$. Then for every $U \subseteq V(G)$ with*

$$|U| \leq \frac{(r-r')(c-c')}{\Delta(G)} \tag{45}$$

there exists an (r', c')-closed set $S \supseteq U$ with

$$|S| \leq \frac{\Delta(G)}{c-c'}|U|.$$

Proof. Similarly to the previous proof, we build up an increasing sequence $U_0 \subset U_1 \subset \ldots \subset U_t \subset \ldots$ of sets of vertices starting with $U_0 \stackrel{\text{def}}{=} U$. We let $D_t \subseteq \partial_G(U_{t-1})$ be such that $G - (U_{t-1} \cup D_t)$ is not an (r', c')-expander, pick an arbitrary set of vertices S_t violating this property (note for the record that $|S_t| \leq r'$ and $S_t \cap (U_{t-1} \cup D_t) = \emptyset$) and let $U_t \stackrel{\text{def}}{=} U_{t-1} \cup S_t$. Note (and this is crucial for the argument) that D_t is *not* included into U_t. As in the previous proof, we continue in this way for as long as possible, and take as $\widehat{U}$

the final set U_T in the chain we have constructed. It only remains to prove that

$$|U_{t-1}| \le \frac{\Delta(G)}{c-c'}|U| \tag{46}$$

implies $|U_t| \le \frac{\Delta(G)}{c-c'}|U|$, as this gives us $\left|\widehat{U}\right| \le \frac{\Delta(G)}{c-c'}|U|$ by induction (the base follows from $\frac{\Delta(G)}{c-c'} \ge \frac{\Delta(G)}{c} \ge 1$).

The assumption (45), along with the inductive assumption (46) imply $|U_{t-1}| \le r - r'$. Hence $|U_t| \le r$, and we can apply to it the expansion property in the original graph G:

$$|\partial_G(U_t)| \ge c|U_t|. \tag{47}$$

The crucial observation is that

$$\partial_G(U_t) \subseteq \partial_G(U) \cup \bigcup_{u=1}^{t} \partial_{G-(U_{u-1}\cup D_u)}(S_u). \tag{48}$$

Let us check this.

Fix $v \in \partial_G(U_t)$, and let $u \in [0,t]$ be the *minimum* index for which $v \in \partial_G(U_u)$. If $u = 0$, we are done. Otherwise $v \in \partial_G(S_u)$ and $v \notin D_u \cup U_{u-1}$ since $D_u \subseteq \partial_G(U_{u-1})$ and $U_{u-1} \subseteq U_t$. Hence $v \in \partial_{G-(U_{u-1}\cup D_u)}(S_u)$, which proves (48).

The rest is standard: (48) gives the bound

$$|\partial_G(U_t)| \le \Delta(G)\cdot|U| + c'\sum_{u=1}^{t}|S_u| = \Delta(G)\cdot|U| + c'|U_t|.$$

from which the required inequality $|U_t| \le \frac{\Delta(G)}{c-c'}|U|$ follows immediately given (47).■

7.2. Random regular graphs

In this section we prove Lemmas 4.15 and 4.17.

Lemma 4.15. *For every integer constant $d > 0$ and any $\varepsilon > 0$ there exists $b > 0$ such that $\boldsymbol{G_{n,d}}$ is (bn, ε)-sparse with probability $1 - o(1)$.*

Proof. First, by a union bound we have

$$\mathbf{P}[\boldsymbol{G_{n,d}} \text{ is not } (bn,\varepsilon)\text{-sparse}] \le \sum_{h=1}^{\lfloor bn\rfloor}\sum_{U\in[n]^h}\mathbf{P}[|E\left(\boldsymbol{G_{n,d}}|_U\right)| \ge (1+\varepsilon)h]\,. \tag{49}$$

Fix $1 \le h \le \lfloor bn \rfloor$ and an individual $U \in [n]^h$. Recall that $\boldsymbol{G_{n,d}}$ is the random variable $G(\boldsymbol{P_{n,d}})$ conditioned by the event that this projected graph is simple. Whenever $G(\boldsymbol{P_{n,d}})$ is simple,

$$|E\left(\boldsymbol{G_{n,d}}|_U\right)| = \left|\boldsymbol{P_{n,d}}|_{\widehat{U}}\right|,$$

where $\widehat{U} \stackrel{\text{def}}{=} \bigcup_{u \in U} V_u$ is the union of the corresponding cells; $\left|\widehat{U}\right| = dh$. As $G(\boldsymbol{P_{n,d}})$ is simple with probability $\Omega(1)$, we have

$$\mathbf{P}[|E\left(\boldsymbol{G_{n,d}}|_U\right)| \ge (1+\varepsilon)h] \le O\left(\mathbf{P}\left[\left|\boldsymbol{P_{n,d}}|_{\widehat{U}}\right| \ge (1+\varepsilon)h\right]\right). \tag{50}$$

In order to save on tiresome manipulations with factorials (we will have an opportunity to practise them in the proof of our next lemma), we apply a simple, and most likely well-known trick. Order all dn vertices in such a way that $\widehat{U}$ makes an initial segment of length dh in this order, and build a random perfect matching step by step, using the following Markov chain M. Its states are matchings (not necessarily perfect), and in a state marked with a non-perfect matching, M picks up the *minimum* unmatched vertex and matches it with another vertex picked uniformly at random among all other unmatched vertices. Then, by symmetry, in $(dn)/2$ steps this Markov chain converges to the *uniform* distribution on perfect matchings, i.e., precisely to $\boldsymbol{P_{n,d}}$.

The set $\boldsymbol{P_{n,d}}|_{\widehat{U}}$ is completely built within at most $\left|\widehat{U}\right| = dh$ steps. After ℓ steps we have at most $dh - \ell$ unmatched vertices inside $\widehat{U}$, and at least $d(n-h)-\ell$ choices outside $\widehat{U}$. Thus, the probability that the new edge added by M will also have the second endpoint in $\widehat{U}$ is bounded by $\frac{dh-\ell}{d(n-h)-\ell}$. Let $x \stackrel{\text{def}}{=} h/(n-h)$; note that $x \le b/(1-b)$. We can also assume w.l.o.g. that $b \le \frac{1}{d+1}$; this in particular implies that $x \le 1/d \le 1$ and thus $\frac{dh-\ell}{d(n-h)-\ell} \le x$. Hence

$$\mathbf{P}\left[\left|\boldsymbol{P_{n,d}}|_{\widehat{U}}\right| \ge (1+\varepsilon)h\right] \le \mathbf{P}[\boldsymbol{S_{x,hd}} \ge (1+\varepsilon)h], \tag{51}$$

where $\boldsymbol{S_{p,m}}$ is the sum of m independent Bernoulli random variables taking values 1 with probability p. Since $x \le 1/d$, we know that $\mathbf{E}[\boldsymbol{S_{x,hd}}] \le h$ and we can bound the right-hand side in (51) by the Chernoff-Hoeffding bound [22, (2.1)]. We conclude that

$$\mathbf{P}[\boldsymbol{S_{x,hd}} \ge (1+\varepsilon)h] \le 2^{-D((1+\varepsilon)/d||x)hd}, \tag{52}$$

where

$$D(p||q) \stackrel{\text{def}}{=} p \log_2 \frac{p}{q} + (1-p) \log_2 \frac{1-p}{1-q}$$

is the *Kullback-Leibler divergence*.

Let now $b \to 0$ (for fixed d, ε), so that also $x \to 0$. Then

$$D\left((1+\varepsilon)/d||x\right) \geq \frac{(1+\varepsilon)}{d} \log_2 \frac{1}{x} - O(1)$$

and thus, recalling that $x = \frac{h}{n-h}$, we get

$$D\left((1+\varepsilon)/d||x\right) hd \geq \left\{(1+\varepsilon)H(x) - O(x)\right\} n,$$

where

$$H(x) \stackrel{\text{def}}{=} x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}$$

is the *binary entropy function*. Therefore, for sufficiently small b (and hence x as well) we have the bound

$$D\left((1+\varepsilon)/d||x\right) hd \geq \left(1 + \frac{\varepsilon}{2}\right) H(x)n \geq \left(1 + \frac{\varepsilon}{2}\right) H(h/n)n. \tag{53}$$

Putting together (49), (50), (51), (52) and (53), we get

$$\mathbf{P}[\boldsymbol{G_{n,d}} \text{ is not } (b,\varepsilon)\text{-sparse}] \leq O\left(\sum_{h=1}^{\lfloor bn \rfloor} \sum_{U \in [n]^h} \binom{n}{h} \cdot 2^{-\left(1+\frac{\varepsilon}{2}\right)H(h/n)n}\right).$$

But $\binom{n}{h} \leq 2^{H(h/n)n}$ (see e.g. [4, Lemma 4.7.1]), and hence we are only left to prove that

$$\sum_{h=1}^{\lfloor bn \rfloor} 2^{-\frac{\varepsilon}{2}H(h/n)n} \leq o(1). \tag{54}$$

The first term in this sum is $O(n^{-\varepsilon/2})$. Also, since we can assume w.l.o.g. that $H'(x) \geq 1$ for $x \in [0, b/(1-b)]$, $H((h+1)/n) \geq H(h/n) + 1/n$, and hence the sum decays as a geometric progression.

This proves (54) and completes the proof of Lemma 4.15.■

As we mentioned before, the most non-trivial part in the proof of Lemma 4.17 is to show that $\boldsymbol{G_{n,d}}$ is an (an, c)-expander for appropriate values of d, a, c, and we could in principle employ for this purpose the same Markov

chain trick as in the previous proof. That, however, would have resulted in a significantly sub-optimal bound on the degree d. Instead, we do an unpretentious factorial calculation in the style of Bassalygo ([6], bipartite graphs with a slightly different random model) and Bollobás ([10], edge expansion). We do not aim at a complete analysis of the resulting expression as it turns out to be even messier than in the two previous cases just mentioned.

We need some more preliminaries. Recall that for $p_1, \dots, p_r \geq 0$ with $\sum_i p_i = 1$, $H[p_1, \dots, p_r]$ is the binary entropy of the corresponding distribution:

$$H[p_1, \dots, p_r] = \sum_{i=1}^{r} p_i \log_2 \frac{1}{p_i}$$

(thus, $H(p) = H[p, 1-p]$). The bound $\binom{n}{pn} \leq 2^{H(p)n}$ used in the proof of Lemma 4.15 is actually two-sided (see e.g. the same source [4, Lemma 4.7.1]):

$$\frac{2^{H(p)n}}{\sqrt{8pn}} \leq \binom{n}{pn} \leq 2^{H(p)n}. \tag{55}$$

The right-hand side here can be generalized, by an obvious induction, to multinomial coefficients as follows:

$$\binom{n}{p_1 n \dots p_r n} \leq 2^{H[p_1,\dots,p_r]n}. \tag{56}$$

Theorem 7.2 *Assume that $c \geq 1$ and $d > 2(1+c)$ are fixed constants, with d being integer. Consider the function*

$$\begin{aligned}\alpha(x,y) \stackrel{\text{def}}{=}\ & H[x, cx, 1-x-cx] \\ & + d\left(cxH(y/cx) + y - H(x) + \frac{1}{2}H[x-y, 2y, 1-x-y] - (1-x)H(y/(1-x))\right),\end{aligned}$$

that is well defined in the region $0 \leq y \leq x \leq \frac{1}{1+c}$.

a) *For $x \in [0, 1/(1+c)]$ we have $c^2x - 4cx - 4x + 4 > 0$. Let*

$$y_0(x) \stackrel{\text{def}}{=} \frac{2cx^{3/2}}{\sqrt{x}(2+c) + \sqrt{c^2x - 4cx - 4x + 4}}.$$

Then $y_0(x)$ is a root of the equation $\frac{\partial}{\partial y}\alpha(x,y) = 0$ lying on the interval $[0, x]$.

b) *Let also* $0 \le a \le \frac{1}{1+c}$*, and assume that* $x \in (0,a] \Longrightarrow \alpha(x, y_0(x)) < 0$*. Then* $\boldsymbol{G_{n,d}}$ *is an* (an, c)*-vertex expander with probability* $1 - o(1)$*.*

Proof. a) Let $L \stackrel{\text{def}}{=} c^2x - 4cx - 4x + 4$. L is a linear function in x, $L|_{x=0} = 4$ and $L|_{x=1/(c+1)} = \frac{c^2}{c+1}$, which implies $L > 0$ whenever $x \in [0, 1/(c+1)]$. $y_0 \ge 0$ is obvious. Next,

$$\frac{\partial \alpha}{\partial y} = \frac{d}{2\ln 2} \cdot \ln \frac{(x-y)(cx-y)^2}{(1-x-y)y^2},$$

and a direct calculation proves that $\frac{\partial \alpha}{\partial y}|_{y=y_0} = 0$. Finally, $x - y_0 = \beta \cdot (\sqrt{x}(2-c) + \sqrt{L})$, where $\beta = \frac{x}{\sqrt{x}(2+c)+\sqrt{L}} \ge 0$. From this, $y_0 \le x$ is immediate when $c \le 2$. For $c \ge 2$ we do one more radical manipulation (recall that $x \in [0, 1/(1+c)]$): $\sqrt{x}(2-c) + \sqrt{L} = \frac{4(1-2x)}{\sqrt{x}(c-2)+\sqrt{L}} \ge 0$.

b) $\boldsymbol{G_{n,d}}$ is *not* an (an, c)-vertex expander if and only if there exists $h \in \{1, 2, \ldots, \lfloor an \rfloor\}$ and two disjoint sets of vertices U, V such that $|U| = h$, $|V| = \lceil ch \rceil - 1$ and $\partial(U) \subseteq V$. For any fixed h, the number of overall choices of the pair (U, V) is $\binom{n}{h \;\; \lceil ch \rceil - 1 \;\; n-h-\lceil ch \rceil + 1}$. Fix any such U, V, and let $\widehat{U}, \widehat{V}$ be the corresponding sets of cells.

We further partition $\mathbf{P}\left[\partial_{\boldsymbol{P_{n,d}}}(\widehat{U}) \subseteq \widehat{V}\right]$ according to the number $\delta_{\boldsymbol{P_{n,d}}}(\widehat{U})$ of cross-edges between $\widehat{U}$ and $\widehat{V}$ in the matching $\boldsymbol{P_{n,d}}$. For any fixed ℓ, the number of configurations with $\partial_{\boldsymbol{P_{n,d}}}(\widehat{U}) \subseteq \widehat{V}$ and $\left|\delta_{\boldsymbol{P_{n,d}}}(\widehat{U})\right| = \ell$ is equal to

$$\binom{dh}{\ell}\binom{d(\lceil ch \rceil - 1)}{\ell}\ell! N\left(\frac{dh-\ell}{2}\right) N\left(\frac{d(n-h)-\ell}{2}\right),$$

where $N(t)$ is the number of perfect matchings on $2t$ vertices:

$$N(t) \stackrel{\text{def}}{=} \frac{(2t)!}{t!} 2^{-t}.$$

Putting it together,

$$\begin{aligned}
&\mathbf{P}[\boldsymbol{G_{n,d}} \text{ is not an } (an, c) - \text{vertex expander}] \\
&\le \sum_{h=1}^{\lfloor an \rfloor} \sum_{\substack{\ell \in \{0,\ldots,hd\} \\ hd-\ell \text{ even}}} \frac{\binom{n}{h \;\; \lceil ch \rceil - 1 \;\; n-h-\lceil ch \rceil + 1}\binom{dh}{\ell}\binom{d(\lceil ch \rceil - 1)}{\ell}\ell! N\left(\frac{dh-\ell}{2}\right) N\left(\frac{d(n-h)-\ell}{2}\right)}{N\left(\frac{dn}{2}\right)} \\
&= \sum_{h=1}^{\lfloor an \rfloor} \sum_{\substack{\ell \in \{0,\ldots,hd\} \\ hd-\ell \text{ even}}} \binom{n}{h \;\; \lceil ch \rceil - 1 \;\; n-h-\lceil ch \rceil + 1} \cdot \binom{d(\lceil ch \rceil - 1)}{\ell} \cdot 2^{\ell} \frac{\binom{\frac{dn}{2}}{\frac{dh-\ell}{2} \;\; \frac{d(n-h)-\ell}{2} \;\; \ell}}{\binom{dn}{dh}\binom{d(n-h)}{\ell}}.
\end{aligned}$$

Let $x \stackrel{\text{def}}{=} h/n$ and $y \stackrel{\text{def}}{=} \ell/(dn)$ so that $0 \le y \le x \le \frac{1}{1+c}$. Then, applying the bounds (55) and (56), we obtain

$$\begin{cases} \mathbf{P}[\boldsymbol{G_{n,d}} \text{ is not an } (an, c) - \text{vertex expander}] \\ \quad \le \sum_{x \in \{1/n, 2/n, \ldots, a\}} \sum_{y \in \{0, (dn)^{-1}, \ldots, x\}} O(nx) 2^{\alpha(x,y)n}. \end{cases} \tag{57}$$

Next,

$$\frac{\partial^2}{\partial y^2} \alpha(x, y) = \gamma \cdot Q(x, y),$$

where $\gamma = \frac{d}{2 \ln 2 (x-y)(cx-y)y(1-x-y)} \ge 0$, and

$$Q(x, y) = 2cx^3 + 2cx^2 y - 2cxy^2 - 2cx^2 + cxy - 2xy^2 + y^2$$

is a polynomial that is quadratic in y. $\frac{\partial Q}{\partial y}$ is linear in y, $\frac{\partial Q}{\partial y}|_{y=0} = cx(1+2x) \ge 0$ and $\frac{\partial Q}{\partial y}|_{y=x} = x(1-2x)(2+c) \ge 0$, hence $\frac{\partial Q}{\partial y} \ge 0$ for all $y \in [0, x]$. Also, $Q(x, x) = -x^2(1-2x)(c-1) \le 0$, hence $Q(x, y) \le 0$ for all $y \in [0, x]$. In other words, $\alpha(x, y)$ is concave in y on the interval $[0, x]$. Since $y_0(x)$ is a root of $\frac{\partial \alpha}{\partial y}$ lying on this interval, it is also the global maximum of $\alpha(x, y)$, that is $\alpha(x, y) \le \alpha(x, y_0(x))$ for all $0 \le y \le x \le \frac{1}{c+1}$, and hence, by our assumption $\alpha(x, y) < 0$ whenever $x \ge 0$. By compactness, for any fixed $\varepsilon > 0$ there exists a constant $\delta(\varepsilon) > 0$ such that $\alpha(x, y) \le -\delta(\varepsilon)$ whenever $x \ge \varepsilon$. Hence (since the pair (h, ℓ) may take only $O(n^2)$ values), (57) implies the required expansion property with probability $1 - o(1)$ for all U with $|U| \ge \varepsilon n$, where ε is an arbitrary constant of our choice.

We still have to analyze the case of very small h. For that we note the behavior of the function $\alpha(x, y_0(x))$ when $x \to 0$:

$$\lim_{x \to 0} \frac{\alpha(x, y_0(x))}{H(x)} = 1 + c - \frac{d}{2}. \tag{58}$$

which is negative since by our assumption $d > 2(1 + c)$. Now the proof for small values of h is completed by the same argument as in the proof of Lemma 4.15: the extra term $O(nx)$ in (57) is constant when the decaying geometric progression (54) starts, and hence its contribution to the sum is negligible.■

Corollary 7.3 *For any fixed $d \ge 14$, $\boldsymbol{G_{n,d}}$ is an $(0.204n, 5/2)$-vertex expander with probability $1 - o(1)$.*

Proof. For $d = 14$ and $x \in (0, 0.204]$, the condition $\alpha(x, y_0(x)) < 0$ is checked by plotting. Since $H[x, cx, 1 - x - cx] \geq 0$, $\alpha(x, y) < 0$ for $d = 14$ implies that $\alpha(x, y) < 0$ for all $d \geq 14$. Thus, all requirements of Theorem 7.2 are fulfilled.■

Lemma 4.17 *Let $d \geq 14$ and $g \geq 0$ be arbitrary integer constants. Then with probability $1 - o(1)$ as $n \to \infty$, $\boldsymbol{G_{n,d}}$ contains a spanning subgraph G with girth$(G) \geq g$ that is an $(0.203n, 5/2)$-vertex expander.*

Proof. G will be obtained from $\boldsymbol{G_{n,d}}$ simply by removing all cycles of length $\leq (g-1)$. Clearly, girth$(G) \geq g$. The number of such cycles is (say) $O(\log n)$, see e.g. [8], which means that for any set U, $\partial_{\boldsymbol{G_{n,d}}}(U)$ and $\partial_G(U)$ differ by a set of size $O(\log n)$. Hence Corollary 7.3 implies the required expansion condition for all sets U with $|U| \geq \varepsilon n$, where ε is an arbitrary constant of our choice. Moreover, if ε is sufficiently small, then by (58), $\boldsymbol{G_{n,d}}$ is actually an $(\varepsilon, 5)$-expander, with probability $1 - o(1)$.

But Lemma 4.15 also implies that all cycles of length $\leq (g-1)$ are vertex disjoint (otherwise, $\boldsymbol{G_{n,d}}$ could not be $\left(2g-3, \frac{2g-2}{2g-3}\right)$-sparse). Hence we have removed at most two edges adjacent to every vertex, and therefore, since $\boldsymbol{G_{n,d}}$ is an $(\varepsilon n, 5)$-expander, G is an $(\varepsilon n, 3)$-expander.■

7.3. Non-emptiness of the polytopes $P[\Delta]$

In this section we prove Lemmas 4.5, 4.20.

Lemma 4.5 *Let A be an $m \times n$ (r, s, c)-boundary expander, where $r \geq 2$ and $rc \geq 2(s+1)$, and let*

$$\ell \stackrel{\text{def}}{=} \lfloor \frac{rc}{2} \rfloor - s - 1.$$

Then for any $b \in \{0,1\}^m$ we have $N_+\left(P_{A,b}[[n]^{\leq \ell}]\right) \neq \emptyset$.

Proof. Let

$$L \stackrel{\text{def}}{=} \lfloor rc/2 \rfloor = s + \ell + 1 \geq 2.$$

Let us call a $\mathbb{F}_2$-linear constraint $f(x) = b$ *derivable* if its width $w(f)$ is at most L, and it can be represented as a $\mathbb{F}_2$-linear combination of at most $r/2$ initial constraints (19).

Claim 7.4 **a)** *All initial constraints* (19) *are derivable.*

b) *If $f = b$ and $f' = b'$ are derivable* **and** $w(f \oplus f') \leq L$ *then $f \oplus f' = b \oplus b'$ is derivable.*

c) $0 = 1$ *is not derivable.*

Proof of Claim 7.4. a) Note that $s \leq L$. Since $r \geq 2$, every initial constraint (19) forms its own derivation.

b) The constraint $f \oplus f' = b \oplus b'$ can be represented as a $\mathbb{F}_2$-linear combination of a set I of initial constraints with $|I| \leq r$. Hence $|\partial_A(I)| \geq c \cdot |I|$. Since all variables in $X_{\partial_A(I)}$ must appear in $f \oplus f'$, we have $w(f \oplus f') \geq |\partial_A(I)| \geq c \cdot |I|$. Thus, in fact $|I| \leq L/c \leq r/2$.

c) is obvious.■

Claim 7.4 implies that derivable constraints of width ≤ 2 induce the following nice structure on the set of variables $\{x_1, \ldots, x_n\}$. We have a subset $X_0 \subseteq \{x_1, \ldots, x_n\}$ consisting of those variables x_j for which there exists a derivable constraint of the form $x_j = \epsilon_j$, for a uniquely (due to part c)) defined ϵ_j. On the remaining set $\{x_1, x_2, \ldots, x_n\} \setminus X_0$ we have an equivalence relation $\approx$ such that for any $x_j \approx x_{j'}$ there exists a derivable constraint of the form $x_j \oplus x_{j'} = \epsilon_{jj'}$, where the constants $\epsilon_{jj'}$ satisfy the consistency relations $\epsilon_{jj'} \oplus \epsilon_{j'j''} = \epsilon_{jj''}$ $(x_j \approx x_{j'} \approx x_{j''})$. Finally, every derivable constraint of width ≤ 2 is either one of $x_j \oplus x_{j'} = \epsilon_{jj'}$, or is the sum of at most two derivable constraints $x_j = \epsilon_j$ with $x_j \in X_0$.

Negating some of the variables x_j if necessary, we can assume w.l.o.g. that all constants $\epsilon_j, \epsilon_{jj'}$ are equal to 0. For every equivalence class C of $\approx$ introduce a new propositional variables y_C. Consider the propositional projection π (see Definition 3.2) defined by

$$\pi(x_j) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } x_j \in X_0 \\ y_C & \text{if } x_j \in C, \end{cases}$$

and apply it to the system (19). Let $\widehat{A}Y = \widehat{b}$ be the matrix representation of the resulting system (it is easy to see from Claim 7.4 that this system does not contain trivial/contradictory equations 0=0, 1=0). Since the operators $P \mapsto P[\Delta]$ and $P \mapsto N^+(P)$ behave well with respect to projections (see (18) and Lemma 3.3), it suffices to show that $N_+(P_{\widehat{A},\widehat{b}}([[n]^{\leq \ell}])) \neq \emptyset$. We will in fact prove that

$$y_{1/2} \in N_+(P_{\widehat{A},\widehat{b}}([[n]^{\leq \ell}])), \tag{59}$$

where $y_{1/2}$ is the vector in which all coordinates are equal to 1/2.

Denote by D the set of all derivable constraints (in the original x-variables), and consider the set $\widehat{D}$ obtained by applying π to the elements of D in a natural way, followed by all possible cancellations. As it turns out, $\widehat{D}$ enjoys all the properties in Claim 7.4, with an extra addition in item c) that is very crucial for our purposes.

Claim 7.5 **a)** *All initial constraints of the system $\widehat{A}Y = \widehat{b}$ are in $\widehat{D}$.*

b) *if $g = b$ and $g' = b'$ are in $\widehat{D}$ and $w(g \oplus g') \leq L$ then $g \oplus g' = b \oplus b'$ is in $\widehat{D}$.*

c) $(0 = 1)$ *is not in $\widehat{D}$ and, moreover, every non-trivial constraint in $\widehat{D}$ has width ≥ 3.*

Proof of Claim 7.5. a) is obvious.

b) Let $f(x_1, \ldots, x_n) = b$ and $f'(x_1, \ldots, x_n) = b'$ be derivable constraints such that $\pi(f) = g$ and $\pi(f') = g'$. y_C occurs in g if and only if f contains an odd number of variables in C. Pick arbitrarily a representative $x_C \in Vars(f)$ in any such class C. Then we note that the constraint $\widetilde{f} \stackrel{\text{def}}{=} \oplus_{y_C \in Vars(g)} x_C = b$ is derivable (and, clearly, $\pi(\widetilde{f}) = g$). Indeed, we can repeatedly use Claim 7.4b) to remove from f, one by one, all variables $x_j \in X_0$ as well appropriate pairs $\{x_j, x_{j'}\}$ with $x_j \approx x_{j'}$ until we are left with $\widetilde{f}$. As the variables only get removed, the width condition is never violated. Apply the same procedure to g' to get $\widetilde{f}'$. Finally, if a variable y_C appears in both g and g', and its pre-images x_C, x'_C turn out to be different in $\widetilde{f}$ and $\widetilde{f}'$, we replace in $\widetilde{f}$, using Claim 7.4b) one more time, x_C with x'_C. Repeating this for all clauses C, we end up with $\widetilde{f}^*$, $\widetilde{f}'^*$ such that $\pi(\widetilde{f}^*) = g$, $\pi(\widetilde{f}'^*) = g'$ and $w(\widetilde{f}^* \oplus \widetilde{f}'^*) = w(g \oplus g') \leq L$. Applying Claim b) one last time, we see that $f \oplus f' = b \oplus b'$ is in D and hence $g \oplus g' = b \oplus b'$ is in $\widehat{D}$.

c) As we have just seen, every constraint $(g = b)$ in $\widehat{D}$ has a pre-image $f = b$ in D with $w(f) = w(g)$. Now the first statement follows from Claim 7.4c), and the width bound follows from the fact that the projection π trivializes all derivable constraints of width ≤ 2.■

With Claim 7.5 at our disposal, finishing the proof of (59) is routine (cf. the last part of the proof of Claim 6.3). By Proposition 2.5, it suffices to show that $y_{1/2}^{(C_0,\epsilon)} \in P_{\widehat{A},\widehat{b}}([[n]^{\leq \ell}])$ for any equivalence class C_0 and ϵ. By Claim

7.5, parts a) and c), all equations in the system $\widehat{A}Y = \widehat{b}$ have width ≥ 3, hence $y_{1/2}^{(C_0,\epsilon)} \in P_{\widehat{A},\widehat{b}}$.

Fix now any set $\mathcal{C}$ of equivalence classes with $|\mathcal{C}| \leq \ell$, and let $\mathcal{C}_0 \stackrel{\text{def}}{=} \mathcal{C} \cup \{C_0\}$. It remains to prove that $\left(y_{1/2}^{(C_0,\epsilon)}\right)|_{\mathcal{C}_0} \in \left(P_{\widehat{A},\widehat{b}}|_{\mathcal{C}_0}\right)_{0-1}$.

For that purpose, let us consider the set D_0 of all constraints $g = b$ in $\widehat{D}$ that satisfy $Vars(g) \subseteq \mathcal{C}_0$. Since $|\mathcal{C}_0| \leq \ell + 1 \leq L$, this set is actually a $\mathbb{F}_2$-linear subspace (by Claim 7.4b)) that, moreover, does not contain any element of width ≤ 2 by Claim 7.4c). Add to it one extra constraint $y_{C_0} = \epsilon$; the resulting linear subspace $\mathbb{A}$ will not contain constraints of width ≤ 1. Hence $\left(y_{1/2}^{(C_0,\epsilon)}\right)|_{\mathcal{C}_0}$ is equal to the uniform convex combination of all restrictions ρ in the dual solution space $\mathbb{A}^*$ of the system $\mathbb{A}$, and we only have to prove that every $\rho \in \mathbb{A}^*$ can be extended to an element in $P_{\widehat{A},\widehat{b}}$.

We inductively define increasing sets of y-variables (identified, for better readability, with their indices) $\mathcal{C}_0 \subset \mathcal{C}_1 \subset \ldots \subset \mathcal{C}_T$ as follows: while there exists a row i_t in the matrix $\widehat{A}$ with $\left|J_{i_t}(\widehat{A}) \setminus \mathcal{C}_{t-1}\right| = 1$ (say, $J_{i_t}(\widehat{A}) \setminus \mathcal{C}_{t-1} = \{C_t\}$), we arbitrarily pick any such i_t and let

$$\mathcal{C}_t \stackrel{\text{def}}{=} \mathcal{C}_{t-1} \cup \{C_t\}.$$

Let us remark that while this construction is analogous to those used in Section 7.1, this time we do *not* have any a priori bound on the number of steps T; neither we need one.

Claim 7.6 *For every $1 \leq t \leq T$ there exists a constraint $g = b$ in $\widehat{D}$ such that $y_{C_t} \in Vars(g) \subseteq \mathcal{C}_0 \cup \{C_t\}$.*

Proof of Claim 7.6. We only have to note that since $\left|J_{i_t}(\widehat{A}) \cup \mathcal{C}_0\right| \leq s+\ell+1 = L$, the set of all constraints $(g = b) \in \widehat{D}$ with $Vars(g) \subseteq J_{i_t}(\widehat{A}) \cup \mathcal{C}_0$ makes a linear subspace. Now the claim is proved by an obvious induction on t.■

Using Claim 7.6, we recursively extend ρ by picking $\rho(y_{C_t})$ in such a way that it satisfies the constraint from that claim; note that this value does not depend on the choice of the constraint. Let $\widehat{\rho}$ be the resulting restriction with $\sup(\widehat{\rho}) = \mathcal{C}_T$ and, as in Section 6, let $x(\widehat{\rho})$ be obtained from $\widehat{\rho}$ by extending it with values $1/2$ outside of $\mathcal{C}_T$. We claim that in fact $x(\widehat{\rho}) \in P_{\widehat{A},\widehat{b}}$.

Indeed, by our construction every row $J_i(\widehat{A})$ has either ≥ 2 or 0 variables outside of $\mathcal{C}_T$. In the first case the corresponding (real) constraint (15) is

satisfied by any two variables in $J_i(\widehat{A}) \setminus \mathcal{C}_T$. In the second case we represent the corresponding initial constraint as a $\mathbb{F}_2$-linear combination of constraints stipulated by Claim 7.6 and a constraint $(g = b) \in \widehat{D}$ with $Vars(g) \subseteq \mathcal{C}_0$. As ρ satisfies all of them, Lemma 4.5 is proved.■

For Lemma 4.20, let us first note one simple combinatorial property of expanders.

Lemma 7.7 *Let G be an (an, c)-vertex expander, where $a \leq 1/4$, and let $c' < 4ac$ (so that in particular $c' < c$). Then for every disjoint sets of vertices $S, D \subseteq V(G)$ with $|S| \leq n/2$, for sufficiently large n we have*

$$|N(S) \setminus N(D)| \geq c'|S| - 2|D|.$$

In particular ($D = \emptyset$),

$$|N(S)| \geq c'|S|,$$

and

$$c' \geq 2 \Longrightarrow |N(S) \setminus N(D)| \geq c'(|S| - |D|).$$

Proof. If $|S| \leq an$ then we have the following, much stronger, bound:

$$|\delta(S) \setminus N(D)| \geq |\partial(S) \setminus D| \geq c|S| - |D|. \tag{60}$$

Thus, let us assume $|S| \geq an$, and let $\ell \stackrel{\text{def}}{=} \lfloor 2a|S| \rfloor$. Note that $\ell \leq \frac{|S|}{2}$ (since $a \leq 1/4$) and $\ell \leq an$ (since $|S| \leq n/2$).

Let A and B be two disjoint subsets of S, of cardinality ℓ each. Then, using (60), we obtain the estimate

$$\begin{cases} |N(S) \setminus N(D)| \geq |N(A \cup B) \setminus N(D)| \\ \quad = |N(A) \setminus N(D)| + |N(B) \setminus N(D)| - |E(A,B)| \\ \quad = |\delta(A) \setminus N(D)| + |E(A)| + |\delta(B) \setminus N(D)| \\ \qquad +|E(B)| - |E(A,B)| \\ \quad \geq 2(c\ell - |D|) + |E(A)| + |E(B)| - |E(A,B)|. \end{cases} \tag{61}$$

Randomizing over all possible choices of the pair (A, B), we see that the expectation of the quantity $|E(\boldsymbol{A}, \boldsymbol{B})| - |E(\boldsymbol{A})| - |E(\boldsymbol{B})|$ is $O\left(\frac{|E(S)|\ell}{n^2}\right)$, which is $O(1)$ since we can assume w.l.o.g. that $|E(S)| \leq O(n)$ (otherwise the

statement is trivial). Applying (61) to the pair minimizing this quantity, we immediately get

$$|N(S) \setminus N(D)| \geq 2(2ac|S| - |D|) - O(1) \geq c'|S| - 2|D|$$

since $c' < 4ac$.■

Lemma 4.20. *Let a, c be constants such that $c > 2$ and $ac > 1/2$. Then there exists $\varepsilon > 0$ such that for sufficiently large n and for any (an, c)-vertex expander G the polytope $FPM(G)$ contains a point x with $||x||_\infty \leq \frac{1-\varepsilon}{2}$.*

Proof. Decreasing a if necessary, we can assume that $a \leq 1/4$. Choose c' such that $2 < c' < 4ac \leq c$, and let

$$\varepsilon \stackrel{\text{def}}{=} 1 - \frac{2}{c'}.$$

We have to prove that the following linear program is feasible:

$$\begin{cases} 0 \leq x_e \leq \frac{1}{c'} \ (e \in E(G)) \\ x(\delta(v)) = 1 \ (v \in V(G)). \end{cases}$$

By a standard duality argument, it suffices to show that for any *fixed* real weights w_v $(v \in V(G))$, there exists $x \in [0, 1/c']^{E(G)}$ that satisfies

$$\sum_{v \in V(G)} w_v x(\delta(v)) \geq \sum_{v \in V(G)} w_v. \tag{62}$$

Let $V(G) = [n]$ and assume w.l.o.g. that $w_1 \geq w_2 \geq \ldots \geq w_n$. For an edge $e = (u, v) \in E(G)$, we let $\alpha(e) \stackrel{\text{def}}{=} \min(u, v)$ and $\beta(e) \stackrel{\text{def}}{=} \max(u, v)$. In this notation, (62) can be re-written as

$$\sum_{e \in E(G)} w_{\alpha(e)} x_e + \sum_{e \in E(G)} w_{\beta(e)} x_e \geq \sum_{v=1}^{n} w_v. \tag{63}$$

Assume for the ease of notation that n is even (if n is odd, we will have to adjust by $\frac{1}{2} w_{(n+1)/2}$ each of the two right-hand sides in (64)). We split the inequality (63) into two parts and show how to construct $x \in [0, 1/c']^{E(G)}$ so that

$$\begin{cases} \sum_{e \in E(G)} w_{\alpha(e)} x_e \geq \sum_{v=1}^{n/2} w_v \\ \sum_{e \in E(G)} w_{\beta(e)} x_e \geq \sum_{v=n/2+1}^{n} w_v. \end{cases} \tag{64}$$

For that purpose we recursively construct a sequence $x^{(0)}, x^{(1)}, \ldots, x^{(n/2)}$ of vertices such that $x^{(0)} \le x^{(1)} \le \ldots \le x^{(n/2)}$ (pointwise) and all $x^{(v)}$ possess the following properties:

$$\begin{cases} x^{(v)} \in [0, 1/c']^{E(G)}; \\ ||x^{(v)}||_1 = v; \\ \left\{ e \in E(G) \,\middle|\, x_e^{(v)} > 0 \right\} \subseteq N([v]) \ (= \{ e \in E(G) \mid \alpha(e) \le v \}). \end{cases}$$

First we let $x^{(0)}$ be the zero vector. For $v > 0$, we note that $|N([v])| \ge c'v$ by Lemma 7.7, hence an extension $x^{(v)} \ge x^{(v-1)}$ with the required properties is always possible. We construct a specific $x^{(v)}$ by consecutively saturating all $x_e^{(v-1)}$ with $e \in N([v])$ to the maximum possible value $1/c'$ *giving priority to edges* $e \in N([v])$ *with smaller values of* $\beta(v)$ until we eventually increase the ℓ_1-norm of $x_e^{(v-1)}$ by 1. In this way we enforce the following property:

For every $e \in N([v])$ *we have either* $x_e^{(v)} = x_e^{(v-1)}$ *or* $\forall f \in N([v])(\beta(f) < \beta(e) \Longrightarrow x_f^{(v)} = (1/c'))$.

Let $x \stackrel{\text{def}}{=} x^{(n/2)}$; we are left to check the conditions (64).

For the first inequality, let

$$X_v = \sum_{\alpha(e) \le v} x_e.$$

Then we estimate the left-hand side as follows:

$$\sum_{e \in E(G)} w_{\alpha(e)} x_e = \sum_{v=1}^{n/2} w_v \sum_{\{ e \mid \alpha(e) = v \}} x_v = \sum_{v=1}^{n/2} w_v (X_v - X_{v-1}) = \sum_{v=1}^{n/2-1} (w_v - w_{v+1}) X_v + w_{n/2} X_{n/2}.$$

But by our construction, $X_v \ge \sum_{\alpha(e) \le v} x_e^{(v)} = v$, with equality for $v = n/2$. Hence, since $w_v - w_{v+1} \ge 0$, we further have

$$\sum_{e \in E(G)} w_{\alpha(e)} x_e \ge \sum_{v=1}^{n/2-1} (w_v - w_{v+1}) v + w_{n/2} \cdot (n/2) = \sum_{v=1}^{n/2} w_v.$$

Similarly, the same "Abelian-summation" trick reduces the second inequality in (64) to proving that for any $u \in [n/2+1, n]$, $\sum_{\beta(e) \le u} x_e \ge (u - n/2)$ or, equivalently, (since $||x||_1 = n/2$) $\sum_{\beta(e) > u} x_e \le n - u$. Assume the contrary, and let $u \in [n/2 + 1, n]$ be an arbitrary vertex for which this inequality is violated. Let $v \in [1, n/2]$ be the *first* vertex for which $\sum_{\beta(e) > u} x_e^{(v)} > n - u$.

Let $S \stackrel{\text{def}}{=} [v]$ and $D \stackrel{\text{def}}{=} \{u+1,\ldots,n\}$. Then $x_e^{(v)} > x_e^{(v-1)}$ for at least one edge with $\beta(e) > u$, which, as we noted when constructing this consequence, is possible only when all edges $e \in N(S) \setminus N(D)$ (that is, precisely those with $\alpha(e) \leq v$ and $\beta(e) \leq u$) satisfy $x_e = 1/c'$. Applying Lemma 7.7 once more, we conclude that in the vector $x^{(v)}$ these edges carry ℓ_1-norm at least $|S| - |D|$. Since $||x^{(v)}||_1 = |S|$, we have $\sum_{\beta(e)>u} x_e^{(v)} \leq |D| = n - u$. This contradiction proves the second inequality in (64) and hence Lemma (4.20).■

8. Conclusion and open questions

In this paper we have attempted to point out a somewhat strange gap in the literature on the complexity of dynamic semi-algebraic proof systems and combinatorial algorithms. Namely we have proposed to quantify the complexity of a proof/algorithm simply by its sparsity, that is the maximum number of variables involved in the cuts it is making. One way to think of this complexity measure (cf. the proof of Theorem 2.9) is by viewing the proof/algorithm as a highly parallel protocol that works independently and exhaustively for all small sets of variables J, thus producing the polytopes $(P_J)|_{0-1}$ in each of them. After that these processors M_J come to a general meeting where they exchange this information by forming the intersection $P[\Delta]$, upon which they return to their own "districts" J to refine their local polytopes on the basis of the information they gained at the meeting. These two stages repeat until there is no new information exchanged.

Our lower bounds are based on capturing this intuition by the notions of *w-integral* and *w-obstructing* polytopes. They are specified by the amount of local information that can be distributed among the processors M_J in such a way that they do not learn anything new by exchanging it. We have developed tools for constructing such polytopes that has turned out to be a not so trivial task in the infeasible case (see the proofs in Section 4.2 and 7.3). We hope that these concepts will become useful in further study of dynamic semi-algebraic proof systems/algorithms. One particular goal that we have in mind is one of the most important open problems in the area: find "direct", "combinatorial" proofs of *size* lower bounds in such systems. As we mentioned in the introduction, the only results currently available are manifestly indirect (and, in the case of LS_+, are only conditional!) and based on the feasible interpolation theorem [32, 33].

In our lower bounds for natural problems (Sections 4, 5) the parallel game described at the beginning of this section terminates in essentially[7] one round. On the other hand, our tradeoff result Theorem 2.12 displays the opposite behavior: at each of the conventions, only a small number of processors learn something new, and they need exponentially many (in w) rounds to converge. Moreover, this lower bound on the number of rounds works for widths that are much larger than the minimum width.

The combinatorial principle underlying Theorem 2.12, however, is all but natural. Are there any "natural" problems in either proof complexity or algorithms that display an "interesting" behavior in terms of the number of rounds required for them to "essentially converge" (we deliberately prefer to leave this question quantitatively loose)?

Speaking of strong tradeoffs, our tradeoff between rank and width (Theorem 2.12) is very similar to the tradeoff between resolution tree-like size and width given in [34]. Thus it would be very natural to ask if these two results can be combined into one. In other words, do there exist similar tradeoffs between tree-like *cutting plane* proof size and width? This seems to be an interesting and clean problem that is left open by our paper.

Another thing I tried for a while was to make the lower bound for the perfect matching principle, Theorem 4.14, work for the Lovász-Schrijver system as well. It looks as if the difficulties here are general and stem from the fact that the whole powerful apparatus of protection matrices, pseudoexpectations etc. works best in the presence of useful structure like a rich group of symmetries, the structure of a linear space over the finite field etc. For example, the only LS_+ lower bounds for the matching principles in the standard (width-unrestricted) rank model we are aware of either pertain to complete graphs and heavily exploit their symmetries (see e.g. [29] and the references therein) or can be achieved via a gadget reduction from Tseitin tautologies [11, Lemma 4] and hence indirectly utilize the $\mathbb{F}_2$-linear structure. In the absence of such structure, it took a considerable pain to prove even the most intuitively obvious fact about the perfect matching polytope (Lemma 4.20). We feel it might potentially turn out to be very useful to develop methods for analyzing more general (that is, less symmetric) situations, both in terms of rank and width lower bounds. Is it true that it takes $n^{\Omega(1)}$ rounds for the

[7]Strictly speaking, the w-obstructing polytopes $P[\Delta]$ used for lower bounds purposes are not exactly $P^{',w}$ or $N_{+,w}(P)$. But they can be placed between, say, $N_{+,w}(P)$ and $N_{+,Cw}(P)$ for an absolute constant $C > 0$, so we view this as a technicality for the purpose of this generic discussion.

LS or LS^+-hierarchy to converge from $FPM(\boldsymbol{G_{n,d}})$ to $PM(\boldsymbol{G_{n,d}})$ a.s.? In fact, I believe it is open even for the cutting planes hierarchy: our Theorem 4.14 does not have any direct bearing for rank lower bounds. Can Theorem 4.14 itself be generalized to the complexity measure w_{LS^+}?

On a more general note, how large can be the gap between $w_{LS^+}(\tau \vdash 0)$ and $w_{CP}(\tau \vdash 0)$ for $O(1)$-CNFs τ?

Finally, our model does not seem to have any straightforward extension to *static* proof systems like Sherali-Adams or Sum-of-Squares. There still are, however, a few dynamic proof systems of interest operating with higher-degree polynomials, notably the systems LS^d_+ LS^d, LS^d_+ introduced in [21]. Can we extend our lower bound for $\mathbb{F}_2$-linear equations, that is Theorem 4.3, to, say, LS^d for some $d > 2$? One reason why it might be interesting is because for this higher-order systems even indirect, interpolation-based size lower bounds completely fall apart [21, Section 4].

Acknowledgment

I am greatly indebted to anonymous referees of this paper for very useful remarks, and, in particular, for pointing out to me the reference [16].

References

[1] M. Alekhnovich, S. Arora, and I. Tourlakis, *Toward strong nonapproximability results in the Lovász-Schrijver hierarchy*, Computational Complexity **20** (2011), no. 4, 615–648.

[2] M. Alekhnovich, E. Ben-Sasson, A. Razborov, and A. Wigderson, *Pseudorandom generators in propositional proof complexity*, SIAM Journal on Computing **34** (2004), no. 1, 67–88.

[3] M. Alekhnovich and A. Razborov, *Lower bounds for the polynomial calculus: non-binomial case*, Proceedings of the Steklov Institute of Mathematics **242** (2003), 18–35.

[4] R. Ash, *Information theory*, Dovers Publications, 1990.

[5] B. Barak and D. Steurer, *Sum-of-squares proofs and the quest toward optimal algorithms*, Proceedings of International Congress of Mathematicians (ICM), vol. IV, 2014, pp. 509–533.

[6] L. A. Bassalygo, *Asymptotically optimal switching circuits*, Problems of Information Transmission **17** (1981), no. 3, 206–211.

[7] E. Ben-Sasson and A. Wigderson, *Short proofs are narrow - resolution made simple*, Journal of the ACM **48** (2001), no. 2, 149–169.

[8] B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European Journal of Combinatorics **1** (1980), 311–316.

[9] ———, *Random graphs*, Academic Press, London, 1985.

[10] ———, *The isoperimetric number of random regular graphs*, European Journal of Combinatorics **9** (1988), 241–244.

[11] S. Buss, D. Grigoriev, R. Impagliazzo, and T. Pitassi, *Linear gaps between degrees for the Polynomial Calculus modulo distinct primes*, Journal of Computer and System Sciences **62** (2001), 267–289.

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

[13] V. Chvátal and E. Szemerédi, *Many hard examples for resolution*, Journal of the ACM **35** (1988), no. 4, 759–768.

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

[15] S. Dash, *Exponential lower bounds on the lengths of some classes of branch-and-cut proofs*, Mathematics of Operation Research **30** (2005), no. 3, 678–700.

[16] S. S. Dey, M. Molinaro, and Q. Wang, *Approximating polyhedra with sparse inequalities*, Mathematical Programming, Ser. B **154** (2015), 329–352.

[17] J. Edmonds, *Maximum matching and a polyhedron with 0,1-vertices*, Journal of Research of the National Bureau of Standards **69** (1965), 125–130.

[18] F. Eisenbrand and A. Schulz, *Bounds on the Chvátal rank of polytopes in the 0-1 cube*, Combinatorica **23** (2003), 245–261.

[19] M. Goemans and L. Tuncel, *When does the positive semidefiniteness constraint help in lifting procedures*, Mathematics of Operation Research **26** (2001), 796–815.

[20] R. E. Gomory, *An algorithm for integer solutions of linear programs*, Recent Advances in Mathematical Programming, McGraw-Hill, 1963, pp. 269–302.

[21] D. Yu. Grigoriev, E. A. Hirsch, and D. V. Pasechnik, *Complexity of semi-algebraic proofs*, Moscow Mathematical Journal **2** (2002), no. 4, 647–679.

[22] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (1963), no. 301, 13–30.

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

[24] S. Jukna, *Boolean function complexity: Advances and frontiers*, Springer-Verlag, 2012.

[25] J. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM Journal on Optimization **11** (2001), 796–817.

[26] J. R. Lee, P. Raghavendra, and D. Steurer, *Lower bounds on the size of semidefinite programming relaxations*, Proceedings of the 47th Annual ACM Symposium on Theory of Computing, 2015, pp. 567–576.

[27] L. Lovász and M. D. Plummer, *Matching theory*, AMS Chelsea Publishing, 2009.

[28] L. Lovász and A. Schrijver, *Cones of matrices and set-functions and 0-1 optimization*, SIAM Journal on Optimization **1** (1991), no. 2, 166–190.

[29] C. Mathieu and A. Sinclair, *Sherali-Adams relaxations of the matching polytope*, Proceedings of the 41st ACM Symposium on the Theory of Computing, 2009, pp. 293–302.

[30] J. Nordström, *Pebble games, proof complexity and time-space trade-offs*, Logical Methods in Computer Science **9** (2013), 1–63.

[31] M. Pinsker, *On the complexity of a concentrator*, 7th Annual Teletraffic Conference (Stockholm), 1973, pp. 318/1–318/4.

[32] P. Pudlák, *Lower bounds for resolution and cutting planes proofs and monotone computations*, Journal of Symbolic Logic **62** (1997), no. 3, 981–998.

[33] ———, *On the complexity of propositional calculus*, Sets and Proofs, Invited papers from Logic Colloquium'97, Cambridge University Press, 1999, pp. 197–218.

[34] A. Razborov, *A new kind of tradeoffs in propositional proof complexity*, Journal of the ACM **62** (2016), no. 3, article 16.

[35] G. Schoenebeck, *Linear level Lasserre lower bounds for certain k-CSPs*, Proceedings of the 49th IEEE Symposium on Foundations of Computer Science, 2008, pp. 593–602.

[36] G. Schoenebeck, L. Trevisan, and M. Tulsiani, *A linear round lower bound for lovász-schrijver lp relaxations of Vertex Cover and Max Cut*, Proceedings of the 39th ACM Symposium on the Theory of Computing, 2007, pp. 302–310.

[37] A. Schrijver, *On cutting planes*, Annals of Discrete Mathematics **9** (1980), 291–296.

[38] M. Tulsiani, *CSP gaps and reductions in the Lasserre hierarchy*, Proceedings of the 41st ACM Symposium on the Theory of Computing, 2009, pp. 303–312.

[39] N. C. Wormald, *Models of random regular graphs*, Surveys in Combinatorics (J. D. Lamb and D. A. Preece, eds.), 1999, pp. 239–298.