Fundamenta Informaticae XXI (2001) 1–14
DOI 10.3233/FI-2012-0000
IOS Press

Tight lower bounds on the resolution complexity of perfect matching principles

Dmitry Itsykson*
Steklov Institute of Mathematics at St. Petersburg
27 Fontanka, St.Petersburg, 191023, Russia
dmitrits@pdmi.ras.ru

Vsevolod Oparin
St. Petersburg Academic University
8 Khlopina, St.Petersburg, 194021, Russia
oparin.vsevolod@gmail.com

Mikhail Slabodkin
St. Petersburg Academic University
8 Khlopina, St.Petersburg, 194021, Russia
slabodkinm@gmail.com

Dmitry Sokolov
Steklov Institute of Mathematics at St. Petersburg
27 Fontanka, St.Petersburg, 191023, Russia
sokolov.dmt@gmail.com

Abstract. The resolution complexity of the perfect matching principle was studied by Razborov [14], who developed a technique for proving its lower bounds for dense graphs. We construct a constant degree bipartite graph G_n such that the resolution complexity of the perfect matching principle for G_n is $2^{\Omega(n)}$ where n is the number of vertices in G_n. This lower bound is tight up to some polynomial. Our result implies the $2^{\Omega(n)}$ lower bounds for the complete graph K_{2n+1} and the complete bipartite graph $K_{n,O(n)}$ that improves the lower bounds following from [14]. We show that for every graph G with n vertices that has no perfect matching there exists a resolution refutation of perfect matching principle for G of size $O(n^2 2^n)$. Thus our lower bounds match upper bounds up to an application of polynomial. Our results also imply the well-known exponential lower bounds on the resolution complexity of the pigeonhole principle, the functional pigeonhole principle and the pigeonhole principle over a graph.

We also prove the following corollary. For every natural number d, for every n large enough, for every function $h : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, d\}$, we construct a graph with n vertices that has the following properties. There exists a constant D such that the degree of the i-th vertex is at least $h(i)$ and at most D, and it is impossible to make all degrees equal to $h(i)$ by removing the graph's edges. Moreover, any proof of this statement in the resolution proof system has size $2^{\Omega(n)}$. This

Address for correspondence: dmitrits@pdmi.ras.ru
*The research is partially supported by the RFBR grant 14-01-00545, by the President's grant MK-2813.2014.1 and by the Government of the Russia (grant 14.Z50.31.0030).

result implies well-known exponential lower bounds on the Tseitin formulas as well as new results: for example, the same property of a complete graph.

Preliminary version of this paper appeared in proceedings of CSR-2015 [8].

Keywords: proof complexity, expander, perfect matching, resolution, width

1. Introduction

Sometimes it is possible to represent combinatorial statements as unsatisfiable CNF formulas. For example, CNF formulas PHP_n^m encode the pigeonhole principle; PHP_n^m states that it is possible to put m pigeons into n holes such that every pigeon is contained in at least one hole and every hole contains at most one pigeon. PHP_n^m depends on variables $p_{i,j}$ for $i \in [m]$ and $j \in [n]$, and $p_{i,j} = 1$ iff the i-th pigeon is in the j-th hole. For every $i \in [m]$, PHP_n^m contains a clause $(p_{i,1} \vee p_{i,2} \vee \cdots \vee p_{i,n})$. For every $j \in [n]$ and every $k \neq l \in [n]$, PHP_n^m contains a clause $(\neg p_{k,j} \vee \neg p_{l,j})$. PHP_n^m is unsatisfiable iff $m > n$.

For an undirected graph $G(V, E)$ we define a CNF formula PMP_G that encodes the fact that G has a perfect matching. We assign a binary variable x_e for all $e \in E$. PMP_G is the conjunction of the following conditions: for all $v \in V$, exactly one edge that is incident to v has value 1. Such conditions can be written as the conjunction of the statement that at least one edge takes value 1: $\bigvee_{(v,u)\in E} x_{(v,u)}$ and the statement that for any pair of edges e_1, e_2 incident to v, at most one of them takes value 1: $\neg x_{e_1} \vee \neg x_{e_2}$. If G has no perfect matchings then PMP_G is an unsatisfiable formula.

For an unsatisfiable CNF formula φ, a resolution refutation, a proof of its unsatisfiability, in the resolution proof system is a sequence of clauses with the following properties: the last clause is an empty clause (we denote it by $\square$); every clause is either a clause of the initial formula φ, or can be obtained from previous ones by the resolution rule. The resolution rule allows to infer a clause $(B \vee C)$ from clauses $(x \vee B)$ and $(\neg x \vee C)$. The size of a resolution refutation is the number of clauses in it. It is well known that the resolution proof system is sound and complete. Soundness means that if a formula has a resolution refutation then it is unsatisfiable. Completeness means that every unsatisfiable CNF formula has a resolution refutation.

Let $K_{m,n}$ denote the complete bipartite graph with m and n vertices in its parts. Note that the formulas $\mathrm{PMP}_{K_{m,n}}$ are easier to refute in the resolution proof system then PHP_n^m, since $\mathrm{PMP}_{K_{m,n}}$ contain more clauses. Therefore any lower bound on the size of a resolution refutation of $\mathrm{PMP}_{K_{m,n}}$ implies the same lower bound on the size of a resolution refutation of PHP_n^m and, conversely, every upper bound on the resolution refutation of PHP_n^m implies the same upper bound on the size of resolution refutation of $\mathrm{PMP}_{K_{m,n}}$.

We say that a family of unsatisfiable CNF formulas F_n is weaker than a family of unsatisfiable formulas H_n if every clause of H_n is an implication of a clause of F_n. In these terms $\mathrm{PMP}_{K_{m,n}}$ is weaker than PHP_n^m. The size of any resolution refutation of H_n is at least the size of the minimal resolution refutation of F_n. Thus it is interesting to prove lower bounds for formulas as weak as possible.

1.1. Known results

Haken [6] proved the lower bound $2^{\Omega(n)}$ on the resolution complexity of PHP_n^{n+1}. Raz [11] proved the lower bound 2^{n^ϵ} on the resolution complexity of PHP_n^m for some positive constant ϵ and an arbitrary $m > n$. The latter lower bound was simplified and improved to $2^{\Omega(n^{1/3})}$ by Razborov [12].

Urquhart [16] and Ben-Sasson, and Wigderson [3] consider formulas $G\text{-}\,\mathrm{PHP}_m^n$ that are defined by a bipartite graph G; the first part of G corresponds to pigeons and consists of m vertices, and the second part corresponds to holes and consists of n vertices. Every pigeon must be contained in one of the adjacent holes. Formulas $G\text{-}\,\mathrm{PHP}_n^m$ can be obtained from PHP_n^m by substituting variables which do not have corresponding edges in G with zeroes. The paper [3] presents the lower bound $2^{\Omega(n)}$ for formulas $G\text{-}\,\mathrm{PHP}_n^m$ where $m = O(n)$ and G is a bipartite constant degree expander.

Razborov [13] considers a functional pigeonhole principle FPHP_n^m that is a weakening of PHP_n^m; the formula FPHP_n^m is the conjunction of PHP_n^m and additional conditions stating that every pigeon is contained in at most one hole. Razborov proved the lower bound $2^{\Omega\left(\frac{n}{(\log m)^2}\right)}$ for FPHP_n^m which implies a lower bound $2^{\Omega\left(n^{1/3}\right)}$ depending only on n.

Razborov [14] proved that if G has no perfect matchings then the resolution complexity of PMP_G is at least $2^{\frac{\delta(G)}{\log^2 n}}$ where $\delta(G)$ is the minimal degree of the graph and n is the number of vertices.

Alekhnovich [1], and Dantchev and Riis [5] consider the graphs of the chessboard $2n \times 2n$ without two opposite corners. The perfect matching principle for such graphs is equivalent to the possibility to tile such chessboards with domino. The strongest lower bound $2^{\Omega(n)}$ was proved in [5] and this lower bound is polynomially connected with the upper bound $2^{O(n)}$. We note that the number of variables in such formulas is $\Theta(n^2)$.

1.2. Our results

For all constant C, all n and all $m \in [n+1, Cn]$ we give an example of a bipartite graph $G_{m,n}$ with m and n vertices in its parts such that all degrees are bounded by a constant and the resolution complexity of $\mathrm{PMP}_{G_{m,n}}$ is $2^{\Omega(n)}$. The number of variables in such formulas is $O(n)$, therefore the lower bound matches (up to an application of a polynomial) the trivial upper bound $2^{O(n)}$ that holds for every formula with $O(n)$ variables. This is the first lower bound for the perfect matching principle, that is exponential in the number of variables. In particular, our results imply that the resolution complexity of $\mathrm{PMP}_{K_{m,n}}$ is $2^{\Omega(n)}$. This lower bound improves the lower bound $2^{\Omega(n/\log^2 n)}$ that follows from [14]. Due to the upper bound $O(n^3 2^n)$ that follows from the upper bound for PHP_n^{n+1} [4], this result is tight up to an application of a polynomial. Our result implies the lower bound $2^{\Omega(n)}$ on the resolution complexity of $\mathrm{PMP}_{K_{2n+1}}$ where K_{2n+1} is a complete graph on n vertices, and it is also better than the lower bound $2^{\Omega(n/\log^2 n)}$ following from [14]. We show that for every graph G with n vertices that does not have a perfect matching there exists a resolution refutation of PMP_G of size $O(n^2 2^n)$. Thus the lower and upper bounds for $\mathrm{PMP}_{K_{2n+1}}$ differ polynomially. We note that $\mathrm{PMP}_{G_{m,n}}$ is weaker than $G_{m,n}\text{-}\,\mathrm{PHP}_n^m$, PHP_n^m and FPHP_n^m, therefore our lower bound implies the same lower bound for $G_{m,n}\text{-}\,\mathrm{PHP}_n^m$, PHP_n^m and FPHP_n^m.

Our proof can be divided into two parts. Firstly, we prove lower bound on the resolution width for perfect matching principles based on bipartite graphs with certain expansion properties. To do this we modify the method introduced by Ben-Sasson and Wigderson, namely, we define a nonstandard measure

on the clauses of a resolution refutation. Secondly, we give a construction of constant degree bipartite graphs that have an appropriate expansion property. We use lossless expanders and similarly to [9] we remove vertices with high degrees from them. For example, we can use the explicit construction of lossless expanders from [10] or the randomized construction from [7]. Finally, we apply the theorem of Ben-Sasson and Wigderson stating that if a formula ϕ in O(1)-CNF has the resolution width at least w, then any resolution refutation of ϕ has the size at least $2^{\Omega(w^2/n)}$ where n is the number of variables in ϕ.

We also prove a more general result. For a graph $G(V, E)$ and a function $h : V \to \{1, 2, \dots, d\}$ we define a formula $\Psi_G^{(h)}$ encoding that $G(V, E)$ has a subgraph $H(V, E')$ such that for all v in H the degree of v equals $h(v)$. Note that if $h \equiv 1$ then $\Psi_G^{(h)}$ is precisely PMP_G. For any $d \in \mathbb{N}$, we show that there exists $D \in \mathbb{N}$ that for all n large enough and every function $h : V \to \{1, 2, \dots, d\}$, where $|V| = n$, it is possible to construct a graph $G(V, E)$ in polynomial time with degrees of vertices at most D, such that the formula $\Psi_G^{(h)}$ is unsatisfiable, and the size of any resolution refutation of $\Psi_G^{(h)}$ is at least $2^{\Omega(n)}$.

If h maps V to $\{1, 2\}$ then $\Psi_G^{(h)}$ is weaker than Tseitin formulas based on the graph G. Thus our result implies the lower bound $2^{\Omega(n)}$ on the resolution complexity of Tseitin formulas that was proved in [15].

2. Preliminaries

We consider simple graphs without loops and multiple edges. The graph G is called bipartite if its vertices can be divided into two disjoint parts X and Y in such a way that any edge is incident to one vertex from X and one vertex from Y. By $G(X, Y, E)$ we denote a bipartite graph with parts X and Y and set of edges E. A matching in a graph $G(V, E)$ is a set of edges $E' \subseteq E$ such that any vertex $v \in V$ has at most one incident edge from E'. A matching E' covers a vertex v if there exists $e \in E'$ incident to v. A perfect matching is a matching that covers all vertices of G. For a bipartite graph $G(X, Y, E)$ and a set $A \subseteq X$ by $\Gamma(A)$ we denote a set of all neighbors of vertices from A.

Theorem 2.1. (Hall)
Consider such a bipartite graph $G(X, Y, E)$ that for some $A \subseteq X$, for all $B \subseteq A$, the following inequality holds: $|\Gamma(B)| \geq |B|$. Then there exists a matching that covers all vertices from A.

In [3] E. Ben-Sasson and A. Wigderson introduced a notion of a formula width. A width of a clause is a number of literals contained in it. For a k-CNF formula φ, the width of φ is the maximum width of its clauses. A width of a resolution refutation is a width of the largest used clause.

Theorem 2.2. ([3])
For any k-CNF unsatisfiable formula φ, the size of a resolution refutation is at least $2^{\Omega\left(\frac{(w-k)^2}{n}\right)}$, where w is a minimal width of a resolution refutation of φ and n is a number of variables used in φ.

A partial substitution is a set that consists of assignments $x := a$, there x is a propositional variable and $a \in \{0, 1\}$ such that all variables are distinct. The result of the application of a partial substitution ρ to a CNF formula φ may be obtained from ϕ by the following procedure: delete all clauses from ϕ that are satisfied by ρ and delete all literals from other clauses that have common variable with some assignment from ρ.

Lemma 2.3. Let φ be a CNF formula that is obtained from an unsatisfiable CNF formula ψ by the application of a partial substitution. Then φ is unsatisfiable and the size of the minimal resolution refutation of ψ is at least the size of the minimal resolution refutation of φ.

3. Upper bound

In this section we prove that for every graph $G(V, E)$ that has no perfect matching the resolution complexity of PMP_G is at most $2^{|V|}poly(|V|)$.

We will use the classical Tutte's criterion of the existence of perfect matching:

Theorem 3.1. (Tutte, 1947)
Graph G has a perfect matching iff for any set $S \subseteq V$:

$$o(G - S) \leq |S|$$

where $G - S$ denotes the graph G without vertices from the set S and $o(G - S)$ denotes the number of connected components with odd cardinality in the obtained graph.

Theorem 3.2. If graph G on n vertices does not have a perfect matching, then the formula PMP_G has a resolution refutation of size $O(n^2 2^n)$.

The plan of the proof of Theorem 3.2 is the following:

1. Observe that if M is odd set of vertices in the graph that has a perfect matching, then every perfect matching contains at least one edge that connects M with $V \setminus M$. We give a resolution derivations of this observation for all odd sets M simultaneously of total size $O(n^2 2^n)$.

2. Tuttes theorem states that if G has no perfect matching, then there exists a set $S \subseteq V$ such that $|o(G - S)| > |S|$. We call odd components from $G - S$ pigeons and elements of S holes. We say that a pigeon M is in a hole s if there is an edge of the perfect matching that connects some vertex from M with s. On the first step we have already derived clause stating that every pigeon is in at least one hole. Every hole contains at most one pigeon by the property of perfect matchings.

3. We use the monotone refutation of the pigeonhole principle by Buss and Pitassi [4] to get a contradiction.

A monotone resolution refutation of the pigeonhole principle PHP_n^m is a sequence of clauses $C_1, C_2, \ldots, C_k$ such that for every j, C_t has only positive occurrences of variables $p_{i,j}$ for $i \in [m], j \in [n]$, C_k is the empty clause, and for every $t \in [k]$ the clause C_t is either a clause of PHP_n^m or may be obtained from previous clauses by the monotone resolution rule:

$$\frac{A \vee \bigvee\limits_{i \in I_1} p_{i,j} \qquad\qquad B \vee \bigvee\limits_{i \in I_2} p_{i,j}}{A \vee B \vee \bigvee\limits_{i \in I_1 \cap I_2} p_{i,j}}.$$

Every monotone resolution rule corresponds to one particular hole j. The monotone resolution rule implicitly uses that every hole contains at most one pigeon. Buss and Pitassi [4] showed that for the pigeonhole principle PHP_n^m monotone resolution refutations and general resolution refutations are polynomially equivalent.

Theorem 3.3. ([4])
For all $m > n$ the formula PHP_n^m has a monotone resolution refutation of size $O(n2^n)$.

Let A and B be two disjoint subsets of V. Let $E(A, B)$ be the set of edges connecting vertices from A with vertices from B. For every $F \subseteq E$ we denote a clause $W_F = \bigvee\limits_{e \in F} x_e$.

Lemma 3.4. All clauses $W_{E(M,V\setminus M)}$ for all $M \subseteq V$ with odd size can be inferred simultaneously from PMP_G with a resolution derivation of size $n^2 2^n$.

Proof:
We prove by the induction on $0 \leq k \leq \frac{n-1}{2}$ that all clauses $W_{E(M,V\setminus M)}$ for all $M \subseteq V$ with $|M| = 2k+1$ can be inferred with a resolution derivation of size $2n^2 \sum\limits_{l=0}^{k} \binom{n}{2l+1}$. The base case $k = 0$ is trivial since the required clauses are in PMP_G.

Induction step. Let $|M| = 2(k+1)+1$, if there are no edges between vertices of M then $W_{E(M,V\setminus M)}$ may be obtained by the weakening rule from the clause $W_{E(\{v\},V\setminus\{v\})}$ for any vertex $v \in M$; the latter clause is in PMP_G.

We show that if $u, v \in M$ are connected by an edge then using already derived clauses we may derive $W_{E(M,V\setminus M)} \vee \neg x_{(u,v)}$ with at most $2n-2$ applications of rules. By the induction hypothesis we already have a clause $W_{E(M\setminus\{u,v\},\{u,v\}\cup V\setminus M}$. By the weakening rule applied to $W_{E(M\setminus\{u,v\},\{u,v\}\cup V\setminus M)}$ we get a clause $D = W_{E(M,V\setminus M)} \vee W_{E(\{u,v\},V\setminus\{u,v\})}$. For every edge $e \in E(\{u,v\}, M \setminus \{u,v\})$ the original formula contains a clause $\neg x_e \vee \neg x_{(u,v)}$. We consequentially apply the resolution rule with D and all such clauses to get $W_{E(M,V\setminus M)} \vee \neg x_{(u,v)}$.

Let u be some vertex from M. Consider the clause $W_{E(\{u\},V\setminus\{u\})}$ and consequentially apply the resolution rules with clauses $W_{E(M,V\setminus M)} \vee \neg x_{(u,w)}$ for $w \in V \setminus \{u\}$ such that $(u,w) \in E$. Finally we get a clause $W_{E(M,V\setminus M)}$. For every $M \subseteq V$ with $|M| = 2k+3$ we use at most n clauses $W_{E(M,V\setminus M)} \vee \neg x_{(u,w)}$. Each of them requires a derivation of size at most $2n-2$ and on the last step we apply at most n rules. Thus in order to derive $W_{E(M,V\setminus M)}$ we add at most $(2n-2)n + n < 2n^2$ new clauses. □

Proof:
[Proof of Theorem 3.2] Graph G does not have a perfect matching. Tutte's criterion implies that there exists $S = \{s_1, \ldots, s_l\} \subseteq V$ such that $o(G - S) > |S|$. Let $C_1, C_2, \ldots, C_m$ be connected components of odd cardinality in graph $G - S$; we know that $m > l$.

By Lemma 3.4 we infer clauses $W_{E(C_i,V\setminus C_i)}$ for all $i \in \{1, 2, \ldots, m\}$. Note that $E(C_i, V \setminus C_i) = E(C_i, S)$.

Let us denote

$$\psi = \bigwedge_{i=1}^{m} W_{E(C_i,S)}$$

$$\phi = \bigwedge_{s \in S} \bigwedge_{\substack{e_1,e_2 \in E(\{s\},V\setminus S) \\ e_1 \neq e_2}} (\neg x_{e_1} \vee \neg x_{e_2}).$$

We will present a refutation of size $O(n^2 l 2^l)$ for the formula $\psi \wedge \phi$. Since $l < n/2$, $l2^l < 2^n$ when n is large enough and there exists a refutation of PMP_G of size $O(n^2 2^n)$.

By Theorem 3.3 PHP_l^m has a monotone resolution refutation $D_1, D_2, \ldots, D_r$ where $r \leq l2^l$. Let H_k be obtained from D_k by substitutions of $p_{i,j}$ by $W_{E(C_i,\{s_j\})}$ for all $i \in [m], j \in [l]$. We will show that $H_1, H_2, \ldots, H_r$ may be extended to a resolution refutation of $\psi \wedge \phi$ of size $2n^2l2^l$.

If D_k is a clause of PHP_l^m (it contains only positive occurrences of variables), then H_k is a clause of ψ. If D_k is the result of monotone resolution rule applied to D_{k_1} and D_{k_2}, then $D_{k_1} = A \vee \bigvee\limits_{i\in I_1} p_{i,j}$, $D_{k_2} = B \vee \bigvee\limits_{i\in I_2} p_{i,j}$ and $D_k = A \vee B \vee \bigvee\limits_{i\in I_1\cap I_2} p_{i,j}$.

We show that there is a resolution derivation of H_k from H_{k_1} and H_{k_2} of size $2n^2$. Let $H_{k_1} = A' \vee \bigvee\limits_{i\in I_1} W_{E(C_i,\{s_j\})}$, where A' is obtained from A by the substitutions. If for every $i \in I_2$, $p_{i,j}$ has occurrence in D_{k_1}, then H_k is the weakening of H_{k_1}. For every $i \in I_2$ such that $p_{i,j}$ has no occurrences in D_{k_1} and for all $v \in C_i$ such that $(v, s_j) \in E$ we derive $F_v = A' \vee \bigvee\limits_{i\in I_1\cap I_2} W_{E(C_i,\{s_j\})} \vee \neg x_{v,s_j}$. To derive F_v we apply the resolution rule at most n times to H_{k_1} with clauses $\neg x_{v,s_j} \vee \neg x_{u,s_j}$ for all $u \in \bigcup\limits_{i\in I_1\setminus I_2} C_i$ such that $(u, s_l) \in E$. Finally, we consequentially resolve all derived F_v with H_{k_2} and get H_k. □

4. Lower bounds for perfect matching principle

Our goal is to prove the following theorem:

Theorem 4.1. There exists a constant D such that for all $C > 1$ there exists $a > 0$ such that for all n large enough and for all $m \in [n + 1, Cn]$ it is possible to construct in polynomial in n time a bipartite graph $G(V, E)$ with parts of size m and n such that all degrees are at most D, the formula PMP_G is unsatisfiable, and the size of any resolution refutation of PMP_G is at least 2^{an}.

We note that the lower bound from Theorem 4.1 is tight up to an application of a polynomial since these formulas contain $O(n)$ variables and thus there is a trivial upper bound $2^{O(n)}$.

Corollary 4.2. For every $C > 1$, there exists $a > 0$ such that for every n and $m \in [n + 1, Cn]$ the resolution complexity of $\mathrm{PMP}_{K_{m,n}}$ is at least 2^{an} where $K_{m,n}$ is the complete bipartite graph with m and n vertices in parts.

Proof:
By Theorem 4.1 there exists a bipartite graph G with n and m vertices in parts such that the resolution complexity of PMP_G is at least 2^{an}. The formula PMP_G may be obtained from $\mathrm{PMP}_{K_{m,n}}$ by substituting zeros for the edges that do not belong to G. Therefore by Lemma 2.3, the resolution complexity of $\mathrm{PMP}_{K_{m,n}}$ is at least the resolution complexity of PMP_G. □

The lower bound from Corollary 4.2 improves the lower bound $2^{n/\log^2 n}$ that follows from [14].

Corollary 4.3. The resolution complexity of $\mathrm{PHP}_{K_{2n+1}}$ is $2^{\Omega(n)}$ where K_{2n+1} is the complete graph on $2n + 1$ vertices.

Proof:
By Theorem 4.1 there exists a bipartite graph G with n and $n+1$ vertices in parts such that the resolution complexity of PMP_G is at least 2^{an}. Formula PMP_G may be obtained from $\mathrm{PMP}_{K_{2n+1}}$ by substituting zeros for edges that do not belong to G. Therefore by Lemma 2.3 the resolution complexity of $\mathrm{PMP}_{K_{2n+1}}$ is at least the resolution complexity of PMP_G. □

The lower bound from Corollary 4.3 improves the lower bound $2^{n/\log^2 n}$ for the resolution complexity of $\mathrm{PMP}_{K_{2n+1}}$ that follows from [14].

By Theorem 3.2 the lower bounds from Corollary 4.2 and Corollary 4.3 are tight up to an application of a polynomial.

The plan of the proof of Theorem 4.1 is the following. In Section 4.1 we prove the lower bound on the resolution width of PMP_G if G is a bipartite graph which has some expansion property. In Section 4.2 we show how to construct a constant degree bipartite graphs with the appropriate expansion property. Note that if degrees of all vertices of G are at most D then PMP_G is D-CNF formula. Finally, in Section 4.3 we conclude the proof by using Theorem 2.2.

4.1. Perfect matching principle for expanders

Definition 4.4. A bipartite graph $G(X, Y, E)$ is (r, c)-boundary expander if for any set $A \subseteq X$ such that $|A| \leq r$ the following inequality holds: $|\delta(A)| \geq c|A|$ where $\delta(A)$ denotes the set of vertices in Y connected with the set A by exactly one edge.

Theorem 4.5. Let $G(X, Y, E)$ be a bipartite (r, c)-boundary expander with $c \geq 1$ and $|X| > |Y|$. Let G have a matching that covers all vertices from the part Y. Then the formula PMP_G is unsatisfiable and the width of its resolution refutation is at least $cr/2$.

Proof:
Parts X and Y have different number of vertices, hence there are no perfect matchings in G, and PMP_G is unsatisfiable.

We call an assignment to variables of PMP_G *proper* if for every vertex $v \in X$ at most one edge incident to v has value 1 and for every $u \in Y$ exactly one edge incident to u has value 1. In other words, proper assignments correspond to matchings that cover all vertices from Y. For some subset $S \subseteq X$ and for a clause C we say that S properly implies C if any proper assignment that satisfies all constraints in vertices from S, also satisfies C. We denote this as $S \vdash C$.

Now we define a measure on clauses from a resolution refutation of PMP_G: $\mu(C) = \min\{|S| \mid S \subseteq X, S \vdash C\}$.

The measure μ is very similar to the measure from [3], where the measure of a clause is the number of local conditions that imply the clause. We consider the implication only on the set of matchings that cover all vertices from Y (proper assignments). In our case conditions in vertices from Y are satisfied by every proper assignment, therefore we consider only conditions in vertices from X.

The measure μ has the following properties:

1. The measure of any clause from PMP_G equals 0 or 1.

2. Semiadditivity: $\mu(C) \leq \mu(C_1) + \mu(C_2)$, if C is obtained by applying the resolution rule to C_1 and C_2. Let $S_1 \vdash C_1$, $|S_1| = \mu(C_1)$ and $S_2 \vdash C_2$, $|S_2| = \mu(C_2)$. Hence $S_1 \cup S_2 \vdash C_1$ and $S_1 \cup S_2 \vdash C_2$, so $S_1 \cup S_2 \vdash C$, therefore $\mu(C) \leq |S_1| + |S_2| = \mu(C_1) + \mu(C_2)$.

3. The measure of the empty clause $\square$ is greater than r. To prove this property we need the following lemma.

Lemma 4.6. Let a bipartite graph $G(X, Y, E)$ have two matchings, the first one covers all vertices from $A \subseteq X$, and the second one covers all vertices from $B \subseteq Y$. Then there exists a matching in G that covers A and B simultaneously.

Proof:
[1] Let M_A and M_B be matchings that cover A and B respectively. It is sufficient to prove the statement for the subgraph $G'(X, Y, E')$, where E' is the union $M_A \cup M_B$. We call all edges from M_A black and all edges from M_B white. Some edges can have both colours.

Every vertex from G' has at most one outgoing black edge and at most one white incident edge. Hence every connected component of G' is either an isolated vertex or a simple path or a simple cycle. We prove separately for every connected component of G' that there exists a matching in this component that covers all vertices in this component from $A \cup B$. Note that every vertex from A has at least one black incident edge and every vertex from B has at least one white incident edge. Consider all cases of connected components.

- An isolated vertex: it cannot be from $A \cup B$, hence the required matching is empty.
- A simple cycle: cycle in a bipartite graph must have even length. The required matching is a set of all black (or white) edges from the cycle.
- An odd-length simple path: the required matching contains all edges from this path with odd numbers.
- An even-length simple path: the first and the last vertices from this path are from the same part of the graph and also the first and the last edges have different colours, hence either the first or the last vertex from the path is not from $A \cup B$, all other vertices can be covered by a matching as it was done in the previous case.

$\square$

Let $\mu(\square) \leq r$, then there is $S \subseteq X$ such that $S \vdash \square$ and $|S| \leq r$. For all $A \subseteq S$ the following holds: $|\Gamma(A)| \geq |\delta(A)| \geq c|A| \geq |A|$, and the Hall's Theorem (Theorem 2.1) implies that there is a matching in G that covers S. G also has a matching covering all vertices of Y, therefore Lemma 4.6 implies that there exists a matching that covers S and Y, hence it corresponds to a proper assignment that satisfies all constraints for vertices from S, but it is impossible to satisfy the empty clause, and we get a contradiction with the fact that $\mu(\square) \leq r$.

Lemma 4.7. Let $S \subseteq X$ be a minimal set that properly implies some clause C. Let $v \in Y$ have exactly one neighbour from S. Then C contains at least one edge incident to v.

[1] This proof was suggested by an anonymous reviewer.

Proof:
Let $u \in S$ be connected with v; denote edge (u, v) by f. Since $|S \setminus \{u\}| < |S|$, clause C is not properly implied from the set $S \setminus \{u\}$, i. e. there exists a proper assignment σ that satisfies all restrictions in the vertices $S \setminus \{u\}$, but refutes the clause C. Such assignment σ can not satisfy the constraint in the vertex u, since otherwise σ would satisfy S and therefore satisfy C. Since σ is a proper assignment, σ assigns value 0 to all edges that are incident with u, and σ satisfies v. There is an edge e incident to v such that $\sigma(e) = 1$. The vertex v is a boundary vertex for S, therefore the other endpoint of e does not belong to S. Consider an assignment σ' that is obtained from σ by changing the values of f and e, σ' is proper and it satisfies all constraints from S, and hence it satisfies C. Thus C contains either e or f. □

The semiadditivity of the measure implies that any resolution refutation of the formula PMP_G contains a clause C with the measure in the interval $\frac{r}{2} \leq \mu(C) \leq r$. We claim that the width of C is at least $rc/2$.

Let $S \vdash C$ and $|S| = \mu(C)$. Since G is a (r, c)-boundary expander, $\delta(S) \geq c|S|$. By Lemma 4.7 for every $v \in \delta(S)$ the clause C contains at least one edge incident to v and all such edges are distinct since $\delta(S) \subseteq Y$. Therefore the size of the clause C is at least $|\delta(S)| \geq c|S| \geq cr/2$. □

Remark 4.8. The condition in Theorem 4.5 that G has a matching covering all vertices from Y cannot be removed for free since for every (r, c)-boundary expander it is possible to add one vertex to X and $\lceil c \rceil$ vertices to Y such that the new vertex in X is connected with all new vertices in Y. The resulting graph is also an (r, c)-boundary expander, but the resulting formula will contain an unsatisfiable subformula that depends on $\lceil c \rceil + 1$ variables, hence it can be refuted with width $\lceil c \rceil + 1$. We do not know whether it is possible to replace the second condition in the theorem by a weaker condition.

4.2. Expanders

In this section we show how to construct a constant degree graph that satisfies the conditions of Theorem 4.5.

Definition 4.9. The bipartite graph G with parts X and Y is an (r, d, c)-expander, if degrees of all vertices from X do not exceed d, and for every set $I \subseteq X, |I| \leq r$ the inequality $|\Gamma(I)| \geq c|I|$ holds. Here $\Gamma(I)$ denotes the set of all vertices that are adjacent with at least one vertex from I.

Lemma 4.10. ([2])
Every (r, d, c)-expander is a $(r, 2c - d)$-boundary expander.

Proof:
Let $A \subseteq Y$, $|A| \leq r$, then $|\Gamma(A)| \geq c|A|$. The number of edges between A and $\Gamma(A)$ may be estimated: $d|A| \geq E(A, \Gamma(A)) \geq |\delta(A)| + 2|\Gamma(A) \setminus \delta(A)| = 2|\Gamma(A)| - |\delta(A)| \geq 2c|A| - |\delta(A)|$. Finally we get $|\delta(A)| \geq (2c - d)|A|$. □

We say that a family of graphs G_n is explicit if it is possible to construct G_n in polynomial in n time.

Theorem 4.11. ([10])
For every $\epsilon > 0$ and every time-constructible function $m(n)$ there exist $k \geq 1, b > 0$ and there exists an explicit construction of a family of d-regular $(\frac{n}{kd}, d, (1 - \epsilon)d)$-expanders with sizes of parts $|X| = m(n)$ and $|Y| = n$, where $d \leq \log^b(\frac{m}{n}))$.

The existence of expanders from Theorem 4.11 can also be proved by the probabilistic method. But Theorem 4.11 gives an explicit construction of such graphs.

Note that we can not use expanders from Theorem 4.11 directly since the vertices in Y may have unbounded degrees. Similarly to [9] we delete vertices with high degrees and some other vertices in such a way that the resulting graph would be a good enough expander.

Theorem 4.12. For every $C \geq 1$ and every $\epsilon > 0$, there exists $k \geq 1$, integer $d \geq 3$ and an explicit construction of a family of $(\frac{n}{kd}, d, (1-\epsilon)d)$-expanders with $|X| = Cn$, $|Y| = n$ and degrees of all vertices from Y do not exceed $5Cd^2k\frac{1}{\epsilon}$.

Proof:
Let us fix $C \geq 1$ and $\epsilon > 0$, we consider $d \geq 3$ and k such that by Theorem 4.11 there exists a family of $(\frac{n}{kd}, d, (1-\frac{\epsilon}{4})d)$-expanders $G(X, Y, E)$ with $|X| = 2Cn$, $|Y| = n$. Let us denote $K = 5Cd^2k\frac{1}{\epsilon}$; we modify this graph in such a way that a resulting graph will be an expander with degrees at most K.

We denote $Y' = \{v \in Y \mid deg(v) \geq K\}$ and $X' = \{v \in X \mid |\Gamma(v) \cap Y'| \geq \frac{\epsilon}{2}d\}$. We will prove that the induced subgraph $G'(X \setminus X', Y \setminus Y', E')$ is $(\frac{n}{kd}, d, (1-\epsilon)d)$-expander. Let $\Gamma_H(Z)$ denote the set of neighbours of the set of vertices Z in graph H. Consider some set $Z \subseteq X \setminus X'$ such that $|Z| \leq \frac{n}{kd}$. We know that $(1-\frac{\epsilon}{4})d|Z| \leq |\Gamma_G(Z)|$ and also $|\Gamma_G(Z)| = |\Gamma_{G'}(Z)| + |\Gamma_G(Z) \cap Y'|$. By the definition of X' we get that $|\Gamma_G(Z) \cap Y'| < \frac{\epsilon}{2}d|Z|$. Therefore $(1-\frac{\epsilon}{4})d|Z| \leq |\Gamma_{G'}(Z)| + \frac{\epsilon}{2}d|Z|$, and we get $|\Gamma_{G'}(Z)| \geq (1-\frac{3}{4}\epsilon)d|Z| > (1-\epsilon)d|Z|$.

Let us estimate the sizes of X' and Y'. Since G is bipartite, $\sum\limits_{v \in X} deg(v) = \sum\limits_{v \in Y} deg(v) \leq Cnd$, hence $|Y'| \leq \frac{Cnd}{K} = \frac{\epsilon n}{5kd}$.

Assume that $|X'| > \frac{n}{kd}$ and consider some subset $X_0 \subseteq X'$ such that $|X_0| = \lfloor \frac{n}{kd} \rfloor$. $|\Gamma_G(X_0)| \leq |\Gamma_G(X_0) \setminus Y'| + |Y'| \leq (1-\frac{\epsilon}{2})d|X_0| + |Y'|$. By the property of G we know that $|\Gamma_G(X_0)| \geq (1-\frac{\epsilon}{4})d|X_0|$, hence $\frac{\epsilon}{4}|X_0| \leq |Y'|$ and $|Y'| \geq \epsilon\lfloor \frac{n}{4kd} \rfloor$; the latter contradicts our bound on Y' for n large enough.

Finally, we add to G' several vertices without edges to part $Y \setminus Y'$ in order to make its size precisely n, and delete several vertices from part $X \setminus X'$ to make its size Cn. Note that this operation does not affect the expander property of the graph. □

4.3. Proof of Theorem 4.1

Proof:
[Proof of Theorem 4.1] We consider $\epsilon = \frac{1}{10}$ and constants k and $d \geq 3$ that exist by Theorem 4.12 for given C and $\epsilon = \frac{1}{10}$. By Theorem 4.12 it is possible to construct in polynomial in n time a bipartite graph H_1 such that H_1 is an $(\frac{n}{kd}, d, \frac{9}{10}d)$-expander with $|X| = Cn$, $|Y| = n$, and degrees of all vertices from Y do not exceed $D = 50Cd^2k$. We delete from the part X arbitrary $Cn - m$ vertices and denote the resulting graph by H_2. We add a matching to the graph H_2 in such a way that the resulting graph G will have a matching that covers Y; this procedure increases degrees in at most one. By Lemma 4.10, graph H_2 is an $(\frac{n}{kd}, \frac{8}{10}d)$-boundary expander, and hence G is an $(\frac{n}{kd}, \frac{8}{10}d - 1)$-boundary expander with degrees at most $D + 1$. The formula PMP_G is unsatisfiable since $m > n$. By Theorem 4.5 the width of any resolution refutation of PMP_G is at least $\frac{2n}{5k}$. By Theorem 2.2 the size of any resolution refutation of PMP_G is at least $2^{\Omega\left(((8d/10-1)n/2kd-D-1)^2/n\right)}$. □

5. Existence of subgraphs with a given degree sequence

Let $G(V, E)$ be an undirected graph and h be a function $V \to \mathbb{N}$ such that for every vertex $v \in V$, $h(v)$ is at most the degree of v. We consider a formula $\Psi_G^{(h)}$ constructed as follows: its variables correspond to edges of G. $\Psi_G^{(h)}$ is a conjunction of the following statements: for every $v \in V$, exactly $h(v)$ edges that are incident to v have value 1. The formula PMP_G is a particular case of $\Psi_G^{(h)}$ for $h \equiv 1$.

Theorem 5.1. For all $d \in \mathbb{N}$ there exists $D \in \mathbb{N}$ such that for all n large enough and for any function $h : V \to \{1, 2, \ldots, d\}$ where V is a set of cardinality n, there exists an explicit graph $G(V, E)$ with maximum degree at most D, such that the formula $\Psi_G^{(h)}$ is unsatisfiable, and the size of any resolution refutation for $\Psi_G^{(h)}$ is $2^{\Omega(n)}$.

To prove the Theorem 5.1 we need the following Lemma:

Lemma 5.2. For all $d \in \mathbb{N}$, for all n large enough, for any set V of cardinality n and for any function $h : V \to \{1, 2, \ldots, d\}$ there exists an explicit construction of a graph $G(V, E)$ with the following properties:

1. V consists of two disjoint sets U and T such that there are no edges between vertices from U.
2. The degree of every vertex $u \in U$ equals $h(u) - 1$ and the degree of every vertex $v \in T$ equals $h(v)$.
3. $|U| \geq \frac{n}{2} - 2d^2$.

Proof:
Let $n \geq 4d^2$ and let the vertices $v_1, v_2, \ldots, v_n$ be arranged in a non-decreasing order of $h(v_i)$. Let k be the largest number that satisfies the inequality $\sum_{i=1}^{k}(h(v_i) - 1) < \sum_{i=k+1}^{n} h(v_i) - d(d-1)$. We denote $U = \{v_1, v_2, \ldots, v_k\}$ and $T = V \setminus U$. Obviously, $|U| = k \geq n/2 - d(d-1)$. Now we construct a graph G based on the set of vertices V. We start with an empty graph and add edges one by one. For every vertex $v \in T$ by the co-degree of v we call the difference between $h(v)$ and the current degree of v. From every $u \in U$ we add $h(u) - 1$ edges to G that lead to distinct vertices of $V \setminus U$. Doing so, we maintain degrees of all $v \in T$ below the value of $h(v)$. This always can be done since by the construction of U the total co-degree of all vertices from T is greater than $d(d-1)$, hence for all big enough n there exists at least d vertices with co-degrees at least 1.

While the number of vertices in T with positive co-degrees is greater than d, we will choose one of those vertices $w \in T$ and add to the graph exactly co-degree of w edges that connect w with other vertices from T. Finally, we will have that T contains at most d vertices with co-degrees at most d. Now we connect them with distinct vertices from the set U, remove that vertices from U, and add them to T. It is possible that in the last step some vertex $v \in T$ is already connected with several vertices from U, in that case we should connect v with new vertices. By this operation we deleted at most d^2 vertices from U, and therefore $|U| \geq n/2 - 2d^2$. □

Proof:
[Proof of Theorem 5.1] By Lemma 5.2 we construct a graph $G_1(V, E_1)$ and a set $U \subseteq V$ of size at least $\frac{n}{2} - 2d^2$ such that for all $v \in U$, the degree of v is equal to $h(v) - 1$ and for all $v \in V \setminus U$ the degree of v is equal to $h(v)$. Consider graph $G(U, E_2)$ from Theorem 4.1 with U as the set of its vertices. Define a new graph $G(V, E)$, where the set of edges E equals $E_1 \cup E_2$. Recall that edges from the set E_2 connect vertices of the set U and edges from E_1 do not connect pairs of vertices from U (that follows from the construction of the graph in Lemma 5.2).

For every vertex $v \in V \setminus U$ its degree equals $h(v)$. Therefore, if $\Psi_G^{(h)}$ is satisfiable then in any satisfying assignment of $\Psi_G^{(h)}$ all edges that are incident to vertices $V \setminus U$ must have the value 1. After substituting the value 1 for all these variables, $\Psi_G^{(h)}$ becomes equal to the formula PMP_{G_2} that is unsatisfiable because of Theorem 4.1.

Formula PMP_{G_2} is obtained from $\Psi_G^{(h)}$ by a substitution of several variables, thus Lemma 2.3 implies that the size of any resolution refutation of $\Psi_G^{(h)}$ is at least the size of the minimal refutation for PMP_G, that is at least $2^{\Omega(n)}$ by Theorem 4.1. □

5.1. Corollaries

Tseitin formulas. A Tseitin formula $T_G^{(f)}$ can be constructed from an arbitrary graph $G(V, E)$ and a function $f : V \to \{0, 1\}$; variables of $T_G^{(f)}$ correspond to edges of G. The formula $T_G^{(f)}$ is a conjunction of the following conditions: for every vertex v we write down a CNF condition that encodes that the parity of the number of edges incident to v that have value 1 is the same as the parity of $f(v)$.

Based on the function $f : V \to \{0, 1\}$ we define a function $h : V \to \{1, 2\}$ in the following way: $h(v) = 2 - f(v)$. In other words, if $f(v) = 1$, then $h(v) = 1$, and if $f(v) = 0$, then $h(v) = 2$. By Theorem 5.1 there exists such a number D, that for all n large enough it is possible to construct a graph G with n vertices of degree at most D such that the size of any resolution refutation of the formula Ψ_G^h is at least $2^{\Omega(n)}$.

Note that every condition corresponding to a vertex of the formula $T_G^{(h)}$ is implied from the condition corresponding to the formula Ψ_G^h. Since the resolution proof system is implication complete, every condition of $T_G^{(h)}$ may be derived from a condition of Ψ_G^h by derivation of size at most 2^D. Hence all clauses of the Tseitin formula may be obtained from clauses of formula Ψ_G^h by the derivation of size $O(n)$. Thus the size of any resolution refutation of $T_G^{(f)}$ is at least $2^{\Omega(n)}$. This lower bound was proved in the paper [15].

Complete graph. Let K_n be a complete graph with n vertices and $h : V \to \{1, \ldots, d\}$, where d is some constant. Let formula $\Psi_{K_n}^{(h)}$ be unsatisfiable. By Theorem 5.1 there exists D such that for all n large enough there exists an explicit graph G with n vertices of degree at most D that the size of any resolution refutation of Ψ_G^h is at least $2^{\Omega(n)}$. The graph G can be obtained from K_n by removing several edges, hence the formula $\Psi_G^{(h)}$ can be obtained from $\Psi_{K_n}^{(h)}$ by substituting zeroes for edges that do not present in G. Therefore, by Lemma 2.3 the size of the resolution refutation of $\Psi_{K_n}^{(h)}$ is at least $2^{\Omega(n)}$.

Acknowledgements

The authors are grateful to Alexander Shen for the suggestions on the presentation of results, and to anonymous reviewers for useful comments.

References

[1] Alekhnovich, M.: Mutilated Chessboard Problem is Exponentially Hard for Resolution, *Theor. Comput. Sci.*, **310**(1-3), January 2004, 513–525, ISSN 0304-3975.

[2] Alekhnovich, M., Hirsch, E. A., Itsykson, D.: Exponential Lower Bounds for the Running Time of DPLL Algorithms on Satisfiable Formulas, *J. Autom. Reason.*, **35**(1-3), 2005, 51–72, ISSN 0168-7433.

[3] Ben-Sasson, E., Wigderson, A.: Short proofs are narrow — resolution made simple, *Journal of ACM*, **48**(2), 2001, 149–169.

[4] Buss, S., Pitassi, T.: Resolution and the weak pigeonhole principle, *Proceedings of the CSL97, Lecture Notes in Computer Science*, 1414, 1997.

[5] Dantchev, S. S., Riis, S.: "Planar" Tautologies Hard for Resolution, *FOCS*, 2001.

[6] Haken, A.: The intractability of resolution, *Theoretical Computer Science*, **39**, 1985, 297–308, ISSN 0168-7433.

[7] Hoory, S., Linial, N., Wigderson, A.: Expander Graphs and Their Applications, *Bulletin of the American Mathematical Society*, **43**, 2006, 439–561.

[8] Itsykson, D., Slabodkin, M., Sokolov, D.: Resolution complexity of perfect matching principles for sparse graphs, *Proceedings of CSR-2015*, 9139, 2015.

[9] Itsykson, D., Sokolov, D.: Lower bounds for myopic DPLL algorithms with a cut heuristic, *Proceedings of the 22nd international conference on Algorithms and Computation*, ISAAC'11, Springer-Verlag, available as ECCC Report TR12-141, Berlin, Heidelberg, 2011, ISBN 978-3-642-25590-8.

[10] M. Capalbo, O. Reingold, S. V., Wigderson, A.: Randomness conductors and constant-degree expansion beyond the degree/2 barrier, *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, 2002.

[11] Raz, R.: *Resolution Lower Bounds for the Weak Pigeonhole Principle*, Technical Report 01-021, Electronic Colloquium on Computational Complexity, 2001.

[12] Razborov, A. A.: *Resolution Lower Bounds for the Weak Pigeonhole Principle*, Technical Report 01-055, Electronic Colloquium on Computational Complexity, 2001.

[13] Razborov, A. A.: Resolution lower bounds for the weak functional pigeonhole principle, *Theoretical Computer Science*, **303**(1), 2003, 233—243.

[14] Razborov, A. A.: Resolution lower bounds for perfect matching principles, *Journal of Computer and System Sciences*, **69**(1), 2004, 3–27.

[15] Urquhart, A.: Hard Examples for Resolution, *JACM*, **34**(1), 1987, 209–219.

[16] Urquhart, A.: Resolution Proofs of Matching Principles, *Annals of Mathematics and Artificial Intelligence*, **37**(3), March 2003, 241–250, ISSN 1012-2443.