Satisfiable Tseitin Formulas Are Hard for Nondeterministic Read-Once Branching Programs*

Ludmila Glinskih[1] and Dmitry Itsykson[2]

1 St. Petersburg Department of V.A. Steklov Institute of Mathematics of the Russian Academy of Sciences and St. Petersburg Academic University, Russia
lglinskih@gmail.com

2 St. Petersburg Department of V.A. Steklov Institute of Mathematics of the Russian Academy of Sciences, Russia
dmitrits@pdmi.ras.ru

Abstract

We consider satisfiable Tseitin formulas $\mathrm{TS}_{G,c}$ based on d-regular expanders G with the absolute value of the second largest eigenvalue less than $\frac{d}{3}$. We prove that any nondeterministic read-once branching program (1-NBP) representing $\mathrm{TS}_{G,c}$ has size $2^{\Omega(n)}$, where n is the number of vertices in G. It extends the recent result by Itsykson at el. [9] from OBDD to 1-NBP.

On the other hand it is easy to see that $\mathrm{TS}_{G,c}$ can be represented as a read-2 branching program (2-BP) of size $O(n)$, as the negation of a nondeterministic read-once branching program (1-coNBP) of size $O(n)$ and as a CNF formula of size $O(n)$. Thus $\mathrm{TS}_{G,c}$ gives the best possible separations (up to a constant in the exponent) between 1-NBP and 2-BP, 1-NBP and 1-coNBP and between 1-NBP and CNF.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes - Relations among complexity measures, F.2.2 Nonnumerical Algorithms and Problems - Complexity of proof procedures

Keywords and phrases Tseitin formula, read-once branching program, expander

Digital Object Identifier 10.4230/LIPIcs.MFCS.2017.26

1 Introduction

1.1 Satisfiable and unsatisfiable Tseitin formulas

A Tseitin formula $\mathrm{TS}_{G,c}$ is defined for every undirected graph $G(V,E)$ and labelling function $c : V \to \{0,1\}$. We introduce a propositional variable for every edge of G. The Tseitin formula $\mathrm{TS}_{G,c}$ represents a linear system over the field GF(2) that for every vertex $v \in V$ states that the sum of all edges adjacent to v equals $c(v)$.

A Tseitin formula is satisfiable if and only if the sum of values of the labeling function for all vertices in every connected component is even [17]. The study of Tseitin formulas is motivated by Proof Complexity. Proof Complexity basically deal with unsatisfiable Tseitin formulas that roughly speaking encode that it is impossible that a graph has an odd number of vertices with odd degree. It is important for Proof Complexity that propositional formulas have small CNF representations; thus it is usually assumed that G has constant degree;

* The research was supported by Russian Science Foundation (Project 16-11-10123).

42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017).
Editors: Kim G. Larsen, Hans L. Bodlaender, and Jean-Francois Raskin; Article No. 26; pp. 26:1–26:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

for such graphs Tseitin formulas have CNF representations of size $O(n)$ and it contains $O(n)$ variables where n is the number of vertices in G. Tseitin formulas were invented by Tseitin in 1968 for the graph of $n \times n$ cellular square and were used for the proving of the first superpolynomial lower bound for regular resolution. In 1987 Urquhart extended this result and proved exponential lower bound on the complexity of resolution refutations of Tseitin formulas based on expanders. Unsatisfiable Tseitin formulas are one of the basic examples of hard formulas for many proof systems; in particular, Tseitin formulas are hard for bounded depth Frege [3], [15], Polynomial Calculus over the field $\mathbb{F}$ with $\mathrm{char}(\mathbb{F}) \neq 2$, tree-like Lovasz-Schrijver proof system [8], etc.

Satisfiable Tseitin formulas have been studied less intensively. In the recent paper by Itsykson at el. [9] satisfiable Tseitin formulas appeared in the proof of an exponential lower bound on the size of the derivation of unsatisfiable Tseitin formulas in the proof system $\mathrm{OBDD}(\mathrm{join}, \mathrm{reordering})$. An OBDD is a partial case of a read-once deterministic branching program, where in every path from the source to a sink all variables appear in the same order. The key step in the proof of the mentioned lower bound is the proof of an exponential lower bound on the OBDD representation of satisfiable Tseitin formulas based on constant degree expanders. The latter lower bound motivated us for the current research. There are known examples of Boolean functions that are easy for read-once branching program but hard for OBDD (see for example Theorem 6.1.2 in [18]). Is it possible to extend the mentioned lower bound from OBDD to read-once branching program?

It is well known that the size of the shortest regular resolution proof of any unsatisfiable CNF formula ϕ equals the size of the minimal read-once branching program for the following search problem $Search_\phi$: given an assignment of variables of ϕ, find a clause that is refuted by this assignment [12]. Thus lower bounds for the size of resolution proofs of ϕ implies lower bounds on the size of read-once branching program for $Search_\phi$. However it is unclear whether the sizes of read-once branching programs for $Search_\phi$ for unsatisfiable Tseitin formula ϕ and for the evaluation of a satisfiable Tseitin formula are connected. The difference is the following:

1. the first case is about unsatisfiable Tseitin formulas while the second case is about satisfiable Tseitin formulas;
2. to find a clause that is refuted may be harder than just to say that the value of a function is 0.

1.2 Results

In this paper we prove that every nondeterministic read-once branching program (1-NBP) representing a satisfiable Tseitin formula $\mathrm{TS}_{G,c}$ based on d-regular expander with the absolute value of the second largest eigenvalue less than $\frac{d}{3}$ has size $2^{\Omega(n)}$, where n is the number of vertices in G and d is a constant. As a corollary we get a lower bound $2^{\Omega(n)}$ on the size of nondeterministic read-once branching programs for Tseitin formulas based on complete graph K_n. All mentioned lower bounds are tight up to a constant in the exponent since every satisfiable Tseitin formula based on graph with n vertices and m edges may be represented as OBDD of size $O(m2^n)$ (see Proposition 3 below).

1.3 Comparison with other works

If we consider a Tseitin formula as a system of linear equations then every variable will have exactly two occurrences. Therefore by straightforward transformation every satisfiable Tseitin formula $\mathrm{TS}_{G,c}$ may be represented as read-2 deterministic branching program

(2-BP) of size $O(m)$, where m is the number of edges in G. Thus satisfiable Tseitin formulas based on constant-degree expanders strongly exponentially separate 1-NBP and 2-BP. And this separation is optimal up to a constant in the exponent. Consider a function $\mathrm{CLIQUE\_ONLY}_n : \{0,1\}^{n(n-1)/2} \to \{0,1\}$ that detect whether a undirected graph on n vertices is exactly a clique on $\lfloor n/2 \rfloor$ vertices. Borodin, Razborov and Smolensky [4] proved that any nondeterministic read-once branching program representing $\mathrm{CLIQUE\_ONLY}_n$ has size $2^{\Omega(n)}$ (note that $\mathrm{CLIQUE\_ONLY}_n$ depends on $\Theta(n^2)$ variables) while there is a deterministic read-twice branching program of size poly(n). Thathachar [16] gave, for every natural k, an explicit function that can be evaluated by deterministic read-$(k+1)$ branching program of linear size but every nondeterministic read-k branching program for this function has size at least $2^{\Omega\left(n^{1/(k+1)}\right)}$. As far as we know, the best previously known gap between sizes of 1-NBP and 2-BP was $2^{\Omega(\sqrt{n})}$ and we improved it to $2^{\Omega(n)}$.

First explicit Boolean function with strongly exponential lower bound on the size of 1-NBP was constructed in [4], however this function was rather artificial. Duros at el. [7] proved strongly exponential lower bounds on the size of 1-NBP for the function $\oplus cl_{3,n}$ that computes the parity of the number of triangles in the graph (and this extends the result of Babai at el.[2] from 1-BP to 1-NBP) and for the function $\Delta_{3,n}$ that is true iff the input graph does not contain triangles. So satisfiable Tseitin formulas based on constant-degree expanders is one more natural example of functions that require strongly exponential 1-NBP.

A satisfiable Tseitin formula based on a d-regular graph on n vertices is a characteristic function of an affine subspace of $\{0,1\}^{dn/2}$. Characteristic functions of affine (linear) subspaces were already studied in the context of complexity of deterministic and nondeterministic reak-k branching programs, namely characteristic functions of linear error-correcting codes were studied by Okolnishnikova [14] and Jukna [10]. Jukna [10] proved lower bound $2^{\Omega(\sqrt{n})}$ on the size of nondeterministic read-k branching program for characteristic functions of error-correcting codes $\mathcal{C} \subseteq \{0,1\}^n$. Duris at el. [7] presented a probabilistic construction of a linear code such that its characteristic function require 1-NBP of the size at least $2^{\Omega(n)}$. Jukna [10] noted that the negation of a characteristic function of an affine subspace may be represented by linear size (in the size of linear system that defines this subspace) nondeterministic read-once branching program. Indeed we just need to guess an equation that is not satisfied and then check this equation. Duris at el. [7] also showed that the characteristic function of a linear subspace of $\{0,1\}^n$ (and hence it is also true for affine subspaces) may be represented by a randomized read-once branching program with one-sided error 2^{-r} of size $O(n^r)$ for all natural r. As far as we know, satisfiable Tseitin formulas based on explicit constant-degree expanders are the only example of *explicit* functions (randomized construction was presented in [7]) that strongly exponentially separate nondeterministic and co-nondeterministic read-once branching program. And also our separation between nondeterministic and randomized read-once branching program seems to be the best known for the explicit functions.

We finally note that Tseitin formulas based on constant degree graphs may be represented as CNF formulas of size $O(n)$, so we get a strongly exponential separation between sizes of 1-NBP and CNF.

2 Preliminaries

2.1 Branching programs

A deterministic branching program (BP) is a form of representation of Boolean functions. A Boolean function $\{0,1\}^n \to \{0,1\}$ is represented by a directed acyclic graph with exactly one source and two sinks. All nodes except sinks are labeled with a variable; every internal node has exactly two outgoing edges: one is labeled with 1 and the other is labeled with 0. One of the sinks is labeled with 1 and the other is labeled with 0. The value of the function for a given values of variables is evaluated as follows: we start a path from the source such that for every node on its path we go along the edge that is labeled with the value of the corresponding variable. This path will end in a sink. The label of this sink is the value of the function.

A nondeterministic branching program (NBP) differs from a deterministic in the way that we also allow guessing nodes that are unlabeled and have two outgoing unlabeled edges. So nondeterministic branching program may have three type of nodes: guessing nodes, nodes labeled with a variable (we call them just labeled nodes) and two sinks; the source may be either a guessing node or labeled node. The result of a function represented by a nondeterministic branching program equals 1, if there exists at least one path from the source to the sink labeled with 1 such that for every node labeled with a variable on its path we go along an edge that is labeled with the value of the corresponding variable, while for guessing nodes we are allowed to choose any of two outgoing edges.

Note that deterministic branching programs constitute a special case of nondeterministic branching programs.

A deterministic or nondeterministic branching program is (syntactic) read-k (k-BP or k-NBP) if every path from the source to a sink contains at most k occurrences of every variable.

An ordered binary decision diagram (OBDD) is a partial case of 1-BP, where on every path from the source to a sink all variables appear in the same order.

2.2 Tseitin formulas

Let $G(V,E)$ be an undirected graph without loops but possibly with multiple edges, $c : V \to \{0,1\}$ be a labeling function that matches every vertex with a boolean value. Let us match every edge $e \in E$ with a propositional variable x_e. Tseitin formula $\mathrm{TS}_{G,c}$ based on a graph G and a labeling function c is the conjunction of the following conditions: for every vertex v the sum of variables x_e for all edges e that are incident to v equals $c(v)$ modulo 2. More formally: $\bigwedge\limits_{v \in V} \left(\sum\limits_{e \text{ is incident to } v} x_e = c(v) \bmod 2 \right)$.

If the maximal degree of a graph G is bounded by a constant d, then a sum modulo 2 can be written as a d-CNF formula with size at most $O(2^d d)$. Hence the size of CNF representation of $\mathrm{TS}_{G,c}$ does not exceed $O(2^d dn)$.

We will use the following criterion of the satisfiability of Tseitin formulas:

▶ **Proposition 1** ([17]). A Tseitin formula $\mathrm{TS}_{G,c}$ is satisfiable if and only if for every connected component U the following holds: $\sum\limits_{v \in U} c(v) = 0 \bmod 2$.

▶ Remark. Note that a substitution of a value to a variable $x_e := \alpha$ transforms Tseitin formula $\mathrm{TS}_{G,c}$ to a Tseitin formula $\mathrm{TS}_{G',c'}$, where graph G' is obtained from the graph G by deleting the edge e, c' equals c in every vertex except two vertices that are incident to edge e. On these two vertices the values of c and c' differ by α. In particular it follows that if $\mathrm{TS}_{G,c}$ is satisfiable and an edge e is not a bridge in the graph G, then the formula $\mathrm{TS}_{G',c'}$ is also satisfiable by Proposition 1 since the parity of sum of labels in G' in every connected component is the same as in G.

▶ **Lemma 2.** *Let $G(V,E)$ be a graph with k connected components. If the Tseitin formula $\mathrm{TS}_{G,c}$ is satisfiable, then the number of its satisfying assignments equals $2^{|E|-|V|+k}$.*

Proof. Let us fix some spanning forest F of the graph G; F contains exactly $|V| - k$ edges. Consider some partial substitution ρ to the edges of G that are not in F. By the Remark we know that after the application of the partial substitution ρ to $\mathrm{TS}_{G,c}$ we will get a satisfiable Tseitin formula based on the graph F. Since F is a forest the resulting Tseitin formula has exactly one satisfying assignment. Indeed a forest always has a vertex with degree 1 which helps us unambiguously determine the value of the incident edge. After that we can delete this edge from the forest and we will get a forest again; and so on. Hence the number of satisfying assignment of $\mathrm{TS}_{G,c}$ equals the number of different partial substitutions to the edges that are not in F; so the number of satisfying assignment equals $2^{|E|-|V|+k}$. ◀

▶ **Proposition 3.** Any satisfiable Tseitin formula based on a graph with n vertices and m edges can be represented as OBDD of size $O(m2^n)$.

Proof. Let us fix some order on the edges of the graph. The described OBDD will have m levels. Nodes on the i-th level are labeled with i-th edge of the graph.

Assume that we already ask for the value of the first $i - 1$ edges. For every vertex of the graph we compute the sum modulo 2 of values of edges from these $i - 1$ that are incident to the vertex. So we will have a vector of n parities. The i-th level of the OBDD contains 2^n nodes corresponding to the all possible values of vector of parities that we get after the reading of the first $(i - 1)$ edges. Every node on the i-th level has two outgoing edges to nodes on the $(i + 1)$-th level corresponding to the way how values on the edges change the parity of vertices. The node on the first level corresponding to all zero values of parities is the source of the OBDD (all nodes that are not reachable from the source should be removed). Outgoing edges for every node on the last level will go to a sink corresponding to the fact, whether the labeling function of the Tseitin formula is consistent with the resulting values of parities. ◀

▶ **Proposition 4.** 1) Every two satisfying assignments of a satisfiable Tseitin formula $\mathrm{TS}_{G,c}$ differ in at least two positions. 2) Every path from the source to the sink labeled with 1 in 1-NBP representing a satisfiable Tseitin formula $\mathrm{TS}_{G,c}$ contains variables for all edges of G.

Proof. 1) If we change a value of any edge in a satisfying assignment of $\mathrm{TS}_{G,c}$, the parity condition will be violated on two ends of this edge. 2) Assume that some acceptance path does not contains a variable x. Then there are two satisfying assignments of $\mathrm{TS}_{G,c}$ that differ only in the value of the variable x; this contradicts item 1. ◀

2.3 Expanders

Let $G(V, E)$ be an undirected graph without loops but possibly with multiple edges. G is an algebraic (n, d, α)-expander if G is d-regular, $|V| = n$ and the absolute value of the second largest eigenvalue of the adjacency matrix of G is not greater than αd.

It is well known that for all $1 > \alpha > 0$ and all large enough constants d there exist natural number n_0 and a family $\{G_n\}_{n=n_0}^{\infty}$ of (n, d, α)-algebraic expanders. There are explicit constructions such that G_n can be constructed in $\mathrm{poly}(n)$ time [13]. Also, it is known that a random d-regular graph is an expander with high probability.

Let us denote by $E(A, B)$ a multiset of edges that have one end in A and another end in B. Note that in the case where both ends of an edge are simultaneously in A and in B, we count this edge twice.

▶ **Lemma 5** (Cheeger inequality [5])**.** *Let $G(V, E)$ be an (n, d, α)-expander. Then for all $A \subseteq V$ such that $|A| \le \frac{n}{2}$ the following inequality holds: $|E(A, V \setminus A)| \ge \frac{1-\alpha}{2} d|A|$.*

▶ **Corollary 6.** *Every (n, d, α)-expander with $0 < \alpha < 1$ is connected.*

Proof. If G is not connected, then we will get a contradiction with Lemma 5 if we choose A to be a smallest connected component. ◀

▶ **Lemma 7** (Expander mixing lemma [1]). *Let $G(V, E)$ be (n, d, α)-expander, $A, B \subseteq V$. Then $\left||E(A, B)| - \frac{d|A||B|}{n}\right| \leq \alpha d\sqrt{|A||B|}$.*

Using Lemma 7 we can improve the estimation of the number of edges that go from A to the complement of A for small sets A.

▶ **Proposition 8.** For every (n, d, α)-expander for every $A \subseteq V$ the following inequality holds: $|E(A, V \setminus A)| \geq d|A|(1 - \frac{|A|}{n} - \alpha)$.

Proof. $|E(A, V \setminus A)| = |E(A, V)| - |E(A, A)| = d \cdot |A| - |E(A, A)| \geq d \cdot |A|(1 - \frac{|A|}{n} - \alpha)$. The last inequality follows from Lemma 7. ◀

3 Lower bound

Our main goal is to prove the following theorem:

▶ **Theorem 9.** *Let $G(V, E)$ be an algebraic (n, d, α)-expander, where $\alpha < \frac{1}{3}$. Let $\mathrm{TS}_{G,c}$ be a satisfiable Tseitin formula. Then the size of every* 1-NBP *that represents $\mathrm{TS}_{G,c}$ is $2^{\Omega(n)}$.*

Let us describe the plan of the proof. Consider a minimal 1-NBP that evaluates $\mathrm{TS}_{G,c}$. For every node of this branching program, except the sink labeled with 0 there exists a path to the sink labeled with 1. In the opposite case this node could be merged with a sink labeled with 0 and it would decrease the size of the 1-NBP.

For nondeterministic branching program, by *the length of a path* we will mean the number of labeled edges in it (i.e. we do not count outgoing edges from guessing nodes). For every labeled node v in a branching program we define its level as the minimal length of paths from the source to v. We choose a level $l = \Omega(n)$ and prove that the minimal 1-NBP contains many label nodes on the level l. The proof consists of two parts:

1. We show that every minimal 1-NBP that evaluates $\mathrm{TS}_{G,c}$ contains at least $2^{C_1 n}$ paths of length l from the source to a labeled node that correspond to different partial substitutions, where C_1 is a constant.
2. We show that in every minimal 1-NBP that evaluates $\mathrm{TS}_{G,c}$ for every labeled node v on the level l there are at most $2^{C_2 n}$ different partial substitutions that correspond to different paths from the source to the vertex v, where $C_2 < C_1$ is a constant.

These two propositions imply that the number of label nodes on the level l is at least $2^{(C_1 - C_2)n}$.

3.1 Lower bound on the number of paths

In this section we perform the first part of the plan and estimate the number of paths of length l from the source of the minimal 1-NBP that end in a labeled node and correspond to different partial substitutions.

▶ **Lemma 10.** *Let $G(V, E)$ be a connected graph. Let k be the maximum number of connected components that can be obtained after deleting of l edges from G. Then every minimal* 1-NBP *evaluating a satisfiable Tseitin formula $\mathrm{TS}_{G,c}$ contains at least $2^{l-(k-1)}$ paths of length l from the source that end in a labeled node and correspond to different partial substitutions.*

Proof. By Lemma 2 the number of satisfying assignments of the formula $\mathrm{TS}_{G,c}$ equals $2^{|E|-|V|+1}$.

For every satisfying assignment of $\mathrm{TS}_{G,c}$ there exists a path in the minimal 1-NBP from the source to the sink labeled with 1 of length $|E|$ that is consistent with the assignment. By Proposition 4 it is impossible that there are paths from the source to the sink labeled with 1 that are shorter than $|E|$. Let P be the set of paths from the source to the sink labeled with 1 such that for every satisfying assignment of $\mathrm{TS}_{G,c}$ there are exactly one path in P that represents this assignment.

We estimate the number of paths in P that define the same partial substitution ρ that corresponds to the first l labeled edges of the path.

If we apply ρ to $\mathrm{TS}_{G,c}$ we will get a Tseitin formula $\mathrm{TS}_{G',c'}$, where G' is obtained from G by deleting l edges corresponding to the path p (the labeling function also changes after the application of ρ, see Remark 2.2 for details). All paths from P that are consistent with ρ satisfy the formula $\mathrm{TS}_{G',c'}$. Recall that all paths from P correspond to different satisfying assignments, hence the number of paths that are consistent with ρ is not greater than the number of satisfying assignments of the formula $\mathrm{TS}_{G',c'}$. By Lemma 2 the number of satisfying assignments of the formula $\mathrm{TS}_{G',c'}$ equals $2^{|E|-l-|V|+m}$, where m is the number of connected components in the graph G'. By the statement of the lemma $m \le k$, therefore the number of satisfying assignments of $\mathrm{TS}_{G',c'}$ is not greater than $2^{|E|-l-|V|+k}$. So we get that every partial substitution of l variables may be a prefix of length l (we assume that prefixes end in labeled nodes) of at most $2^{|E|-l-|V|+k}$ paths from P. Hence there are at least $\frac{2^{|E|-|V|+1}}{2^{|E|-l-|V|+k}} = 2^{l-(k-1)}$ different partial substitutions that correspond to prefixes of length l of paths from P, and these prefixes we will consider as the paths which number we estimate in the lemma. ◀

▶ **Lemma 11.** *Every graph that can be obtained by deleting $l \le \frac{n}{4}$ edges from an algebraic (n, d, α)-expander G contains at most $\frac{2l}{d(1-\alpha)} + 1$ connected components.*

Proof.

▶ **Claim 12.** Let graph $H(V, E)$ have n vertices and k connected components, where $1 < k \le \frac{n}{4} + 1$. Then there exists $M \subseteq V$ such that M consists of the union of all vertices of several connected components and $k - 1 \le |M| \le \frac{n}{2}$.

Proof of Claim 12. We construct M iteratively. Assume that initially M is empty. Let us sort all connected components in the increasing order of their sizes: $s_1, s_2, \dots, s_k$. We add connected components to M starting from the smallest one while the sum of the sizes of these components is less than $k - 1$. Let i be the number of connected components that we added to M. If $|M| \le \frac{n}{2}$ then we are already done. Assume that $|M| > \frac{n}{2}$. Note that $|M \setminus s_i| < k - 1$ by the construction of M. Hence $|s_i| > \frac{n}{2} - (k - 1) \ge k - 1$ since $k - 1 \le \frac{n}{4}$. If $|s_i| \le \frac{n}{2}$ then the we can take $M = s_i$. So we may assume that $|s_i| > n/2$, therefore s_i is the biggest connected component and $i = k$. Since every connected component contains at least one vertex the number of vertices in $M \setminus s_i$ should be at least $k - 1$ that contradicts the construction of M. ◀

Consider some subgraph H that may be obtained from G by deleting of at most l edges. By Corollary 6 G is connected, hence H contains at most $\frac{n}{4} + 1$ connected components. Consider the set M from Claim 12. Let us estimate the number of edges that we need to delete from G in order to separate M from other vertices of the graph. By Lemma 5 $l \ge \frac{|M| \cdot d \cdot (1-\alpha)}{2} \ge \frac{(k-1) \cdot d \cdot (1-\alpha)}{2}$. Hence $k - 1 \le \frac{2l}{d \cdot (1-\alpha)}$. ◀

Altogether Lemma 10 and Lemma 11 imply the following lemma:

▶ **Lemma 13.** *In every minimal* 1-NBP *that represents a satisfiable Tseitin formula based on an* (n, d, α)*-expander for every* $l \leq \frac{n}{4}$ *there are at least* $2^{l\left(1-\frac{2}{d(1-\alpha)}\right)}$ *paths of length* l *from the source to a labeled node that correspond to different partial substitutions.*

3.2 Upper bound on the number of paths that end at the same vertex

In this section we estimate the maximum number of paths with length l that ends in a fixed labeled node v and correspond to different partial substitutions. In particular we prove the following lemma:

▶ **Lemma 14.** *For every minimal* 1-NBP *that evaluates a satisfiable Tseitin formula* $\mathrm{TS}_{G,c}$ *based on an* (n, d, α)*-expander* G *for every* $\beta \in (0; 1)$ *for every labeled node* v *of the* 1-NBP *there are at most* $2^{l\left(1-\frac{1}{d(\alpha+\beta)}\right)}$ *different partial substitutions that correspond to paths of length* l *from the source to* v*, where* $l \leq \beta n - 1$*.*

Proof.

▶ **Claim 15.** Consider some labeled node v of the 1-NBP. Let p_1 and p_2 be two different paths from the source to the node v. Then

1. the sets of variables that correspond to labeled nodes on the paths p_1 and p_2 are equal;
2. if we apply to $\mathrm{TS}_{G,c}$ a partial substitution corresponding to p_1, we get the same Tseitin formula as if we apply to $\mathrm{TS}_{G,c}$ a partial substitution corresponding to p_2.

Proof of Claim 15. 1. Since v is a labeled node and the 1-NBP is minimal there is a path s from v to the sink labeled with 1. Both paths p_1s and p_2s go from the source to the sink labeled with 1. Every variable appears in both of these paths at most once. Let x be a variable that appears in p_1 but doesn't appear in p_2 then the substitution corresponding to the path p_2s satisfy $\mathrm{TS}_{G,c}$. By Proposition 4 p_2 should contain the variable x.

2. By Remark 2.2 if we apply a partial substitution to a Tseitin formula we also get a Tseitin formula. The sets of satisfying assignments of two different Tseitin formulas do not intersect, because every satisfying assignment of variables unambiguously determines the labeling function of Tseitin formula. Paths p_1s and p_2s satisfy the initial formula hence the path s should satisfy both Tseitin formulas, the one corresponding to the path p_1 and the one corresponding to the path p_2. Hence these two Tseitin formulas should be equal. ◀

Let v be some labeled node that has level l. By Claim 15 every path from the source to the node v contains the same set of variables and if we apply to $\mathrm{TS}_{G,c}$ any of the substitutions corresponding to these paths we get the same Tseitin formula $\mathrm{TS}_{H,c'}$. Consider some path from the source to v of length l and denote the set of labels (i.e. variables) of the first l labeled nodes on this path by I. By Claim 15 I does not depend on the choice of the path. Let F be a set of edges that correspond to variables from I. Then $I = \{x_e \mid e \in F\}$ and H is obtained from G by deleting of all edges from F.

We define a system of linear equations depending on variables from I. This system states that the substitution to variables from I change labeling function from c to c' as follows:

$$\bigwedge_{u \in V} \left(\sum_{\substack{e \in F: \\ e \text{ is incident to } u}} x_e = c(u) + c'(u) \bmod 2 \right) \tag{1}$$

For every path from the source to v a partial substitution corresponding to this path is a solution of the system (1). The opposite is not always true since that it is not necessary that a path corresponding to the solution of the system (1) exists in the branching program.

▶ **Claim 16.** The number of solutions of the system (1) is equal to 2^{l-t}, where t is the number of the edges in the spanning forest of a graph $H(V, F)$ that is obtained from G by the deletion of all edges that are not in F.

Proof. Notice that the system (1) is precisely the Tseitin formula $\mathrm{TS}_{H,c+c'}$ based on the graph H and labelling function $c + c'$. We know that the system (1) has solutions, hence the number of its solutions by Lemma 2 equals $2^{|F|-|V|+k}$, where k is the number of connected components in H. The claim is proved since $|F| = l$ and $t = |V| - k$. ◀

▶ **Corollary 17.** *The number of different partial substitutions that correspond to paths going from the source to v is at most 2^{l-t}.*

▶ **Claim 18.** Let G be an algebraic (n, d, α)-expander. Assume that we deleted all edges from the graph except l edges, where $l = \beta n - 1$ and $0 < \beta < 1$. Then the number of edges in the spanning forest of the resulting graph H is at most $\frac{l}{d\cdot(\alpha+\beta)}$.

Proof. Consider any connected component $C \subseteq V$ in the resulting graph H. Let m be the number of edges and t be the number of vertices in C. We estimate the maximal number of edges that connect two vertices from C in the original graph G.

Since G is an algebraic (n, d, α)-expander by Proposition 8 there are at least $dt(1 - \frac{t}{n} - \alpha)$ edges connecting vertices from C with vertices from $V \setminus C$ in the graph G. Hence there are at most $\frac{dt - dt + \frac{dt^2}{n} + \alpha\cdot dt}{2} = \frac{\frac{dt^2}{n} + \alpha\cdot dt}{2}$ edges in G that connect two vertices from C.

Let us note that $t \le m + 1 \le l + 1 \le \beta n$, hence $m \le \frac{dt\cdot(\alpha+\beta)}{2}$. The latter implies that

$$t \ge \frac{2m}{d\cdot(\alpha+\beta)}. \tag{2}$$

Let t_i and m_i be the numbers of vertices and edges in the i-th connected component respectively. Note that the size of the spanning forest in H equals $\sum_i (t_i - 1) = \sum_{i:t_i\ge 2} (t_i - 1) \ge \sum_{i:t_i\ge 2} \frac{t_i}{2}$. Note that all edges of H are in the components of size at least two.

By the inequality (2) we get $\sum_{i:t_i\ge 2} t_i \ge \sum_{i:t_i\ge 2} \frac{2m_i}{d\cdot(\alpha+\beta)} = \frac{2l}{d\cdot(\alpha+\beta)}$. Hence the resulting size of the spanning forest is at least $\frac{l}{d\cdot(\alpha+\beta)}$. ◀

Lemma 14 follows from Corollary 17 and Claim 18. ◀

3.3 Proof of Theorem 9

Proof of Theorem 9. Let $\beta = \min\{\frac{1}{4}, \frac{1-3\alpha}{3}\}$ and $l = \beta n - 1$. Consider the minimal 1-NBP for the Tseitin formula $\mathrm{TS}_{G,c}$.

By Lemma 13 there exist at least $2^{l\left(1-\frac{2}{d(1-\alpha)}\right)}$ paths of length l from the source that end in a labeled node that correspond to different partial substitutions. By Lemma 14 for every labeled node v on the level l there are at most $2^{l\left(1-\frac{1}{d(\alpha+\beta)}\right)}$ different partial substitutions that correspond to paths from the source to v.

Hence there are at least $2^{\frac{l}{d}\left(\frac{1}{\alpha+\beta}-\frac{2}{1-\alpha}\right)}$ labeled nodes on the distance l from the source. The latter is $2^{\Omega(n)}$ since $\beta < \frac{1-3\alpha}{2}$. ◀

3.4 Tseitin formula for complete graph

▶ **Corollary 19.** *Let* $\mathrm{TS}_{K_n,c}$ *be a satisfiable Tseitin formula, where* K_n *is a complete graph on* n *vertices. Then the size of every* 1-NBP *for* $\mathrm{TS}_{K_n,c}$ *is* $2^{\Omega(n)}$.

Proof. Consider a 1-NBP D that evaluates the formula $\mathrm{TS}_{K_n,c}$. Consider a graph G on n vertices that is an algebraic (n, d, α)-expander with $\alpha < \frac{1}{3}$ (note that G may have multiple edges). Consider a partial substitution ρ that assigns 0 for every edge that is in K_n but is not in G. Let D' be a 1-NBP that represents the result of the application of ρ to D. It is straightforward that the size of D' is at most the size of D. D' evaluates satisfiable Tseitin formula $\mathrm{TS}_{G',c}$, where G' is a graph that differs from G only by the fact that G may contain multiple edges (G' does not contain multiple edges). I.e., between every two vertices in the graph G' there is an edge if and only if there is at least one edge between these two vertices in G. Now we show how to obtain the diagram for $\mathrm{TS}_{G,c}$ from the diagram D'. Let graph G contain k edges between vertices u and v: $e_1, e_2, \ldots, e_k$. Note that $k \leq d$. It is well known that there exists a read-once deterministic branching program that evaluates $x_{e_1} + x_{e_2} + \cdots + x_{e_k}$ of size $k + 2$. Let us denote this branching program by R. We put the source of R in the nodes labeled with variable $x_{u,v}$; the sink labeled with 0 in R should be identified with the end of the edge that correspond to the decision $x_{u,v} = 0$. And similarly we do with the sink labeled with 1. We do such substitutions for every pair of vertices that has multiple edges. The resulting program will be read-once because the original diagram was read-once. The size of the resulting program is at most d times greater than the size of the original branching program. By Theorem 9 the size of the resulting program is $2^{\Omega(n)}$ hence the size of D is $2^{\Omega(n)}$. ◀

3.5 Lower bound for arbitrary graphs

Let for connected graph $G(V, E)$ the value $k_G(l)$ denote the maximal number of connected components that can be obtained from G by deleting of l edges.

Lemma 10 and Corollary 17 imply:

▶ **Corollary 20.** *For all connected graphs* $G(V, E)$ *and arbitrary* $1 \leq l \leq |E|$ *the size of any* 1-NBP *evaluating a satisfiable Tseitin formula* $\mathrm{TS}_{G,c}$ *is at least* $2^{|V|-k_G(l)-k_G(|E|-l)+1}$.

Proof. Consider the minimal 1-NBP for the Tseitin formula $\mathrm{TS}_{G,c}$.

By Lemma 10 there exist at least $2^{l-k_G(l)+1}$ paths of length l from the source that end in a labeled node that correspond to different partial substitutions. By Corollary 17 for every labeled node v on the level l there are at most $2^{l-|V|+k_G(|E|-l)}$ different partial substitutions that correspond to paths from the source to v.

Hence there are at least $2^{|V|-k_G(l)-k_G(|E|-l)+1}$ labeled nodes on the distance l from the source. ◀

In the proof of Theorem 9 we actually show that for (n, d, α)-expander with $\alpha < \frac{1}{3}$ $k_G(l) - k_G(|E| - l) < (1 - \epsilon)n$ for some l and some constant $\epsilon > 0$. It implies that Theorem 9 also holds for graphs that differ from (n, d, α) expander by at most $\epsilon n/4$ edges since modification of $\epsilon n/4$ edges changes $k_G(l) + k_G(|E| - l)$ by at most $\epsilon n/2$.

It was proved in the paper [9] that for all connected graphs $G(V, E)$ and arbitrary $1 \leq l \leq |E|$ the size of OBDD evaluating a satisfiable Tseitin formula $\mathrm{TS}_{G,c}$ is at least $2^{|V|-k'_G(l)}$, where $k'_G(l)$ is the maximum over all sets $E' \subseteq E$ of size l of the total number of connected components in graphs G' and G'', where G' is a graph with vertices V and edges E', G'' is a graph with vertices V and edges $E \setminus E'$. It is straightforward that

$k_G(l) + k_G(|E| - l) \geq k'_G(l)$. Thus theoretically the lower bound on the size of OBDD from [9] may be slightly stronger then the lower bound from Corollary 20 for some specific graphs.

4 Futher research

Jukna [10] defined the notion of semantic nondeterministic read-k branching programs that have weaker requirement about occurrences of variables. Namely on every consistent path from the source to a sink labeled with 1 every variable should be tested in at most k times. Jukna showed that semantic nondeterministic read-once branching programs are strictly stronger than syntactic ones and formulated an open question to prove superpolynomial lower bound on the size of semantic 1-NBP. Currently such lower bounds are known only for explicit functions from $D^n \to \{0, 1\}$ with non-binary domains D of size at least 3 [6, 11]. Perhaps a satisfiable Tseitin formula is a good candidate for the binary case.

Acknowledgements. The authors are grateful to Mikhail Slabodkin and reviewers for useful comments and to Alexander Knop and Dmitry Sokolov for fruitful discussions.

References

1. Noga Alon and Fan R. K. Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 306(10-11):1068–1071, 2006. doi:10.1016/j.disc.2006.03.025.
2. L. Babai, P. Hajnal, E. Szemeredi, and G. Turan. A lower bound for read-once-only branching programs. *Journal of Computer and System Sciences*, 35:153–162, 1987.
3. Eli Ben-Sasson. Hard examples for bounded depth Frege. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 563–572, 2002. doi:10.1145/509907.509988.
4. Allan Borodin, Alexander A. Razborov, and Roman Smolensky. On lower bounds for read-k-times branching programs. *Computational Complexity*, 3:1–18, 1993. doi:10.1007/BF01200404.
5. J Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems Anal.*, page 195, 1970.
6. Stephen A. Cook, Jeff Edmonds, Venkatesh Medabalimi, and Toniann Pitassi. Lower bounds for nondeterministic semantic read-once branching programs. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 36:1–36:13, 2016. doi:10.4230/LIPIcs.ICALP.2016.36.
7. Pavol Duris, Juraj Hromkovic, Stasys Jukna, Martin Sauerhoff, and Georg Schnitger. On multi-partition communication complexity. *Inf. Comput.*, 194(1):49–75, 2004. doi:10.1016/j.ic.2004.05.002.
8. D.M. Itsykson and A.A. Kojevnikov. Lower bounds of static Lovasz-Schrijver calculus proofs for Tseitin tautologies. *Zapiski Nauchnykh Seminarov POMI*, 340:10–32, 2006.
9. Dmitry Itsykson, Alexander Knop, Andrei Romashchenko, and Dmitry Sokolov. On obdd-based algorithms and proof systems that dynamically change order of variables. In *34th Symposium on Theoretical Aspects of Computer Science, STACS 2017, March 8-11, 2017, Hannover, Germany*, pages 43:1–43:14, 2017. doi:10.4230/LIPIcs.STACS.2017.43.
10. Stasys Jukna. A note on read-k times branching programs. *ITA*, 29(1):75–83, 1995.
11. Stasys Jukna. A nondeterministic space-time tradeoff for linear codes. *Inf. Process. Lett.*, 109(5):286–289, 2009. doi:10.1016/j.ipl.2008.11.001.
12. László Lovász, Moni Naor, Ilan Newman, and Avi Wigderson. Search Problems in the Decision Tree Model. *SIAM J. Discrete Math.*, 8(1):119–132, 1995. doi:10.1137/S0895480192233867.

13 A Lubotzky, R Phillips, and P Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

14 EA Okolnishnikova. On lower bounds for branching programs. *Siberian Advances in Mathematics*, 3(1):152–166, 1993.

15 Toniann Pitassi, Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. Poly-logarithmic Frege depth lower bounds via an expander switching lemma. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 644–657, 2016. doi:10.1145/2897518.2897637.

16 Jayram S. Thathachar. On separating the read-k-times branching program hierarchy. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 653–662, 1998. doi:10.1145/276698.276881.

17 A. Urquhart. Hard examples for resolution. *JACM*, 34(1):209–219, 1987.

18 Ingo Wegener. *Branching Programs and Binary Decision Diagrams*. SIAM, 2000.