

Finite Limits and Monotone Computations: The Lower Bounds Criterion*

Stasys Jukna

Department of Computer Science, University of Trier[†]
D-54286 Trier, Germany
jukna@ti.uni-trier.de

Abstract

Our main result is a combinatorial lower bounds criterion for monotone circuits over the reals. We allow any unbounded fanin non-decreasing real-valued functions as gates. The only requirement is their "locality". Unbounded fanin AND and OR gates, as well as any threshold gate $T_s^m(x_1, \dots, x_m)$ with small enough threshold value $\min\{s, m - s + 1\}$, are simplest examples of local gates. The proof is relatively simple and direct, and combines the bottlenecks counting approach of Haken with the idea of finite limit due to Sipser. Apparently this is the first combinatorial lower bounds criterion for monotone computations. It is symmetric and yields (in a uniform and easy way) exponential lower bounds.

1. Introduction

The question of determining how much economy the universal non-monotone basis $\{\wedge, \vee, \neg\}$ provides over the monotone basis $\{\wedge, \vee\}$ has been a long standing open problem in Boolean circuit complexity. The breakthrough in the field was made by Razborov in his seminal paper [23] where the first super-polynomial lower bound was proved. Shortly after, such (and even exponential) lower bounds were obtained for different Boolean functions [24, 3, 1, 30, 4, 31], including those whose non-monotone circuits are polynomial [24, 30].

After this impressing progress one principal question still remained unclear: is there a tractable lower bounds *criterion* for monotone circuits? Razborov raised this problem as a candidate for a "final chord" in that direction (see [25], Problem 4). The point is that the combinatorial parts of all the above mentioned lower bounds proofs depend heavily on *specific* properties of concrete Boolean functions, and it was unclear if there are some *common* combinatorial properties of Boolean functions that do actually force their hardness.

In this paper we resolve this problem, and do this in quite general setting. We consider the following general model of monotone computations: gates may be arbitrary non-decreasing real-valued functions $\phi : \mathbf{R}^m \rightarrow \mathbf{R}$ ($m \geq 1$). We do not bound the fanin m . Rather, we require that these functions have bounded "degree". The degree of a gate is formally defined in Section 2. Here we mention only that it does not exceed the fanin but may be much smaller. In particular, a Boolean gate $\phi : \{0, 1\}^m \rightarrow \{0, 1\}$ has degree $\leq d$ if either all minterms or all maxterms (or both) have length at most d . For example, unbounded fanin AND and OR gates have degree 1. The degree of a threshold gate $T_s^m(x_1, \dots, x_m)$ does not exceed threshold value $\min\{s, m - s + 1\}$. We call a circuit *d-local* if all its gates have degree at most d .

Our main result is a general combinatorial lower bounds criteria for bounded degree unbounded fanin monotone real circuits (Theorems 1 and 2). Its restricted version has the following transparent form. For $A \subseteq \{0, 1\}^n$ and $\varepsilon \in \{0, 1\}$, let $\text{Min}_k(A, \varepsilon)$ and $\text{Max}_k(A, \varepsilon)$ denote respectively the minimum and the maximum possible number of vectors in A , all of which have the value ε on some fixed set of k coordinates.

*The work was supported by a DFG grant Me 1077/10-1. Preliminary version of this work appeared as an ECCC Technical Report #96-026 (March 1996)

[†]On leave from Institute of Mathematics, Vilnius, Lithuania.

Criterion: Let f be a monotone Boolean function in n variables, and let C be a monotone d -local real circuit. If C computes f then, for any $A \subseteq f^{-1}(0)$, $B \subseteq f^{-1}(1)$ and $1 \leq a, b \leq n/d$, C has size at least the minimum of

$$\frac{\text{Min}_b(A, 1)}{(db)^a \cdot \text{Max}_a(A, 0)} \quad \text{and} \quad \frac{\text{Min}_a(B, 0)}{(da)^b \cdot \text{Max}_b(B, 1)}$$

Apparently, this is the first such easy (and symmetric) lower bounds criterion for monotone computations. When applied to concrete Boolean functions, it directly yields exponential lower bounds for explicit functions in NP (we demonstrate this in Section 6). The simplicity of the whole argument (as well as of the criterion itself) may be somewhat surprising because the model we are dealing with is quite powerful: it is shown in [9] that *every* slice function has a linear size circuit with fanin-2 non-decreasing real gates, whereas easy counting yields that most of slice functions require exponential size Boolean circuits over the complete basis $\{\wedge, \vee, \neg\}$. Moreover, in the case of *unbounded fanin* gates previous lower bounds were known only for AND/OR gates under additional restriction that circuits have *constant-depth* (cf. [21]).

Our proof combines two ideas: the *bottlenecks counting* idea of Haken [11, 12, 13] and Sipser's idea of *finite limits* [28, 29]. The resulting argument becomes extremely simple and is different from Razborov's *method of approximations* [23, 24, 26], although the general idea remains the same: we try to map a large set of input vectors to gates in the circuit so that not too many vectors are mapped to any one gate. The mapping manages to hit "bottlenecks" in the circuit by sending an input vector to the first gate in the circuit for which this input is "hard" and which nevertheless classifies this input correctly. This "bottlenecks counting" idea is, of course, only a general scheme: its realization depends heavily on what inputs are declared to be "hard". To measure the "hardness" Haken [12] and Haken and Cook [13] use the concept of, so-called, *fences*. We take for this purpose another (apparently, more transparent) concept of *finite limit* due to Sipser [28, 29].

A vector x is a k -limit for a set of vectors A if on every subset of k coordinates, x coincides with at least one vector from A . Thus, if x is a k -limit for the set $f^{-1}(f(x) \oplus 1)$, then x is a "hard" instance for any circuit computing f since the value $f(x)$ cannot be determined when looking at only k bits of x . The key

of the whole argument is one simple "limit lemma" (Lemma 3) implying that in monotone circuits no single gate can make too large progress in classifying k -limits. If the function f is such that $f^{-1}(0)$ has many k -limits for $f^{-1}(1)$ (and vice versa) then the progress made by the whole circuit must be large, and hence, there must be many gates.

The paper is organised as follows. In Section 2 we describe the model of monotone circuits over the reals and define their *locality*. We formulate the criterion (Theorem 1) in Section 3 and prove it in Section 4. The proof is based on two lemmas: Reduction Lemma (Lemma 1) and Limit Lemma (Lemma 3). Just like in Razborov's method of approximations ([23, 26], the first lemma reduces the lower bounds problem for monotone local circuits to an appropriate SET COVER problem. Roughly, we prove that f requires circuits of size $\Omega(t)$ if neither $f^{-1}(0)$ nor $f^{-1}(1)$ may be covered by t "closed" sets. The difference from Razborov's method is the different nature of these "closed" sets: we define them in terms of finite limits. The main goal of the next Section 4.3 is to prove an explicit *upper bounds* on the size of closed sets. These bounds, together with the above mentioned reduction to a cover problem, directly yield the desired lower bounds criteria. In Section 5 we present the criterion in full generality.

In the last section we apply the criteria to explicit Boolean functions and derive exponential lower bounds for them. The difference from known lower bounds [21, 23, 3, 1, 31, 13, 22] for monotone circuits is twofold. First, we achieve these lower bounds in a *uniform and easy way*: all we need is to compute several very simple combinatorial characteristics of a given function. Second, and more important, our bounds hold for more general model of *local* monotone circuits over the reals. In particular, these circuits may contain as gates:

- arbitrary nondecreasing real-valued functions of growing (up to n^ϵ) fanin, and
- arbitrary *unbounded* fanin monotone Boolean functions whose minterms or maxterms (or both) have length $\leq n^\epsilon$.

In Boolean case these lower bounds supplement previous result due to Yao [31] that one needs super-polynomial size to compute $\text{CLIQUE}_{m,k}$ (which outputs 1 iff the input m -vertex graph has a clique on k vertices) even allowing gates capable to perform arbitrary monotone Boolean operation of fanin at most

m^ε . We show (cf. Corollary 2) that $\text{CLIQUE}_{m,k}$ requires super-polynomial size even if we allow gates be arbitrary monotone Boolean functions, whose minterms or maxterms (or both) have length at most $d = o\left(\frac{k}{\log^3 n}\right)$. Note that for $d = \binom{k}{2}$ the whole such circuit for $\text{CLIQUE}_{m,k}$ would consist of just one gate (with minterms corresponding to k -cliques).

Finally, let us mention that the results in the present paper have also an application to *cutting plane proofs* [8] in the propositional calculus. Cutting plane proofs provide a complete refutation system for unsatisfiable sets of propositional clauses. They efficiently simulate resolution proofs, and in fact are known to provide exponentially shorter proofs on some examples (the pigeonhole clauses). Bonet *et al* [6] and Pudlák [22] reduced the problem to lower bounds for circuits with nondecreasing real functions of fanin 2 as gates. Thus, our general lower bound for such circuits (Theorem 2), as well as lower bounds for explicit functions, are also lower bounds for the length of cutting plane proofs.

2. Preliminaries

In this section we recall some (more or less standard) notions concerning Boolean functions and circuits.

Let N be a set of n elements, called *bits*. Subsets of N are called *bit sets*. An *input* or a (binary) *vector* is a mapping $x : N \rightarrow \{0, 1\}$. A *Boolean function* is a mapping $f : \{0, 1\}^n \rightarrow \{0, 1\}$. The value $f(x)$ of f on an input x is defined by $f(x) = f(x(1), \dots, x(n))$. A *circuit* (or straight-line program) is a sequence $C = (g_1, \dots, g_t)$ of mappings¹ (called *gates*) $g_i : \{0, 1\}^n \rightarrow \mathbf{R}$ such that for every $i = 1, \dots, t$, gate g_i has the form $g_i = \phi(h_1, \dots, h_{m_i})$ where $\phi : \mathbf{R}^{m_i} \rightarrow \mathbf{R}$ is some mapping (called the *basis function*) and each h_j is either a Boolean variable or one of the previous gates g_1, \dots, g_{i-1} . The number m_i is a *fanin* of g_i . The number t of gates is the *size* of C . The function computed by C is the function g_t computed by the last gate.

Monotone circuits. Let, in what follows, f be arbitrary (but fixed) Boolean function. In order to define the monotonicity of circuits (computing this particular function f), we look at the behaviour of their gates g

¹The arithmetic structure of real numbers \mathbf{R} will not be used for the lower bounds, one can take any totally ordered set instead of \mathbf{R} .

on bipartite graphs G_g defined by:

$$G_g \Rightarrow \{(x, y) \in f^{-1}(0) \times f^{-1}(1) : g(x) \neq g(y)\}.$$

Intuitively, the larger G_g is, the better g "approximates" the function f . We say that a subgraph $E_g \subseteq G_g$ is *monotone* if it is possible to order all the inputs x_1, \dots, x_p in $f^{-1}(0)$ so that $E_g(x_1) \subseteq \dots \subseteq E_g(x_p)$.

Note 1. The monotonicity means that we actually can order inputs from *both* sides $f^{-1}(0)$ and $f^{-1}(1)$. Indeed, take the order y_1, \dots, y_q of inputs in $f^{-1}(1)$ where inputs from $E_g(x_p) \setminus E_g(x_{p-1})$ are listed first (in an arbitrary order), then list the inputs from $E_g(x_{p-1}) \setminus E_g(x_{p-2})$, etc., so that we again have that $E_g(y_1) \subseteq \dots \subseteq E_g(y_q)$.

Note 2. In Boolean case the graph G_g consists of two bipartite complete subgraphs (cliques):

$$\left(g^{-1}(\varepsilon) \cap f^{-1}(0)\right) \times \left(g^{-1}(\varepsilon \oplus 1) \cap f^{-1}(1)\right), \quad \varepsilon = 0, 1$$

whereas each monotone subgraph $E_g \subseteq G_g$ lacks at least one of these two cliques.

Let now C be a circuit, computing f . A *skeleton* of C is a set of bipartite graphs $\mathcal{M} = \langle E_g : g \in C \rangle$, one for each gate of C . Such a skeleton is *monotone* if all these graphs E_g are monotone. A skeleton is *legal* if for every gate $g = \phi(h_1, \dots, h_m)$ of C we have that $E_g \subseteq E_{h_1} \cup \dots \cup E_{h_m}$. Note that the trivial skeleton $\langle G_g : g \in C \rangle$ is always legal, because $(x, y) \in G_g$ implies that $g(x) \neq g(y)$, and hence, $h_i(x) \neq h_i(y)$ for at least one i .

Definition 1. Given a circuit $C = (g_1, \dots, g_t)$, which computes a Boolean function f , we say that C is *monotone* if there exists a legal monotone skeleton $\mathcal{M} = \langle E_g : g \in C \rangle$ such that $E_{g_t} = f^{-1}(0) \times f^{-1}(1)$.

Note that we do not require the basis functions $\phi : \mathbf{R}^m \rightarrow \mathbf{R}$ themselves be nondecreasing (this would be the standard definition). It is enough that the gates *behave* in a monotone manner during the computations (so that a legal monotone skeletons is possible). In general, finding such a skeleton is not an easy task. However, if we *a priori* know that all basis functions ϕ are nondecreasing then the task becomes trivial:

Proposition 1. Let C be a circuit computing f . If all basis functions $\phi : \mathbf{R}^m \rightarrow \mathbf{R}$ used in C are monotone nondecreasing, then C is monotone.

Proof: Take the following (standard) skeleton $\mathcal{M}_0 \rightleftharpoons \langle E_g : g \in C \rangle$ where

$$E_g \rightleftharpoons \{(x, y) \in f^{-1}(0) \times f^{-1}(1) : g(x) < g(y)\}.$$

It is easy to see that this skeleton is monotone: order the inputs $f^{-1}(0) = \{x_1, x_2, \dots, x_p\}$ in such a way that $g(x_1) \geq g(x_2) \geq \dots \geq g(x_p)$; then $E_g(x_1) \subseteq E_g(x_2) \subseteq \dots \subseteq E_g(x_p)$, as desired. Moreover, the graph E_{g_t} , associated with the last gate g_t , is complete since C computes f . So, it remains to verify the legality, i.e. that $E_g \subseteq \bigcup_{i=1}^m E_{h_i}$ for any gate $g = \phi(h_1, \dots, h_m)$. Indeed, if some edge (x, y) appears in *no* of the graphs E_{h_1}, \dots, E_{h_m} then $h_i(x) \geq h_i(y)$ for all $i = 1, \dots, m$. Since ϕ is nondecreasing, this implies that $g(x) \geq g(y)$, i.e. that $(x, y) \notin E_g$, as desired. ■

Degree and locality. Instead of the fanin, we will be interested in other characteristic of gates - their "degree". This characteristic depends on the function f computed by the whole circuit C , and on the choosed skeleton $\mathcal{M} = \langle E_g : g \in C \rangle$. Moreover, it is well defined only for legal skeletons.

We will denote by $E_g(x)$ the set of all neighbours of x in the graph E_g , i.e. the set of all inputs $y \in f^{-1}(f(x) \oplus 1)$ which are joined by an edge with input x in the graph E_g .

Let $g = \phi(h_1, \dots, h_m)$ be a gate in C . Since the skeleton \mathcal{M} is legal, every edge (x, y) of E_g is covered by (i.e. appears in) at least one of the input graphs E_{h_1}, \dots, E_{h_m} . We are interested in how many of these graphs are really needed for this.

Definition 2. The *degree* $\deg(x, g)$ of a gate $g = \phi(h_1, \dots, h_m)$ at an input x is the minimum of $|I|$ over all subsets $I \subseteq [m]$ such that $E_g(x) \subseteq \bigcup_{i \in I} E_{h_i}(x)$. The ε -*degree*, $\deg_\varepsilon(g)$ of a gate g is the maximum of $\deg(x, g)$ over all inputs $x \in f^{-1}(\varepsilon)$. The *degree* $\deg(g)$ of a gate g is the minimum of $\deg_0(g)$ and $\deg_1(g)$. A circuit C is d -*local* if $\deg(g) \leq d$ for all gates $g \in C$.

The degree of a gate $g = \phi(h_1, \dots, h_m)$ may heavily depend on its place in the whole circuit C and on the function f , computed by C . This is the general definition for which our argument works. It is, however, possible (at least in the Boolean case) to give a "circuit-independent" definition.

Let ϕ be a monotone Boolean function. Recall that a *minterm* (*maxterm*) of ϕ is a minimal subset $I \subseteq [m] \rightleftharpoons \{1, \dots, m\}$ such that setting some constants

to the variables x_i ($i \in I$), forces the value of ϕ to 1 (resp., to 0), independent on the values of the remaining variables. The length of a minterm (maxterm) is the number of variables in it.

Proposition 2. Let $g = \phi(h_1, \dots, h_m)$ be a gate in C and ϕ be a monotone Boolean function. If either all the minterms or all the maxterms of ϕ (or both) have length at most d then $\deg(g) \leq d$.

Proof: Assume w.l.o.g. that all the maxterms of ϕ have length at most d (the case of minterms is dual). Take an input $x \in f^{-1}(0)$. We want to show that $\deg(x, g) \leq d$. Since $g(x) = 0$, there is maxterm $I \subseteq [m]$ of ϕ such that $|I| \leq d$ and $\bigvee_{i \in I} h_i(x) = 0$ whereas $\bigvee_{i \in I} h_i(y) = 1$ for every input $y \in g^{-1}(1)$, and hence, for every input $y \in E_g(x)$. Thus, $E_g(x) \subseteq \bigcup_{i \in I} E_{h_i}(x)$, as desired. ■

If $g = \phi(h_1, \dots, h_m)$ where ϕ is the Boolean AND, then the maxterms of ϕ have length 1, and hence, $\deg_0(g) = 1$ (although $\deg_1(g)$ may be as large as the fanin m). For OR gates we have the dual situation. More generally, if $g = T_s^m(h_1, \dots, h_m)$ is a threshold gate (i.e. $g(x) = 1$ iff $\sum_{i=1}^m h_i(x) \geq s$) then $\deg(g) \leq \min\{s, m - s + 1\}$.

3. The criterion

Given a Boolean function f and a random n -bit string \mathbf{x} in $\{0, 1\}^n$, let

$$\text{Min}_k[\mathbf{x}, \varepsilon] \rightleftharpoons \min_{|S| \leq k} \Pr[f(\mathbf{x}) = \varepsilon \text{ and } \mathbf{x}(S) \equiv \varepsilon \oplus 1],$$

$$\text{Max}_k[\mathbf{x}, \varepsilon] \rightleftharpoons \max_{|S| \geq k} \Pr[f(\mathbf{x}) = \varepsilon \text{ and } \mathbf{x}(S) \equiv \varepsilon].$$

Define

$$H_f^\varepsilon(\mathbf{x}, a, b, d) \rightleftharpoons \frac{\text{Min}_b[\mathbf{x}, \varepsilon]}{(db)^a \cdot \text{Max}_a[\mathbf{x}, \varepsilon]}.$$

Note that

$$\text{Min}_b[\mathbf{x}, \varepsilon] \geq \Pr[f(\mathbf{x}) = \varepsilon] - b \cdot p(\mathbf{x}, \varepsilon) \quad (1)$$

where $p(\mathbf{x}, \varepsilon)$ is the maximum of $\Pr[\mathbf{x}(e) = \varepsilon]$ over all bits $e \in N$. Hence,

$$H_f^\varepsilon(\mathbf{x}, a, b, d) \geq \frac{\Pr[f(\mathbf{x}) = \varepsilon] - b \cdot p(\mathbf{x}, \varepsilon)}{(db)^a \cdot \text{Max}_a[\mathbf{x}, \varepsilon]}. \quad (2)$$

Theorem 1. Let f be a monotone Boolean function on n variables and let C be a monotone d -local real circuit

computing f . Then for any random inputs x, y and any integers $1 \leq a, b \leq n$, we have that

$$\text{size}(C) \geq \min \{H_f^0(x, a, b, d), H_f^1(y, b, a, d)\} \quad (3)$$

This theorem follows directly from Lemmas 1, 2 and 3 proved in the next section.

Remark 1. The criterion, stated in the introduction is a special instance of Theorem 1 for the case when inputs x and y are uniformly distributed in A and B , respectively.

Remark 2. We have shown in [15] that (at least for the case of fanin-2 AND/OR circuits) similar criterion can be derived using Razborov's method of approximations. This criterion was also based on $\text{Max}_a[x, \varepsilon]$ (although notation is slightly different). It is therefore remarkable that both methods – the method of approximations and the bottlenecks counting method – employ essentially the same combinatorial characteristic: the maximal possible number of minterms (maxterms) with a given number of literals in common. Thus, bottlenecks counting idea is not a new approach but rather an approximation method, although more symmetric and simpler. Quite recently, this connection between the two methods was made more explicit in [7, 2, 27]

4. The proof

Just like in Razborov's method of approximations, our first goal is to reduce the lower bounds problem to an appropriate SET COVER problem. For this purpose we first recall Sipser's notion of "finite limits", and use them to define "closed" subsets of $\{0, 1\}^n$. The Reduction Lemma (Lemma 1) reduces the lower bounds problem for the circuit size of f to the following question: how many closed subsets we need to cover $f^{-1}(0)$ and $f^{-1}(1)$? The desired criterion (Theorem 1) then follows immediately from the Limit Lemma (Lemma 3) which gives a (parameterised) *upper* bound on the size of closed sets.

4.1. Finite limits and closed sets

Definition 3. A *witness* of an input x against a set of inputs A is a set of bits $S \subseteq N$ such that for every $y \in A$ there is a bit $e \in S$ for which $y(e) \neq x(e)$. The *length*² of a witness S is its cardinality $|S|$. A witness

S is *legal* if $S \subseteq I(x)$ where $I(x)$ denotes the set of all bits e such that $x(e) = f(x)$. A *k-limit* for a set A is an input x such that $|S| \geq k + 1$ for any legal witness S of x against A .

Proposition 3. If $A \subseteq B$ and x is a *k-limit* for A then x is also a *k-limit* for B . If $A = A_1 \cup \dots \cup A_d$ and x is a *k-limit* for the whole set A then x is a $\lfloor k/d \rfloor$ -limit for at least one of the sets A_1, \dots, A_d .

Proof: The first claim is obvious. For the second claim, observe that if x would have a (legal) witness S_i of length $\lfloor k/d \rfloor$ against A_i , for all $i = 1, \dots, d$, then $S = S_1 \cup \dots \cup S_d$ would be a (legal) witness of x against the whole set A , and (by the monotonicity of norms) this witness would have length at most k . ■

Definition 4. A set of vectors $A = \{x_1, \dots, x_m\} \subseteq \{0, 1\}^n$ is (r, s) -closed if there exists a sequence of sets $\emptyset \neq B_1 \subseteq \dots \subseteq B_m \subseteq \{0, 1\}^n$ such that, for every $\nu = 1, \dots, t$

(i) input x_ν is an r -limit for B_ν ;

(ii) no input from B_ν is an s -limit for the set $\{x_\nu, \dots, x_m\}$.

Remark 3. This (somewhat wicked) notion of closure is necessary to capture the real case. For the case of (unbounded fanin!) *Boolean* circuits more transparent notion works (see [18]). Namely, in this case it is enough to declare A to be (r, s) -closed iff there is a set $B \neq \emptyset$ such that: (i) every input of A is an r -limit for B , but (ii) no input of B is a s -limit for A . With such a notion of closure the argument is much easier but the resulting criterion covers only the Boolean case.

4.2. Reduction lemma

The goal of this section is to reduce the lower bounds problem to an appropriate set-covering problem.

We say that a set of vectors $A \subseteq f^{-1}(\varepsilon)$ is *k-simple* if at least one vector in $f^{-1}(\varepsilon \oplus 1)$ has a legal witness of length k against A . Put otherwise, A is *k-simple* if there exists a set S of k bits such that every vector $x \in A$ takes the value $\varepsilon \oplus 1$ on at least one bit in S .

Definition 5. Let $\rho(f; a, b, d)$ denote the minimal t for which there exists a b -simple set $Y \subseteq f^{-1}(0)$ and (a, db) -closed sets $A_1, \dots, A_t \subseteq f^{-1}(0)$ such that

$$f^{-1}(0) \subseteq Y \cup A_1 \cup \dots \cup A_t.$$

²In Section 5 we will consider more general length measures.

Lemma 1. (Reduction Lemma) *Let f be a monotone function on n variables and let C be a monotone d -local real circuit computing f . Then for any integers $1 \leq k_0, k_1 \leq n/d$, we have that*

$$\text{size}(C) \geq \min \{ \rho(f; k_0, k_1, d), \rho(\neg f; k_1, k_0, d) \}.$$

Proof: Let $C = (g_1, \dots, g_t)$ be a monotone d -local real circuit, and suppose that C computes f , i.e. that $g_t = f$. Since C is monotone, there is a legal skeleton $\mathcal{M} = \langle E_i : i = 1, \dots, t \rangle$ which associates with each of its gates g_i ($i = 1, \dots, t$) a monotone graph (to simplify notation we write E_i instead of E_{g_i}):

$$E_i \subseteq \{ (x, y) \in f^{-1}(0) \times f^{-1}(1) : g_i(x) \neq g_i(y) \}$$

so that $E_t = f^{-1}(0) \times f^{-1}(1)$. If some input x is a k -limit for the set $E_i(x)$ of all its neighbours in the i -th graph E_i , then we can treat x as a "hard instance" for the i -th gate because this gate correctly separates x from all its neighbours, even though this requires knowledge of more than k bits. We will use this property (of being a limit for the set of own neighbours) to colour the nodes of the graph $f^{-1}(0) \times f^{-1}(1)$. We do this step-by-step going through the graphs E_1, \dots, E_t .

Initially no node is coloured.

At the i -th step ($i = 1, \dots, t$) we do the following. Let $\text{Easy}_{i-1} \subseteq \{0, 1\}^n$ denote the set of inputs which were *not* coloured during the first $i - 1$ steps, and let

$$E_i^*(x) \rightleftharpoons E_i(x) \cap \text{Easy}_{i-1}$$

stand for the set of those neighbours of x in E_i that survived all the previous $i - 1$ steps uncoloured. Due to the locality of C , we know that at least one of the degrees (0-degree or 1-degree) of the gate g_i is $\leq d$. Let

$$\sigma(g_i) \rightleftharpoons \begin{cases} 0 & \text{if } \deg_1(g_i) \leq d \\ 1 & \text{otherwise.} \end{cases}$$

We first colour some inputs from $f^{-1}(\sigma)$, where $\sigma = \sigma(g_i)$, and only then we turn to the other part $f^{-1}(\sigma \oplus 1)$:

- Colour a node $x \in f^{-1}(\sigma)$ iff $x \in \text{Easy}_{i-1}$ (i.e. x is still uncoloured) and x is a k_σ -limit for the set $E_i^*(x)$. Let Hard_i^σ denote the set of those inputs in $f^{-1}(\sigma)$ which were coloured during this phase of the i -th step.
- Colour a node $x \in f^{-1}(\sigma \oplus 1)$ iff $x \in \text{Easy}_{i-1}$ and x is a $k_{\sigma \oplus 1}$ -limit for the set $E_i^*(x) \setminus \text{Hard}_i^\sigma$

(of all those neighbours of x in E_i which were not coloured so far – neither during some of the previous $i - 1$ steps nor during the first phase of the i -th step).

For $\varepsilon \in \{0, 1\}$, let $\text{Hard}_i^\varepsilon \subseteq \text{Easy}_{i-1}$ denote the set of those inputs in $f^{-1}(\varepsilon)$ which were coloured during the i -th step. (These are the "hard instances" for the i -th gate g_i). We have that, for both $\varepsilon = 0$ and $\varepsilon = 1$,

$$f^{-1}(\varepsilon) \subseteq Y^\varepsilon \cup \text{Hard}_1^\varepsilon \cup \dots \cup \text{Hard}_t^\varepsilon,$$

where

$$Y^\varepsilon \rightleftharpoons f^{-1}(\varepsilon) \setminus \left(\bigcup_{i=1}^t \text{Hard}_i^\varepsilon \right)$$

is the set of all inputs in $f^{-1}(\varepsilon)$ which were hard for *no* of the gates in C .

Thus, Lemma 1 follows directly from the following two claims:

Claim 1: For some $\varepsilon \in \{0, 1\}$, the set Y^ε of uncoloured inputs in $f^{-1}(\varepsilon)$ is $k_{\varepsilon \oplus 1}$ -simple.

Claim 2: For every $\varepsilon \in \{0, 1\}$ and every $i = 1, \dots, t$, the set $\text{Hard}_i^\varepsilon$ is $(k_\varepsilon, dk_{\varepsilon \oplus 1})$ -closed.

Proof of Claim 1. If all the nodes in at least one of the parts $f^{-1}(0)$ or $f^{-1}(1)$ were coloured, there is nothing to do (empty set is simple). Suppose therefore that both parts have uncoloured nodes, and take an uncoloured node $x \in f^{-1}(\varepsilon \oplus 1)$ where $\varepsilon \rightleftharpoons \sigma(g_t)$. Since C computes f , the graph E_t , associated with the last gate of C , is complete. Thus,

$$E_t^*(x) = f^{-1}(\varepsilon) \setminus \left(\bigcup_{i=1}^{t-1} \text{Hard}_i^\varepsilon \right) \supseteq Y^\varepsilon.$$

The fact that x remained uncoloured means, in particular, that x was *not* a $k_{\varepsilon \oplus 1}$ -limit for this set $E_t^*(x)$. Thus, x must have a legal witness of length $\leq k_{\varepsilon \oplus 1}$ against $E_t^*(x)$, and hence – also against the set Y^ε , i.e. Y^ε must be $k_{\varepsilon \oplus 1}$ -simple.

Proof of Claim 2. First of all observe that $\text{Hard}_1^0 = \text{Hard}_1^1 = \emptyset$, i.e. that *no* node $x \in \{0, 1\}^n$ was coloured during the first step. Indeed, the first gate g_1 is just an l -th variable ($l \in \{1, \dots, n\}$), and hence, for every input $x \in \{0, 1\}^n$ we have that either $E_{g_1}(x)$ is empty or x has a legal witness $S = \{l\}$ against $E_{g_1}(x)$.

Let us now consider the remaining steps $i = 2, \dots, t$. We prove the claim for $\varepsilon = 0$ (the case $\varepsilon = 1$ is dual). To simplify notations, let $g_i = \phi(\{g_j : j \in J\})$ (hence $i > j$ for all $j \in J$) and set $A \rightleftharpoons \text{Hard}_i^0$. Our goal is to prove that A is (k_0, dk_1) -closed.

Since the graph E_i is monotone, we can list the inputs $A = \{x_1, \dots, x_m\}$ in such a way that

$$E_i(x_1) \subseteq \dots \subseteq E_i(x_m). \quad (4)$$

Since C is d -local, we know that either $\deg_1(g) \leq d$ or $\deg_0(g) \leq d$ (or both). Let us consider these two cases separately.

Case 1: $\deg_1(g) \leq d$, i.e. $\sigma(g_i) = 0$.

In this case A is the set of inputs which were coloured during the *first* phase of the i -th step. Take $B_\nu \rightleftharpoons E_i^*(x_\nu)$, for $\nu = 1, \dots, m$. The first condition (i) of Definition 4 is then satisfied by the definition of Hard_i^0 .

Let us now verify the second condition (ii) that no input from B_ν is a (dk_1) -limit for $\{x_\nu, \dots, x_m\}$. To see this, take an input $y \in B_\nu$ and suppose the opposite that y is a (dk_1) -limit for A . Since $y \in E_i(x_\nu)$, we have by (4) that $E_i^*(y) \supseteq \{x_\nu, \dots, x_m\}$. Hence, y is also a (dk_1) -limit for $E_i^*(y)$. On the other hand, since $f(y) = 1$ and $\deg_1(g_i) \leq d$, the degree of this input at the gate g does not exceed d . This means that we can find a subset $I \subseteq J$ such that $|I| \leq d$ and $E_i^*(y) \subseteq \bigcup_{j \in I} E_j^*(y)$. By Proposition 3, the input y must be a k_1 -limit for at least one of the sets $E_j^*(y)$ ($j \in J$), and hence, would already be coloured during the first $i - 1$ steps, a contradiction with $y \in \text{Easy}_{i-1}$.

Case 2: $\deg_1(g_i) > d$, i.e. $\sigma(g_i) = 1$.

We claim that in this case the set $A = \text{Hard}_i^0$ is (k_0, k_1) -closed (and hence, also (k_0, dk_1) -closed). Since $\sigma(g_i) = 1$, we know that A consists of those inputs in $f^{-1}(0)$ which were coloured during the *second* phase of the i -th step. That is, every input $x_\nu \in A$ was a k_0 -limit for the set $E_i^*(x) \setminus \text{Hard}_i^1$. Thus, the first condition (i) of Definition 4 is immediately fulfilled with $B_\nu \rightleftharpoons E_i^*(x_\nu) \setminus \text{Hard}_i^1$, for $\nu = 1, \dots, m$. To verify the second condition (ii), take an arbitrary $y \in B_\nu$. Since $y \in E_i^*(x_\nu)$ we know that $y \in \text{Easy}_{i-1}^1$, i.e. y was not coloured during the first $i - 1$ steps. Moreover, $y \notin \text{Hard}_i^1$ means that this input was not coloured also during the *first* phase of the i -th step, which means that y was not a k_1 -limit for $E_i^*(y)$. But again, $y \in E_i(x_\nu)$ together with (4) implies that $E_i^*(y) \supseteq \{x_\nu, x_{\nu+1}, \dots, x_m\}$. Thus, no input $y \in B$

is a k_1 -limit for the set $\{x_\nu, x_{\nu+1}, \dots, x_m\}$, i.e., A is (k_0, k_1) -closed, as desired.

This completes the proof of Claim 2, and thus the proof of Lemma 1. \blacksquare

4.3. Limit lemma

The goal of this section is to give a tractable *upper bound* on the size of closed and simple sets.

Lemma 2. *Let $Y \subseteq f^{-1}(\varepsilon)$ be a k -simple set. Then*

$$\Pr[x \in f^{-1}(\varepsilon) \setminus Y] \geq \text{Min}_k[x, \varepsilon].$$

Proof: Since Y is k -simple, there must be at least one set of bits S such that $|S| \leq k$ and $x(S) \neq \varepsilon \oplus 1$ for every $x \in Y$. Thus $\Pr[x \in f^{-1}(\varepsilon) \setminus Y] \geq \Pr[x \in f^{-1}(\varepsilon) \text{ and } x(S) \equiv \varepsilon] \geq \text{Min}_k[x, \varepsilon]$. \blacksquare

Lemma 3. (Limit Lemma) *Let $A \subseteq f^{-1}(\varepsilon)$ be an (r, s) -closed set. Then*

$$\Pr[x \in A] \leq s^r \cdot \text{Max}_r[x, \varepsilon].$$

To prove this lemma, we need the following simple fact about transversals. A k -critical transversal for a sequence of sets $\mathcal{F} = \{S_1, \dots, S_t\}$ is a set T for which there is an index $l \in \{1, \dots, t\}$ such that T intersects all the sets S_1, \dots, S_l but no its subset $T' \subseteq T$ with $|T'| \leq r$, does this.

Lemma 4. *Let $\mathcal{F} = \{S_1, \dots, S_t\}$ be a sequence of sets, each of cardinality at most s . Let \mathcal{T} be a family of r -critical transversals for \mathcal{F} . Then there is a family \mathcal{H}_r of at most s^r r -element sets such that every set from \mathcal{T} contains at least one set from \mathcal{H}_r .*

Proof: We will construct the desired family \mathcal{H}_r by induction on r . For $r = 1$ we can take as \mathcal{H}_1 the family of all one element sets $\{e\}$ with $e \in S_1$. This family has at most $|S_1| \leq r$ 1-element sets, as desired.

Suppose now that the family \mathcal{H}_{r-1} is already constructed. For a set H , let $\text{ext}(H) \rightleftharpoons \{T \in \mathcal{T} : T \supseteq H\}$. We can assume w.l.o.g. that $\text{ext}(H) \neq \emptyset$ for every set H in \mathcal{H}_{r-1} (if not, remove these redundant sets from \mathcal{H}_{r-1}). We construct the desired family \mathcal{H}_r by applying the following procedure to the family \mathcal{H}_{r-1} .

Take a set $H \in \mathcal{H}_{r-1}$ and choose the first index i such that $H \cap S_i = \emptyset$ but $T \cap S_i \neq \emptyset$ for all $T \in \text{ext}(H)$ (such an i exists since $|H| \leq r - 1$ and H is a subset of an r -critical transversal). Include in \mathcal{H}_r all the sets

$H \cup \{e\}$ with $e \in S_i$, remove H from \mathcal{H}_{r-1} and repeat the procedure to this smaller family $\mathcal{H}_{r-1} \setminus \{H\}$. No transversal in \mathcal{T} gets lost during this step, since every such transversal, containing this $((r-1)$ -element) set H , must contain at least one of the sets $H \cup \{e\}$ with $e \in S_i$. Since every set in \mathcal{H}_{r-1} produces at most $|S_i| \leq s$ new sets, the resulting family \mathcal{H}_s will have at most $s \cdot |\mathcal{H}_{r-1}| \leq s^r$ sets, as desired. ■

Proof of Lemma 3. Let $A = \{x_1, \dots, x_m\}$. Recall that the closeness of A means that there exists a sequence of sets $\emptyset \neq B_1 \subseteq \dots \subseteq B_t \subseteq f^{-1}(\varepsilon \oplus 1)$ such that

- (i) input x_i is an r -limit for B_i , and
- (ii) no input from B_i is an s -limit for the set $A_i \triangleq \{x_i, \dots, x_t\}$.

By (ii), every input from B_i has a legal witness of length at most s against the set A_i . That is, for every input $y \in B_i$ there is a subset of bits $S_{i,y} \subseteq I(y) = \{e : y(e) = \varepsilon \oplus 1\}$ such that $|S_{i,y}| \leq s$ and every input $x \in A_i$ takes the value $y(e) \oplus 1 = \varepsilon$ on at least one bit $e \in S_{i,y}$. This, in particular, means that for every $x \in A_i$, the set $I(x)$ intersects all the sets in the sequence $\mathcal{F}_i = \{S_{i,y} : y \in B_i\}$ (with sets $S_{i,y}$ arranged in arbitrary order). Now, for each $j = 1, \dots, m$ the input x_j belongs to all the sets A_1, \dots, A_j , and hence, the set $I(x_j)$ must intersect all the sets in the sequence $\mathcal{F}^j = \{\mathcal{F}_1, \dots, \mathcal{F}_j\}$. On the other hand, by (i), no r -element subset of $I(x_j)$ can do this, since any such subset would be a legal witness of x_j against B_j . Thus, for every $j = 1, \dots, m$, the set $I(x_j)$ is an r -critical transversal for the sequence \mathcal{F}^j , and hence, is such a transversal for the whole sequence \mathcal{F}^t . By Lemma 4 there must be a family \mathcal{H} of s^r r -element sets such that every set $I(x)$ with $x \in A$, contains at least one of them. Thus, $\Pr[x \in A] \leq$

$$\sum_{x \in A} \Pr[x \in A \text{ and } x(e) = \varepsilon, \forall e \in I(x)] \leq \sum_{H \in \mathcal{H}} \Pr[x \in A \text{ and } x(H) \equiv \varepsilon] \leq s^r \cdot \text{Max}_r[x, \varepsilon].$$

■

5. Generalization

Although Theorem 1 already yields exponential lower bounds for some explicit functions (cf. Corollary 1), we will state and sketch one more flexible its variant. The only difference is that we will allow one to

use arbitrary norms to measure the *length* of bit sets. Recall that up to now the length of a set S was simply its cardinality $|S|$ (cf. Definition 3).

By a *norm* we will mean any mapping $\mu : 2^N \rightarrow \{0, 1, \dots\}$ which is monotone under the set-theoretic inclusion, i.e. $S \subseteq T$ implies $\mu(S) \leq \mu(T)$. Given such a norm, the *length* of a set S is the number $\mu(S)$. The *deviation* of μ is the function $\lambda(t) = \max\{|S| : \mu(S) \leq t\}$. The *defect* of μ is the maximal length $c = \max\{\mu(\{e\}) : e \in N\}$ of a single bit. These two characteristics connect the length $\mu(S)$ of S with its cardinality:

$$\mu(S)/c \leq |S| \leq \lambda(\mu(S)). \quad (5)$$

We say that a bit-set T *respects* a norm μ if we cannot add a bit from outside the set T to no of its subsets without increasing their length, i.e. if $\mu(S \cup \{e\}) \geq \mu(S) + 1$ for any subset $S \subseteq T$ and any bit $e \notin T$. We say that an input x *respects* μ if the set $I(x)$ does this (recall that $I(x)$ is the set of those bits on which input x takes the value $f(x)$).

For example, if we take the trivial norm $\mu(S) = |S|$ then $c = 1$, $\lambda(t) = t$ and *every* input respects μ . In case of graphs, bits correspond to edges and one can, for example, take $\mu(S)$ to be the number of vertices incident to at least one edge from S . In this case $c = 2$, $\lambda(t) = \binom{t}{2}$ and only inputs, corresponding to cliques, will respect such a norm.

Given random input x , a norm μ and a set of inputs $A \subseteq f^{-1}(\varepsilon)$, let

- $\text{Min}_b[x, A, \mu]$ be the minimum over all sets S with $\mu(S) \leq b$, of $\Pr[x \in A \text{ and } x(S) \equiv \varepsilon \oplus 1]$;
- $\text{Max}_a[x, A, \mu]$ be the maximum over all sets S with $\mu(S) \geq a$, of $\Pr[x \in A \text{ and } x(S) \equiv \varepsilon]$.

Given a pair (μ_0, μ_1) of (not necessarily different) norms, we will be interested in the following characteristic of x :

$$F_f^\varepsilon(x, a, b, d) \rightleftharpoons \frac{\text{Min}_b[x, X^\varepsilon, \mu_{\varepsilon \oplus 1}]}{(d \cdot \lambda(bc))^a \cdot \text{Max}_a[x, X^\varepsilon, \mu_\varepsilon]} \quad (6)$$

where X^ε denotes the set of all inputs from $f^{-1}(\varepsilon)$ respecting the norm μ_ε ; c and λ are the defect and the deviation of $\mu_{\varepsilon \oplus 1}$. Given a random input x it is an easy task to find a lower bound for this characteristic. In particular, the numerator in (6) can be estimated by

$$\text{Min}_b[x, X^\varepsilon, \mu_{\varepsilon \oplus 1}] \geq \Pr[x \in X^\varepsilon] - \lambda(bc) \cdot p(x, \varepsilon) \quad (7)$$

where $p(x, \varepsilon)$ is the maximum of $\Pr[x(e) = \varepsilon]$ over all bits $e \in N$ (cf. the estimate (1)).

Theorem 2. Let f be a monotone Boolean function on n variables and let C be a monotone d -local real circuit computing f . Then for any random inputs \mathbf{x}, \mathbf{y} , any norms μ_0, μ_1 and any integers $1 \leq a, b \leq n$,

$$\text{size}(C) \geq \min \{F_f^0(\mathbf{x}, a, b, d), F_f^1(\mathbf{y}, b, a, d)\}. \quad (8)$$

The proof of Theorem 2 is similar to that of Theorem 1 using more general notion of finite limit (depending this time on the norm), and therefore – the following more general Limit Lemma instead of Lemma 3.

Definition 6. Let μ be a norm. A k -limit for a set A under μ is an input x such that $\mu(S) \geq k + 1$ for any legal witness S of x against A .

Note that the only difference from Definition 3 is that now we take $\mu(S) \geq k + 1$ instead of $|S| \geq k + 1$.

Lemma 5. (Limit Lemma; General Form) Let μ_1 and μ_2 be norms; c be the defect of μ_1 and λ be the deviation of μ_2 . Let $A = \{x_1, \dots, x_m\}$ be a sequence of inputs from $f^{-1}(\varepsilon)$, each of which respects the norm μ_1 , and suppose that there is a sequence of sets $\emptyset \neq B_1 \subseteq \dots \subseteq B_m \subseteq f^{-1}(\varepsilon \oplus 1)$ such that, for every $i = 1, \dots, m$

- (i) input x_i is an (rc) -limit for B_i under the norm μ_1 ,
- (ii) no input from B_i is an s -limit for the set $\{x_i, \dots, x_t\}$ under the norm μ_2 .

Then, for any random input \mathbf{x} ,

$$\Pr[\mathbf{x} \in A] \leq \lambda(s)^r \cdot \text{Max}_r[\mathbf{x}, A, \mu_1].$$

The proof of this lemma is almost the same as that of Lemma 3, taking more care on the possible deviation between the norm $\mu(S)$ and the cardinality $|S|$. Actually, the only place in the proof of Lemma 3 where the norm plays its role is the lemma about transversals (Lemma 2). Using the estimates (5) one can easily modify the proof of this lemma to the case of arbitrary norms.

Let μ be a norm. A k -critical transversal for \mathcal{F} under the norm μ is a set T , which respects μ and for which there is an index $l \in \{1, \dots, t\}$ such that T intersects all the sets S_1, \dots, S_l but no its subset $T' \subseteq T$ with $\mu(T') \leq k$ does this.

Lemma 6. Let \mathcal{F} be a sequence of bit sets, each of cardinality at most s . Let μ be a norm and c be its defect. Let \mathcal{T} be a family of (rc) -critical under μ transversals

for \mathcal{F} . Then there is a family \mathcal{H}_r of bit sets such that:

- (i) $|\mathcal{H}_r| \leq s^r$,
- (ii) $r \leq \mu(H) \leq rc$ for all $H \in \mathcal{H}_r$,
- (iii) every set from \mathcal{T} contains at least one set $H \in \mathcal{H}_r$.

(A proof of Lemmas 5 and 6 can be found in an earlier version of this paper [18]).

6. Two applications

In this section we demonstrate how using the general criteria (refcriterion-simple and Theorem 2) one can easily derive exponential lower bounds for explicit Boolean function. The main goal of these examples is to stress the *tractability*: given an explicit function, we need only to make some elementary computations.

6.1. Drawing polynomials

Andreev in [3] introduced the following monotone function on the variables $X = \{x_{i,j} : i, j \in GF(q)\}$ (q is a prime power): $\text{POLY}_{q,s}(X) = 1 \iff$ there is a polynomial $p(z)$ of degree at most $s-1$ over $GF(q)$ such that $x_{i,p(i)} = 1$ for all $i \in GF(q)$. He proved that any fanin 2 AND/OR circuit computing this function (for appropriate values of s) requires size at least $\exp(\Omega(n^{1/8-\epsilon}))$. Using Razborov's method of approximations, Alon and Boppana [1] were able to essentially improve this bound until $q^{\Omega(s)}$ for any $s \leq (q/\ln q)^{1/2}/2$; for maximal possible s this bound is exponential in $\Omega(n^{1/4}\sqrt{\ln n})$, and this is the largest known lower bound for 'natural' function in NP. This bound is almost optimal because q^{s+1} is the trivial upper bound for $\text{POLY}_{q,s}$ (this function is an OR of q^s monomials, each of q literals). Using our criterion we extend this lower bound to circuits with bounded degree but unbounded fanin monotone circuits over the reals.

Corollary 1. Let $s \leq (q/\ln q)^{1/2}/2$ and let C be a monotone d -local real circuit computing $\text{POLY}_{q,s}$. Then C has size $q^{\Omega(s/d)}$.

Proof: Let $f = \text{POLY}_{q,s}$. We are going to apply the criterion in its simple form (Theorem 1). Let \mathbf{x} be a random input, which on each point $e = (i, j)$ of $GF(q) \times GF(q)$ independently takes the value 0 with probability γ and takes the value 1 with probability $1 - \gamma$ (where γ is a parameter to be fixed later). Let \mathbf{y} be a random input distributed uniformly on the set of graphs of all polynomials over $GF(q)$ of degree at most

$s - 1$. We have only to calculate the values of H_f^0 and H_f^1 in (3).

For the first input \mathbf{x} we have that

$$\begin{aligned} \text{Min}_b [\mathbf{x}, 0] &\geq \Pr[f(\mathbf{x}) = 0] - b \cdot \Pr[\mathbf{x}(e) = 0] \\ &= 1 - q^s(1 - \gamma)^q - b\gamma. \\ \text{Max}_a [\mathbf{x}, 0] &= \max_{|S| \geq a} \Pr[f(\mathbf{x}) = 0 \text{ and } \mathbf{x}(S) \equiv 0] \\ &\leq \gamma^a. \end{aligned}$$

($\text{Max}_a [\mathbf{x}, 0]$ is the maximum probability that \mathbf{x} avoids a fixed set S of a points in $GF(q) \times GF(q)$).

For the second input \mathbf{y} we have that

$$\begin{aligned} \text{Min}_a [\mathbf{y}, 1] &\geq \Pr[f(\mathbf{y}) = 1] - a \cdot \Pr[\mathbf{y}(e) = 1] \\ &= 1 - a/q. \\ \text{Max}_b [\mathbf{y}, 1] &= \max_{|S| \geq b} \Pr[f(\mathbf{y}) = 1 \text{ and } \mathbf{y}(S) \equiv 1] \\ &\leq q^{-b}. \end{aligned}$$

($\text{Max}_b [\mathbf{y}, 1]$ is the maximum fraction of polynomials of degree at most $s - 1$, all of which coincide on some fixed set of b elements from $GF(q)$)

Taking $a \Leftarrow \lceil (s \ln q)/d \rceil$, $b \Leftarrow \lceil s/d \rceil$, and $\gamma \Leftarrow (2s \ln q)/q$ we get

$$H_f^0(\mathbf{x}, a, b, d) \geq \frac{1/2 - b \cdot \gamma}{(db)^a \gamma^a} \geq q^{\Omega(s/d)}$$

and

$$H_f^1(\mathbf{y}, b, a, d) \geq \frac{1 - a/q}{(da)^b q^{-b}} \geq q^{\Omega(s/d)},$$

and Theorem 1 gives the desired lower bound. ■

6.2. Detecting cliques

Let N be the family of all $n = \binom{m}{2}$ 2-element subsets (*edges*) of some set V of m vertices. This way every input $x : N \rightarrow \{0, 1\}$ can be identified with the undirected graph $G_x = (V, E)$ where $(u, v) \in E$ iff $x(u, v) = 1$. The *clique function* $\text{CLIQUE}_{m,k}$ is a monotone Boolean function on n variables, which given an input x computes 1 iff the graph G_x contains a k -clique, i.e. a complete subgraph on k vertices.

Using the method of approximations, Razborov in [23] proved the first super-polynomial lower bound $n^{\Omega(\log n)}$ for this function. Subsequently, Alon and Boppana [1], by strengthening the combinatorial part of Razborov's proof, were able to extend this bound until $2^{\Omega(k^{1/2})}$ for any $k \leq (m/8 \log m)^{2/3}$. For maximal possible k , this bound is exponential in $\Omega(n^{1/6}/(\log n)^{1/3})$.

These bounds in [1] were proved for usual model of fanin 2 AND/OR gates but Pudlák in [22] has shown that this proof can be extended to circuits with arbitrary monotone fanin 2 real functions as gates. Yao [31] considered monotone circuits with arbitrary monotone Boolean functions of fanin $\leq n^{1/100}$ as gates, and proved that any such circuits computing $\text{CLIQUE}_{m,k}$ with $k = \log \log m$, requires super-polynomial size. It appears that in case bounded degree, even *unbounded* fanin does not help.

Corollary 2. *Let $k \leq m^{2/3}(\ln m)^{1/3}/d^{1/3}$ and C be a monotone d -local real circuit. If C computes $\text{CLIQUE}_{m,k}$ then it has size exponential in $\Omega\left(\sqrt{\frac{k}{d \ln m}}\right)$.*

Proof: Let $f = \text{CLIQUE}_{m,k}$. We are going to apply Theorem 2 with the following pair of norms: $\mu_0(S) = |S|$ and $\mu_1(S)$ is the number of vertices incident to at least one edge from S . The defect and deviation for these norms are: $c_0 = 1$ and $\lambda_0(t) = t$ for μ_0 , and $c_1 = 2$ and $\lambda_1(t) = \binom{t}{2}$ for μ_1 .

Let \mathbf{x} be a random input, which on every bit takes independently the value 1 with probability $1 - \gamma$ where $\gamma = 4k^{-1} \ln(m/k)$. This input corresponds to a random graph $G_{\mathbf{x}}$ on m vertices, in which every edge appears independently with probability $1 - \gamma$. Let \mathbf{y} be a random input, uniformly distributed in the set of all (inputs corresponding to) k -cliques; thus \mathbf{y} is k -clique with probability $\binom{m}{k}^{-1}$. We have only to calculate the values of F_f^0 and F_f^1 in (8).

For the first input \mathbf{x} we have that $p(\mathbf{x}, 0)$ is the probability that the graph $G_{\mathbf{x}}$ avoids a single edge, and hence, $p(\mathbf{x}, 0) = \gamma$; $\text{Max}_a [\mathbf{x}, X^0, \mu_0]$ is at most the probability that $G_{\mathbf{x}}$ avoids some fixed set of a edges, and hence, is at most γ^a . Since $f(\mathbf{x}) = 1$ iff $G_{\mathbf{x}}$ contains a k -clique, we have, by the choice of γ , that $\Pr[f(\mathbf{x}) = 0] \geq 1 - \binom{m}{k}(1 - \gamma)^{\binom{k}{2}} \geq 2/3$. Moreover, $X^0 = f^{-1}(0)$ since μ_0 is the trivial norm. Since $\lambda_1(2b) \cdot \gamma \leq 1/3$ for any $b \leq (k/24 \ln(m/k))^{1/2}$, we have by (7) that the first term $F_f^0(\mathbf{x}, a, b, d)$ in (8) is at least

$$\frac{1}{3} \left(\frac{k}{8db^2 \ln(m/k)} \right)^a.$$

For the second input \mathbf{y} we have that $\Pr[\mathbf{y} \in X^1] = 1$ (since cliques respect the norm μ_1), and $p(\mathbf{y}, 1)$ is the probability that a random k -clique contains a fixed edge, and hence, is at most $\binom{m-2}{k-2}/\binom{m}{k} \leq (k/m)^2$;

$\text{Max}_b [y, X^1, \mu_1]$ is the probability that a random k -clique contains some fixed set of b vertices, and hence, is at most $\binom{m-b}{k-b} / \binom{m}{k}$. Thus, for any $a \leq (m/k)^2/2$, the second term $F_f^1(y, b, a, d)$ in (8) is at least

$$\frac{1}{2} \left(\frac{m}{dak} \right)^b.$$

Take $a \Rightarrow \lceil m/(2dk) \rceil$, $b \Rightarrow \lceil (k/24d \ln(m/k))^{1/2} \rceil$. For these values of a and b , we have that $F_f^0 \geq 2^{\Omega(a)} = 2^{\Omega(m/dk)}$ and $F_f^1 \geq 2^{\Omega(b)} = 2^{\Omega(\sqrt{k/d \ln m})}$, which by (8) gives the desired lower bound. ■

7. Conclusion and open problems

Finite limits have already been shown to provide a convenient framework in which to prove lower bounds for different models of computation: AC^0 -circuits [14], depth-three threshold circuits [17], multi-party protocols and syntactic read- k -times branching programs [16]. All these applications are based on an appropriate *Limit Lemma* about the existence of inputs in $f^{-1}(0)$ which are limits for $f^{-1}(1)$ (and vice versa). In each case this lemma captures the main combinatorial properties which make functions hard for that particular model. In some cases (like bounded depth circuits or read- k -times branching programs) this leads to new lower bounds, in other (like multi-party games) we get simpler proofs of known bounds.

In this paper we have shown that finite limits work also for monotone circuits. The whole argument is again based on appropriate limit lemma (cf. Lemma 3). Quite remarkably, this simple combinatorial lemma is enough not only to get exponential lower bounds but also to formulate a transparent lower bounds criterion for quite general models of monotone computations. It would be therefore interesting to understand to what extent limits can help in the presence of negation. One possibility here would be to relax the legality constraint for witnesses. The legality we used in this paper enables one to treat differently 0's and 1's in inputs from different parts $f^{-1}(0)$ and $f^{-1}(1)$. This makes the criterion easy to apply, but cannot handle negation gates, i.e. gates switching the role of 0's and 1's.

Karchmer in [19] and Karchmer & Wigderson [20] established an interesting connection between Razborov's *generalised method of approximations* [26] and the *ultraproduct construction* in model theory. In this approach one uses ultra-filters (or filters, in mono-

tone case) to construct a consistent accepting computation for an input which should be rejected. This 'diagonal' computation can be also looked as a limit for the set of all rejecting computations. It would be interesting to re-state our argument in this frame to better understand the different (yet equal) methods.

There are also some more concrete open problems. Let us mention two of them.

1. Large vs. small degree circuits. We have seen that monotone circuit complexity of $\text{CLIQUE}_{m,k}$ depends heavily on the degree d of gates used in a circuit: if $d = \binom{k}{2}$ then $\text{CLIQUE}_{m,k}$ has monotone d -local Boolean circuit of size 1, whereas it requires super-polynomial size if $d = o(k/\log^3 n)$. Can this gap be essentially improved?

2. Boolean vs. real circuits. We have already mentioned in the introduction that some monotone Boolean functions (namely – some slice functions) have small real monotone circuits of constant fanin but require exponential size Boolean circuits (even over the complete basis $\{\wedge, \vee, \neg\}$). Unfortunately, we are not able exhibit any of such functions explicitly (this could imply $P \neq NP$). It would be therefore interesting to find an *explicit* monotone Boolean function f which has small real monotone circuits (of constant fanin) but requires large Boolean monotone circuits. As a possible candidates let us mention the following two functions on $n = m^2$ variables: the *perfect matching* function PM_n and the *odd factor* function OF_n . The input for both functions is an $m \times m$ (0,1)-matrix representing a bipartite graph X with m vertices in each part. Graph X is accepted by PM_n if it has a perfect matching, whereas for X to be accepted by the second function OF_n it is enough that X has an odd factor, i.e. a spanning subgraph such that all vertices have odd degree in the subgraph (thus X is rejected by OF_n exactly if it has a component with an odd number of vertices). Both these functions require monotone fanin-2 Boolean circuit of size $n^{\Omega(\log n)}$ (see [24, 5]) whereas their real circuit size is not clear.

Acknowledgments

I thank Armin Haken, Alexander Razborov and Avi Wigderson for interesting discussions and remarks.

References

- [1] N. Alon and R. Boppana, The monotone circuit com-

- plexity of Boolean functions, *Combinatorica*, 7:1 (1987), pp. 1-22.
- [2] K. Amano and A. Maruoka, Potential of the approximation method. In: *Proc. of the 37th Ann. IEEE Symp. Found. Comput. Sci.*, 1996.
- [3] A.E. Andreev, On a method for obtaining lower bounds for the complexity of individual monotone functions. *Soviet Math. Dokl.* 31(3):530-534, 1985.
- [4] ———, On one method of obtaining effective lower bounds of monotone complexity. *Algebra i logika*, 26(1):3-21, 1987. In Russian.
- [5] L. Babai, A. Gál and A. Wigderson, Superpolynomial lower bounds for monotone span programs. Manuscript, 1996.
- [6] M. Bonet, T. Pitassi, and R. Raz, Lower bounds for cutting planes proofs with small coefficients. In: *Proc. Twenty-seventh Ann. ACM Symp. Theor. Comput.*, (1995), 575-584.
- [7] Ch. Berg and S. Ulfberg, Symmetric approximation arguments for monotone lower bounds without sunflowers, (manuscript, 1996)
- [8] W. Cook, C. R. Coullard, and Gy. Turán, On the complexity of cutting plane proofs. *Disc. Appl. Math.*, (1987), 25-38.
- [9] S. Cook and A. Rosenbloom, Some results on monotone real circuits, (manuscript, 1996)
- [10] P. Erdős and R. Rado, Intersection theorems for systems of sets. *Journal of London Math. Society* **35** (1969), 85-90.
- [11] A. Haken, The intractability of resolution, *Theor. Comp. Sci.*, **39** (1985), 297-308.
- [12] ———, Counting Bottlenecks to Show Monotone $P \neq NP$, In *Proc. of the 36th Ann. IEEE Symp. Found. Comput. Sci.*, 1995.
- [13] A. Haken and S. Cook, An exponential lower bound for the size of monotone real circuits, Manuscript, 1995.
- [14] J. Håstad, S. Jukna and P. Pudlák, Top-down lower bounds for depth-three circuits, *Computational Complexity*, **5** (1995), 99-112.
- [15] S. Jukna, A criterion for monotone circuit complexity, (unpublished manuscript, 1991) [Available at: <http://www.informatik.uni-trier.de/~jukna/papers/>]
- [16] ———, Finite limits and lower bounds for circuit size, Tech. Rep. Nr. 94-06, Informatik, University of Trier, 1994.
- [17] ———, Computing threshold functions by depth-3 threshold circuits with smaller thresholds of their gates, *Information Processing Letters*, **56** (1995), 147-150.
- [18] ———, Finite limits and monotone computations over the reals. ECCC Technical Report #96-026-01 (April 1996)
- [19] M. Karchmer, On proving lower bounds for circuit size, In *Proc. of the 8th Ann. Symp. on Structure in Complexity Theory*, (1993).
- [20] M. Karchmer and A. Wigderson, Characterizing non-deterministic circuit size. In *Proc. of the 25th ACM Symposium on Theory of Computing*, (1993)
- [21] M. Klawe, W. Paul, N. Pippenger, and M. Yannakakis, On monotone formulae with restricted depth. In *Proc. of 16th ACM Symp. on Theory of Computing* (1984), 480-487.
- [22] P. Pudlák, Lower bounds for resolution and cutting planes proofs and monotone computations. Submitted to *Journal of Symbolic Logic*, 1995.
- [23] A. A. Razborov, Lower bounds for the monotone complexity of some Boolean functions, *Soviet Math. Dokl.*, 31:354-357, 1985.
- [24] ———, Lower bounds of monotone complexity of the logical permanent function. *Mathem. Notes of the Academy of Sci. of the USSR*, 37:485-493, 1985.
- [25] ———, Lower bounds on the monotone complexity of Boolean functions. In: *Proc. of Int. Congress of Mathematicians* (Berkeley, California, USA, 1986), 1987, pp. 1478-1487.
- [26] ———, On the method of approximations. In *Proc. of the 21th Ann. ACM Symp. Theor. Comput.*, (1989), pp. 167-176.
- [27] J. Simon and S. Tsai, A note on the bottlenecks counting argument. In: *Proc. of 12-th Ann. IEEE Conf. on Computational Complexity*, 1997 (this volume).
- [28] M. Sipser, A topological view of some problems in complexity theory. In *Colloquia Mathematica Societatis János Bolyai* **44** (1985), pp 387-391.
- [29] ———, *Personal communication*, (1991)
- [30] É. Tardos, The gap between monotone and non-monotone circuit complexity is exponential, *Combinatorica*, 7:4 (1987), pp. 141-142
- [31] A. C. Yao, Circuits and local computations, In *Proc. 21th Ann. ACM Symp. Theor. Comput.*, (1989), 186-196.