Localizability of the approximation method

Ján Pich[*]

University of Oxford

June 2024

Abstract

We use the approximation method of Razborov to analyze the locality barrier which arose from the investigation of the hardness magnification approach to complexity lower bounds. Adapting a limitation of the approximation method obtained by Razborov, we show that in many cases it is not possible to combine the approximation method with typical (localizable) hardness magnification theorems to derive strong circuit lower bounds. In particular, one cannot use the approximation method to derive an extremely strong constant-depth circuit lower bound and then magnify it to an NC^1 lower bound for an explicit function.

To prove this we show that lower bounds obtained by the approximation method are in many cases localizable in the sense that they imply lower bounds for circuits which are allowed to use arbitrarily powerful oracles with small fan-in.

1 Introduction

Approximation method. In the 1980s Razborov [15, 16] initiated an approach to proving lower bounds on the size of Boolean circuits known as the *approximation method*. The idea of the method was to approximate each small circuit by an 'approximating' circuit, by replacing each connective of the original circuit by an 'approximating' connective, and then show that the approximating circuits are too weak to compute a given Boolean function f. Consequently, each small circuit had to make an error in computing f as well. This strategy was successfully implemented in lower bounds for restricted classes of circuits such as monotone circuits [15, 16] or circuits of constant depth [17, 21].

The formal framework of the approximation method was suitable also for a systematic analysis of potential ways of proving circuit lower bounds. In [18] Razborov gave a remarkable example of such meta-analysis of circuit complexity by showing that the

---

[*]jan.pich@cs.ox.ac.uk

approximation method cannot prove better than $O(nn_0)$-size lower bounds for general circuits, for any Boolean function f, where n_0 is the number of essential variables of f. The essential variables are the variables on which the function depends. Moreover, if we consider a particular form of the approximation method which was used to derive lower bounds for monotone and constant-depth circuits, the best lower bound we can hope for is $24(n_0 + 1)$. Despite these limitations, Razborov [18] showed that the approximation method is in certain sense complete: If a Boolean function f is not computable by any circuit of size s^3, one can use the approximation method to prove an $\Omega(s)$-size circuit lower bound for f. This requires introducing many inessential variables and proving the lower bound for the new function with extra variables. The complete version of the approximation method became known as the fusion method [22].

Natural proofs. Unfortunately, the promise of the fusion method has not been fulfilled as strong complexity lower bounds for explicit Boolean functions remain elusive. A significant part of the reason for this situation can be attributed to the natural proofs barrier of Razborov and Rudich discovered in the 1990s [20]. Intuitively, natural proofs say that the existing circuit lower bounds do not prove just that a particular function requires big circuits but they yield even efficient algorithms rejecting all easy functions while accepting many hard functions. The existence of such an algorithm for general Boolean circuits (formally, a P/poly-natural property against P/poly) would, however, break cryptographic pseudorandom generators.

Hardness magnification. A generic approach to lower bounds which seems to avoid the natural proofs barrier was investigated in more recent years in the area of *hardness magnification*. Hardness magnification is an approach to strong complexity lower bounds by reducing them to lower bounds against weak computational models. While some forms of magnification can be traced back to early 2000s, the term was coined by Oliveira and Santhanam [13] in 2018 to refer to its newer instantiations, see [6] for a more comprehensive exposition. In particular, some subsequent papers established so called HM frontiers, which refer to theorems including three statements of the following kind.

HM frontier for AC^0-XOR:

1. If $MCSP[n^c, 2n^c] \notin AC^0\text{-}XOR[N^{1.01}]$ for some c, then $EXP \not\subseteq NC^1$ [6, §3.2.2].
2. $MAJ \notin AC^0\text{-}XOR[2^{n^{o(1)}}]$, cf. [17, 21].
3. $MCSP[n^c, 2n^c] \notin AC^0$, if c is sufficiently large [6, Theorem 52].

Here, $MCSP[s, t]$ is the promise problem of determining if an $N := 2^n$-bit input has circuit complexity at most s versus at least t. $AC^0\text{-}XOR[s]$ denotes the class of s-size constant-depth circuits with XOR gates at the bottom layer, where the XOR gates compute parity functions with arbitrary fan-in.

Items 2 and 3 thus indicate that proving the lower bound required in Item 1 could be within reach as both the corresponding circuit class and computational problem are prone to a nontrivial analysis. HM frontiers can be established for many computational models instead of $\mathsf{AC}^0\text{-}\mathsf{XOR}$ and for various conclusions in Item 1, e.g. $\mathsf{NP} \not\subseteq \mathsf{NC}^1$ [6].

The reason why magnification theorems seem to overcome the natural proofs barrier is that Item 1 relies on specific properties of the function in question (in our case, MCSP), it is not clear how to generalize it to a substantial fraction of functions. In fact, this intuition was formally supported in [6] by showing that hardness magnification is in certain cases inherently nonnaturalizable in the sense that the conclusion in Item 1 is the non-existence of natural proofs against $\mathsf{P/poly}$.

Locality barrier. Unsurprisingly, hardness magnification suffers from its own barrier, the *locality barrier*.[1] As it turns out, Items 1 in HM frontiers are always obtained by constructing an oracle upper bound. For example, in our case, Item 1 follows from showing that $\mathsf{MCSP}[n^c, 2n^c]$ can be computed by an $\mathsf{AC}^0\text{-}\mathsf{XOR}[O(N)]$-circuit using oracles with fan-in N^ϵ, for some $\epsilon < 1$. On the other hand, Items 2 and 3 can be generalized so that they work against circuits with such oracles. We refer to oracles with small fan-in as *local* oracles and to magnification theorems obtained by constructing oracle upper bounds as *localizable* magnifications.

The fact that many known circuit lower bounds can be extended to models allowing arbitrary local oracles is interesting independently of hardness magnification. Proving the non-existence of subexponential-size learning algorithms for $\mathsf{P/poly}$ would imply the non-existence of $\mathsf{P/poly}$-natural properties against $\mathsf{P/poly}$ [5], but it is not hard to see that natural properties against $\mathsf{P/poly}$ are computable by a single local oracle applied on a prefix of the input. Overcoming the locality barrier is thus essential for proving strong complexity lower bounds of our interest.

There are examples of non-localizable lower bounds but they suffer from drawbacks such as that they do not achieve an HM-frontier or they yield at best uniform lower bounds or 'non-explicit' lower bounds of the form $\mathsf{QP} \not\subseteq \mathsf{P/poly}$, where QP stands for quasipolynomial time. We refer to [14, §5] for a more detailed discussion of these exceptions including non-localizable magnification theorems.

The motivating question of the present paper is to understand the extent of localizability of circuit lower bounds. Do 'concrete' sufficiently strong complexity lower bounds always localize? We address the question by using the approximation method as a convenient substitute for the informal notion of a concrete lower bound.

---

[1]Lower bounds from Item 3 of HM frontiers are affected also by the black-box natural proofs barrier [7].

1.1 Our contribution

Unlike Razborov [18, 19], who investigated general properties of the approximation method for unrestricted circuits, we focus primarily on analyzing the approximation method for weak circuit classes such as AC^0. This is motivated by hardness magnification. Nevertheless, we start by observing a couple of properties of the approximation method for unrestricted circuits.

1.1.1 Limitations: Localizability

Lower bounds based on the approximation method are proved by showing lower bounds for a measure $\rho(f, \mathcal{M})$ which expresses a distance of the target function f for which we want to prove a lower bound and a suitable approximation model $\mathcal{M}$. In fact, $\rho(f, \mathcal{M})$ is a lower bound on the circuit complexity of f. The above-mentioned barrier result of Razborov [18] says that for each legitimate model $\mathcal{M}$ and each Boolean function f, $\rho(f, \mathcal{M}) \leq O(nn_0)$. This leaves open the possibility of proving a superpolynomial circuit lower bound for f by proving a superlinear lower bound for $\rho(h, \mathcal{M})$, for another Boolean function h, and then magnifying it.

Localizability of general circuit lower bounds. We observe that such attacks are ruled out for localizable magnification theorems under additional assumptions such as that the number of oracles in the oracle circuit implying the magnification theorem is sublinear. In other words, we observe that superlinear lower bounds for general circuits based on the approximation method yield lower bounds for general circuits with a limited number of local oracles, cf. Theorem 3.

Limitations of the approximation method for constant-depth circuits. Next, we consider the question of deriving NC^1 lower bounds via subexponential-size constant-depth circuit lower bounds based on the approximation method. We do not rule out such possibility but we show that Razborov's barrier result can be adapted to the case of constant-depth circuits with nontrivial conclusions. Informally, we show that for each legitimate model $\mathcal{M}$ and each Boolean function f, $\rho_d(f, \mathcal{M}) \leq O(2^{8n/d})$, cf. Theorem 4. Here, $\rho_d(f, \mathcal{M})$ denotes a version of ρ for circuits of depth d. In particular, $\rho_d(f, \mathcal{M})$ is a lower bound on the size of d-depth circuits computing f.

Localizability of constant-depth circuit lower bounds. Our main theorem shows that localizability is an inherent property of all sufficiently strong lower bounds for circuits of constant depth obtained via the approximation method.

Theorem 1 (Localizability of the approximation method for constant-depth circuits - Informal, cf. Theorem 5). *Let $\mathcal{M}$ be any legitimate model of constant-depth circuits and f any Boolean function with n inputs. Suppose that $\rho_{O(d^2)}(f, \mathcal{M}) \geq s$. Then f is not computable by depth d circuits of size $s - poly(kdn(2^{m/d} + 1))$ using k arbitrarily powerful*

oracles of arity m. The constants in the poly($\cdot$)-notation are absolute.[2]

Note that showing that f is not computable by $poly(n2^m)$-size circuits of depth d implies that f is not computable by p-size circuits of depth $d/3$ with oracles of arity m because we can replace such oracles by DNFs of size $2^{O(m)}$. Theorem 1 shows that even $poly(n2^{m/d})$-size lower bounds localize, if they are obtained by the approximation method.

1.1.2 Completeness: Extent of Fusion

In order to prove an $s(n)$-size circuit lower bound for a Boolean function f with n inputs using the approximation method, it is necessary to introduce inessential variables, i.e. to consider function $f'(x_1, \ldots, x_N) := f(x_1, \ldots, x_n)$, for $N \geq s(n)$, and prove $\rho(f', \mathcal{M}) \geq s(n)$. Razborov [18] showed that this strategy can be adapted (using the asymmetric definition of ρ) so that it becomes complete. Wigderson [22], following Karchmer's [9] interpretation of Razborov's construction, coined the term *fusion method* for the resulting approach to circuit lower bounds.

Feasibility of fusion. The construction from [18] uses $N = O(s2^n)$. We observe that the fusion method can be adapted to the case of partial Boolean functions so that the corresponding approximation models use $N = poly(s)$, cf. Theorem 7. The adaptation relies on the notion of anticheckers of Lipton and Young [10]. The resulting fusion method inherits completeness from the standard fusion method. If anticheckers can be generated efficiently (which can be done in the case of SAT under standard hardness assumptions), the corresponding approximation models are constructive, see §4.1.

Fusion for constant-depth circuits. Finally, we define a version of the fusion method for constant-depth circuits and show that it is complete for constant-depth circuit lower bounds, cf. Theorem 8.

1.1.3 Proof methods

Informally, the measure $\rho(f, \mathcal{M})$ is defined as the minimal number of 'error sets' needed to cover all inputs on which f differs from its 'approximator'. To get a good lower bound, one wants to define a model $\mathcal{M}$ with small error sets but with approximators making a lot of errors. If f is computable by an s-size circuit W, the approximator of W and error sets corresponding to the gates of W can be used to witness that $\rho(f, \mathcal{M}) \leq s$.

Razborov's proof of $\rho(f, \mathcal{M}) \leq O(nn_0)$ proceeds by constructing suitable distributions on error sets and approximators, which allow us to derive the desired conclusion by a case analysis: For each input x, either I. the probability that a random approximator fails to compute f on x is small or II. the probability that a random error set covers x is high. Then, we can take a majority of random approximators and obtain an approximator

---

[2]We work with a 'symmetric' definition of ρ_d, cf. Definitions 4 and 6, but Theorem 1 holds for 'asymmetric' ρ_d as well, if we adjust the parameters appropriately.

coinciding with f on all inputs from case I. The remaining inputs from case II are covered by a small number of error sets. The construction of the distribution of error sets and approximators is based on an exponentially-big 'universal' circuit which plays the role of circuit W from the previous paragraph.

Theorem 1 is proved by the contrapositive. We assume that there is a small constant-depth oracle circuit computing f and we want to bound $\rho_{O(d^2)}(f, \mathcal{M})$. If the constant-depth circuit did not contain any oracles, its approximator would yield a small set of error sets covering all inputs on which the approximator fails to compute f. The problem is to deal with local oracles.

The first step of the proof is to adapt Razborov's barrier to ρ_d, i.e. to show $\rho_d(h, \mathcal{M}) \leq O(2^{8n/d})$ for every Boolean function h with n inputs. To make this work we need to use suitable constant-depth circuits as an alternative to the exponential-size circuits in Razborov's construction.

Having $\rho_d(h, \mathcal{M}) \leq O(2^{8n/d})$ and the assumption that there is a small oracle circuit C for f, we want to obtain an upper bound on $\rho_{O(d^2)}(f, \mathcal{M})$. Intuitively, we want to replace each oracle in C computing a function h with m inputs by $\rho_d(h, \mathcal{M}) \leq O(2^{8m/d})$. Unfortunately, we do not have a generic way to derive such upper bound just from an upper bound on $\rho_d(h, \mathcal{M})$. To make the proof work we need to exploit the inner structure of Razborov's proof. More specifically, we need to observe that Razborov's construction can be applied 'inside' the oracle circuit so that we can replace each oracle by the majority of its random approximators and still make the case analysis work. This requires one more modification: we need to use Ajtai's p-size constant-depth circuits for the approximate majority instead of general circuits [1].

1.2   Related work

Our results are motivated primarily by the development of hardness magnification and the approximation method, but the notion of locality was considered already by Yao [23], who observed that some concrete monotone circuit lower bounds such as the lower bound of Razborov [15] localize. Yao's observations are incomparable to our results: Our results are not restricted to specific instantiations of the approximation method, but they work only for non-monotone circuits. Hrubeš and Rao [8] considered the model of circuits with oracle gates of restricted fan-in and showed that some known lower bounds such as Nechiporuk's lower bound for formulas localize. Again, these results differ from ours in that they derive localizability only for specific examples of lower bounds.

1.3   Open problems

Optimal localizability. Do superlinear lower bounds for general circuits based on the approximation method localize even with a linear number of oracles (cf. Problem 1)?

Regarding constant-depth circuits, Theorem 1 localizes lower bounds for depth d circuits of size $poly(n2^{8m/d})$. In the HM frontier in the introduction, Item 1 uses $m = poly(\log n)$ while the lower bound in Item 3 is just polynomial. Is it possible to localize lower bounds for depth d circuits of size $poly(n2^{m^{O(1/d)}})$? A stronger localizability of this kind could follow from a positive answer to Problem 2, which asks for a barrier for deriving NC^1 lower bounds via the approximation method. It would be interesting to obtain such localizations even under additional assumptions of a constructivity of approximation models, e.g. assuming that the lower bounds based on these models are provable in S^1_2, a theory of bounded arithmetic formalizing p-time reasoning [4].

Random restrictions as the approximation method. Is it possible to show that the approximation method is complete for constant-depth circuit lower bounds? Can we show that the method of random restrictions is formalizable in the framework of the approximation method? In both cases, we allow using inessential variables. Theorem 8 suggests that such models should exist, but we have not managed to construct them. If we obtained such models with $O(\log s)$ inessential variables, they would be subject to the localizability from Theorem 1 (with the asymmetric version of ρ_d). This would extend the generality of the approximation method and the impact of the localizability from Theorem 1. In particular, it would provide a unifying explanation for the localizability of both Razborov-Smolensky lower bound method and the method of random restrictions.

Feasible and structured models. Razborov [19] introduced a generalization of the approximation method which can prove s-size lower bounds for almost all functions f with error sets of size $poly(s)$. This required detaching errors sets from the inputs of f so that the earlier barrier result would not apply. A drawback of the generalized approximation method is that it is not known to be complete and the models from [19] were obtained nonconstructively. Nevertheless, the elegance and restrictiveness of Razborov's generalized approximation models could be instrumental when searching for concrete examples of approximation models. It could thus be useful to find a constructive version of Razborov's models. Here, constructivity means the existence of a p-time algorithm which given truth-tables of functions f_1, f_2 of the model, outputs the truth-table of the 'approximated' conjuction/disjunction of f_1 and f_2. This question was posed already in [19]. In order for the approximation models obtained from the fusion method for partial functions to be constructive, we need efficient algorithms generating anticheckers. Can we obtain such models unconditionally?

Nonlocalizable lower bounds. I believe that Barrington-Straubing [3] lower bound for $\Omega(n \log \log n)$-size formulas of constant-depth with parity gates can be used to show the hardness of $\mathsf{MCSP}[n^c, 2n^c]$ for $\Omega(n \log \log n)$-size formulas of constant-depth with parity gates at the bottom. Since $\mathsf{MCSP}[n^c, 2n^c]$ is computable by $\mathsf{AC}^0\text{-}\mathsf{XOR}[O(N)]$ with local oracles, if we could make the lower bound work for circuits instead of formulas, it would

give us an example of a concrete nonlocalizable lower bound. There are several other lower bounds which work only for computational models of nearly linear size. Is it possible that the weakness of these lower bounds is a result of their sensitivity to the presence of local oracles? If so, their 'weakness' would turn out to be their strength.

2 Preliminaries

$[n]$ denotes $\{1, \ldots, n\}$. Let $\mathcal{F}_n$ be the set of all Boolean functions on n inputs. We identify a Boolean function $f \in \mathcal{F}_n$ with the set of its ones $\{x \mid f(x) = 1\}$.

A Boolean circuit with inputs $x_1, \ldots, x_n$ is a directed acyclic graph (without multi-edges) with vertices labeled by inputs $x_1, \ldots, x_n$, constants 0,1 or symmetric[3] connectives $\circ(y_1, \ldots, y_m) : \{0, 1\}^m \mapsto \{0, 1\}$ so that there are exactly m edges directed to the vertex labeled by $\circ(y_1, \ldots, y_m)$. Non-input vertices are called gates. One of the gates of the circuit is designed as the output gate. A Boolean circuit thus computes a Boolean function. We consider only Boolean circuits with a single output gate. The depth of a circuit is the number of edges on the longest path from the output to an input or to a constant. The size of a circuit is the number of its gates.

A Boolean formula is a Boolean circuit with a tree as the underlying graph and with logical connectives $\neg$, binary $\wedge$, binary $\vee$. We make an exception and measure the size of a Boolean formula by the number of leafs of its underlying tree.

A circuit class $\mathcal{C}$ is just a set of circuits - we do not impose any 'composition' properties on it. $\circ(\mathcal{C})$ denotes the set of connectives of circuits from $\mathcal{C}$. For example, for a class $\mathcal{C}$ of constant-depth circuits, $\circ(\mathcal{C})$ can be the set of logical connectives of unbounded arity which is, formally, defined as the set of connectives $\neg, \vee_m, \wedge_m$, for each possible arity $m \geq 2$. We say that circuits from $\mathcal{C}$ are 'Boolean circuits over $\circ(\mathcal{C})$' and '$\mathcal{C}$-circuits'.

Circuit$[s]$ denotes the class of Boolean circuits over $\{\neg, \wedge_2, \vee_2\}$ of size at most s. Given $f \in \mathcal{F}_n$, $Size_{\mathcal{C}}(f)$ denotes the size of the smallest $\mathcal{C}$-circuit computing f, if one exists. If there is no such circuit, we let $Size_{\mathcal{C}}(f) := \infty$.

2.1 Approximation models

Intuitively, the approximation method formalizes the following strategy for proving circuit lower bounds: First, show that each s-size circuit C can be approximated by an s-size circuit $\overline{C}$ so that $\Pr_x[C(x) \neq \overline{C}(x)]$ is small. The approximating circuit $\overline{C}$ is obtained from C by replacing each gate of C by an approximating gate. Then show that each s-size approximating circuit $\overline{C}$ is far from f in the sense that $\Pr_x[\overline{C}(x) = f(x)]$ is small. Consequently, no s-size circuit C computes f either. We proceed with formal definitions.

---

[3]As the underlying acyclic graph of a circuit does not specify the order of inputs to a gate, for the computation of the circuit to be well-defined, we need the connectives to be symmetric. That is, $\circ(y_1, \ldots, y_m) = \circ(y_{\pi(1)}, \ldots, y_{\pi(m)})$ for each permutation π on $[m]$.

Definition 1 (Legitimate model). *A legitimate model (of order n) of a circuit class $\mathcal{C}$ is a pair $\langle \mathcal{M}, \overline{\circ}(\mathcal{M}) \rangle$, where $\{0, 1, x_i \mid i \geq 1\} \subseteq \mathcal{M} \subseteq \mathcal{F}_n$ and $\overline{\circ}(\mathcal{M})$ is a set of symmetric operations*

$$\overline{\circ} : \underbrace{\mathcal{M} \times \cdots \times \mathcal{M}}_{m \text{ times}} \mapsto \mathcal{M},$$

for each $\circ \in \circ(\mathcal{C})$ of arity m. We abuse the notation and denote $\langle \mathcal{M}, \overline{\circ}(\mathcal{M}) \rangle$ by $\mathcal{M}$.

Note that the 'approximator' $\overline{\circ}$ of $\circ \in \circ(\mathcal{C})$ depends on the input functions. This means that a single connective of a circuit can be approximated by different functions in different positions in the circuit.

Definition 2 (Approximating circuit). *Given a legitimate model (of order n) of a circuit class $\mathcal{C}$ and a $\mathcal{C}$-circuit C with inputs from $\{x_1, \ldots, x_n\}$, we define an approximating circuit $\overline{C}$ inductively by replacing each gate $\circ(C_1, \ldots, C_m)$ in C by $\overline{\circ}([[\overline{C_1}]], \ldots, [[\overline{C_m}]])$ and setting $\overline{1} := 1, \overline{0} := 0, \overline{x_i} := x_i$. Here, $[[B]] \in \mathcal{F}_n$ denotes the function computed by circuit (or approximating circuit) B. In order to simplify the notation, we will assume that B is $[[B]]$, if it is used in a context which asks for a function.*

Definition 3 (Error sets). *Given a legitimate model $\mathcal{M}$ of a circuit class $\mathcal{C}$, we define the error sets of $\mathcal{M}$ by*

$$\delta_\circ^+(f_1, \ldots, f_m) := \circ(f_1, \ldots, f_m) \setminus \overline{\circ}(f_1, \ldots, f_m),$$
$$\delta_\circ^-(f_1, \ldots, f_m) := \overline{\circ}(f_1, \ldots, f_m) \setminus \circ(f_1, \ldots, f_m),$$
$$\delta_\circ(f_1, \ldots, f_m) := \delta_\circ^+(f_1, \ldots, f_m) \cup \delta_\circ^-(f_1, \ldots, f_m),$$

where $\circ \in \circ(\mathcal{C})$ has arity m and $f_1, \ldots, f_m \in \mathcal{M}$. Further, we let

$$\Delta := \{\delta_\circ(f_1, \ldots, f_m) \mid f_1, \ldots, f_m \in \mathcal{M}, \circ \in \circ(\mathcal{C}) \text{ of arity } m\}.$$

Definition 4 (Distance). *The distance $\rho(f, \mathcal{M})$ of a legitimate model $\mathcal{M}$ of a circuit class $\mathcal{C}$ from a Boolean function $f \in \mathcal{F}_n$ is the minimal t such that there are tuples $\langle \overline{\circ_i}, f_1^i, \ldots, f_{m_i}^i \rangle$, where m_i is the arity of $\circ_i \in \circ(\mathcal{C})$, $i = 1, \ldots, t$ and $f_j^i \in \mathcal{M}$, such that for some $g \in \mathcal{M}$,*

$$f \oplus g \subseteq \bigcup_{i=1,\ldots,t} \delta_{\circ_i}(f_1^i, \ldots, f_{m_i}^i).$$

If there is no such t, we set $\rho(f, \mathcal{M}) := \infty$.[4]

---

[4]Razborov [18] defined the distance $\rho(f, \mathcal{M})$ as the minimal t such that

$$f \setminus g \subseteq \bigcup_{i=1,\ldots,t} \delta_{\circ_i}^+(f_1^i, \ldots, f_{m_i}^i) \text{ and } g \setminus f \subseteq \bigcup_{i=1,\ldots,t} \delta_{\circ_i}^-(f_1^i, \ldots, f_{m_i}^i).$$

We use a 'symmetric' definition of ρ, which simplifies proofs, but our results work for Razborov's 'asymmetric' ρ as well (with an appropriate adjustment of parameters). The asymmetry in Razborov's definition of ρ is used when defining the fusion method, see §4.

Proposition 1. *For each $f \in \mathcal{F}_n$ and each legitimate model $\mathcal{M}$ of a circuit class $\mathcal{C}$,*

$$\rho(f, \mathcal{M}) \leq Size_{\mathcal{C}}(f).$$

Proof. Let C be a $\mathcal{C}$-circuit computing f and $g := \overline{C} \in \mathcal{F}_n$. Consider the set of tuples $\langle \overline{\circ_i}, \overline{B_1}, \ldots, \overline{B_{m_i}} \rangle$ for all gates $\circ_i$ of C together with the subcircuits $B_1, \ldots, B_{m_i}$ which act as inputs of $\circ_i$. If $f(x) \neq g(x)$, we can compare computations of C and $\overline{C}$ on x. We start at the output gate of C and proceed to some input of the output gate which preserves the error, if such an input exists. We proceed in this way until we reach a gate $\circ$ of C such that $\circ(D_1(x), \ldots, D_m(x)) \neq \overline{\circ}(\overline{D_1}(x), \ldots, \overline{D_m}(x))$ but $D_j(x) = \overline{D_j}(x)$, for all $j \in [m]$, where $D_1, \ldots, D_m$ are subcircuits acting as inputs of $\circ$. Therefore, $x \in \delta_\circ(\overline{D_1}, \ldots, \overline{D_m})$. $\square$

Probabilistic approach. A particularly useful instantiation of the approximation method can be described as the following probabilistic approach. For a random variable $\boldsymbol{x}$ on $\{0,1\}^n$, let

$$d := \max \left\{ \Pr[\boldsymbol{x} \in \delta_\circ(f_1, \ldots, f_m)] \mid f_1, \ldots, f_m \in \mathcal{M}, \circ \in \circ(\mathcal{C}) \text{ of arity } m \right\}.$$

If $d \neq 0$, define

$$\rho(f, \mathcal{M}, \boldsymbol{x}) := \min_{g \in \mathcal{M}} \left\{ \frac{\Pr[f(\boldsymbol{x}) \neq g(\boldsymbol{x})]}{d} \right\}.$$

Assuming $\rho(f, \mathcal{M}) \leq t$ is witnessed by g and a set T of error sets with $|T| = t$, we have $\Pr[f(\boldsymbol{x}) \neq g(\boldsymbol{x})] \leq \sum_{\delta \in T} \Pr[\boldsymbol{x} \in \delta] \leq dt$. Therefore, $\rho(f, \mathcal{M}, \boldsymbol{x}) \leq \rho(f, \mathcal{M})$.

Examples. Lower bounds of Razborov-Smolensky for $\mathsf{AC}^0[p]$ and monotone circuit lower bounds of Razborov can be naturally formulated in the framework of the approximation method. In fact, they use the probabilistic approach described above.

Monotone circuit lower bounds: We refer to the presentation in Arora-Barak [2, Chapter 14.3]. The approximation model $\mathcal{M}$ consists of functions computed by small disjunctions of indicator functions. The operations $\overline{\vee}, \overline{\wedge}$ are defined so that probabilities $\Pr[\boldsymbol{x} \in \delta_\circ]$ are small, for a suitable $\boldsymbol{x}$. Nevertheless, for each $g \in \mathcal{M}$, $\Pr[f(\boldsymbol{x}) \neq g(\boldsymbol{x})]$ is high, if f is a suitable function, so $\rho(f, \mathcal{M}, \boldsymbol{x})$ yields the desired lower bound.

$\mathsf{AC}^0[p]$ *lower bounds:* We refer to the presentation in [11] or [2, Chapter 14.2]. The lower bound for $\mathsf{AC}^0[p]$ circuits of depth d is obtained by a lower bound for a version of $\rho(f, \mathcal{M}, \boldsymbol{x})$, where we consider the minimum only over functions $g \in \mathcal{M}$ such that for some circuit C of depth d, $g = \overline{C}$. The approximation model $\mathcal{M}$ consists of functions computed by polynomials over a finite field. The operations $\neg, MOD_p$ are approximated by polynomials which make no errors and the operations $\wedge_m, \vee_m$ (of unbounded arity) are approximated by polynomials which err only on a small number of inputs (The approximating polynomials actually depend on the position of the gate in the given circuit.).

On the other hand, functions g such that $g = \overline{C}$, for a circuit of depth d, are low-degree polynomials and the probability that a low-degree polynomial errs to compute MOD_q, for $q \neq p$, is high.

3 Limitations

3.1 General circuits

Razborov [18] showed inherent limitations of the approximation method, see also [12]. We present Razborov's proof and then adapt it to derive localizations of lower bounds based on the approximation method.

Theorem 2 (Razborov [18])**.** *For each legitimate model $\mathcal{M}$ of Boolean circuits over $\{\neg, \vee_2, \wedge_2\}$ such that $\overline{\neg} := \neg$, and for each $f \in \mathcal{F}_n$,*

$$\rho(f, \mathcal{M}) \leq O(n_0 n),$$

where n_0 is the number of essential inputs of f, i.e. inputs of f such that switching the value of the input affects the output of f on some assignment of the remaining inputs. Moreover, for every random variable $\boldsymbol{x}$ on $\{0,1\}^n$,

$$\rho(f, \mathcal{M}, \boldsymbol{x}) \leq 12(n_0 + 1).$$

Proof. We start with the 'moreover' part. For each legitimate model $\mathcal{M}$, a Boolean function f and a random variable $\boldsymbol{x}$ on $\{0,1\}^n$, we want to find a function $g \in \mathcal{M}$ such that $\Pr[g(\boldsymbol{x}) \neq f(\boldsymbol{x})] \leq 12(n_0 + 1)d$. We use bold font to denote random variables. The existence of such g follows from the following inequality,

$$\Pr[\boldsymbol{h}(\boldsymbol{x}) \neq f(\boldsymbol{x})] \leq 12(n_0 + 1) \Pr[\boldsymbol{x} \in \boldsymbol{\delta}] \leq 12(n_0 + 1)d.$$

Here, the second inequality holds by the definition of d. The first inequality is a direct corollary of Lemma 1, which also specifies the random variables $\boldsymbol{h}$ and $\boldsymbol{\delta}$.

Lemma 1. *For each legitimate model $\mathcal{M}$ of Boolean circuits over $\{\vee_2, \wedge_2, \neg\}$ such that $\overline{\neg} := \neg$, and for each $f \in \mathcal{F}_n$, there is a random variable $\boldsymbol{\delta}$ on Δ and a random variable $\boldsymbol{h}$ on $\mathcal{M}$ such that for each x,*

$$\Pr[\boldsymbol{h}(x) \neq f(x)] \leq 12(n_0 + 1) \Pr[x \in \boldsymbol{\delta}].$$

Before proving Lemma 1 we show how it implies the rest of the theorem. Given a legitimate model $\mathcal{M}$ and a Boolean function f, we want to find a function $g \in \mathcal{M}$ and $O(nn_0)$ tuples $\langle \overline{\circ_i}, f_1^i, \ldots, f_{m_i}^i \rangle$ which cover $f \oplus g$. Observe that there are two ways of 'upper-bounding' $\rho(f, \mathcal{M})$.

11

Case I (random approximators). Consider inputs x such that $\Pr[\boldsymbol{h}(x) \neq f(x)] < 1/3$. By (the multiplicative) Chernoff's bound,

$$\Pr[\mathsf{MAJ}(\boldsymbol{h}_1, \ldots, \boldsymbol{h}_k)(x) \neq f(x)] \leq 1/e^{k/48},$$

where each $\boldsymbol{h}_i$ is an independent copy of $\boldsymbol{h}$. Consequently, we can fix $h_1, \ldots, h_{O(n)}$ such that $\mathsf{MAJ}(h_1, \ldots, h_{O(n)})$ coincides with f on all such inputs x. As the majority function MAJ on m inputs is computable by an $O(m \log m)$-size circuit[5] M, we set $g := \overline{M}(h_1, \ldots, h_{O(n)})$ and conclude that all inputs $x \in g \oplus f$ considered in Case I are covered by $O(n \log n) \leq O(nn_0)$ tuples. Here, w.l.o.g. $n_0 \geq \log n$ since otherwise $\rho(f, \mathcal{M}) \leq 2^{n_0} \leq n$.

Case II (random error sets). For inputs x such that $\Pr[\boldsymbol{h}(x) \neq f(x)] \geq 1/3$, we use the fact that random δ covers such x with high probability. That is, by Lemma 1,

$$\Pr[x \in \boldsymbol{\delta}_1 \vee \cdots \vee x \in \boldsymbol{\delta}_{k'}] \geq 1 - 1/e^{k'/36(n_0+1)},$$

where each $\boldsymbol{\delta}_i$ is an independent copy of $\boldsymbol{\delta}$. Therefore, there are $\delta_1, \ldots, \delta_{O(nn_0)}$ covering all x's considered in Case II.

Cases I and II together imply the inequality $\rho(f, \mathcal{M}) \leq O(nn_0)$.

It remains to prove Lemma 1.

Given $h \in \mathcal{F}_m$, let $h^a := h(x_1, \ldots, x_{m-1}, a) \in \mathcal{F}_{m-1}$, for $a \in \{0, 1\}$. Further, let C_h be a trivial exponential-size circuit computing h defined inductively as

$$C_h := (C_{h^1} \wedge x_m) \vee (C_{h^0} \wedge \neg x_m),$$

with $C_{h(b)} := h(b)$, for $b \in \{0, 1\}^m$. Define the random variable $\boldsymbol{h} := \overline{D_{\boldsymbol{g}}}$, where

$$D_{\boldsymbol{g}} := (C_{f \oplus \boldsymbol{g}} \wedge C_{\neg \boldsymbol{g}}) \vee (C_{\neg(f \oplus \boldsymbol{g})} \wedge C_{\boldsymbol{g}}),$$

for uniformly random $\boldsymbol{g}$ on $\mathcal{F}_n$. Note that for each $g \in \mathcal{F}_n$, $D_g(x) = (f \oplus g \oplus g)(x) = f(x)$.

The random variable $\boldsymbol{\delta}$ is defined as a random error set of $\boldsymbol{h}$ as follows. Let $m \in [n] \cup \{\oplus\}, t \in \{0, 1, \vee\}$ and $\boldsymbol{g}_m \in \mathcal{F}_m$, with $\boldsymbol{g}_\oplus := \boldsymbol{g}_n \in \mathcal{F}_\oplus := \mathcal{F}_n$, be chosen independently and uniformly at random.

$$\boldsymbol{\delta} := \begin{cases} \delta_\vee(\overline{C_{f \oplus \boldsymbol{g}_m} \wedge C_{\neg \boldsymbol{g}_m}}, \overline{C_{\neg(f \oplus \boldsymbol{g}_m)} \wedge C_{\boldsymbol{g}_m}}) & \text{if } m = \oplus, t = \vee, \\ \delta_\wedge(\overline{C_{t \oplus f \oplus \boldsymbol{g}_m}}, \overline{C_{t \oplus \neg \boldsymbol{g}_m}}) & \text{if } m = \oplus, t \in \{0, 1\}, \\ \delta_\vee(\overline{C_{\boldsymbol{g}_m^1} \wedge x_m}, \overline{C_{\boldsymbol{g}_m^0} \wedge \neg x_m}) & \text{if } m \in [n], t = \vee, \\ \delta_\wedge(\overline{C_{\boldsymbol{g}_m^{1-t}}}, \underbrace{\overline{\neg x_m}}_{t \text{ times}}) & \text{if } m \in [n], t = \{0, 1\}. \end{cases}$$

The crucial observation is provided by Claim 3.1.

---

[5]Razborov uses monotone $O(m \log m)$-size circuits for MAJ, which is needed to make the proof work for his definition of ρ.

Claim 3.1. *For each x, there are $3 + 12n$ positions of gates in D_g (the underlying graph of D_g is independent of g) such that for each $g \in \mathcal{F}_n$ satisfying $\overline{D_g}(x) \neq f(x)$, we have $x \in \delta_\circ(\overline{e_1}, \overline{e_2})$, where $\circ$ is the gate of D_g corresponding to one of these $3 + 12n$ positions and e_1, e_2 are its inputs.*

To prove Claim 3.1, note that the error can occur either in one of the top 3 gates of D_g or in one of the remaining 4 subcircuits. Consider C_g, other circuits are treated analogously. Suppose x is a '+error' in the sense that $C_g(x) > \overline{C_g}(x)$. Then, there are gates e_1, e_2, e_3 in C_g such that $e_3 = e_1 \circ e_2$ and $\overline{e_3}$ makes a +error on x, but $\overline{e_1}, \overline{e_2}$ do not, i.e. $e_3(x) > \overline{e_3}(x)$, $e_1(x) \leq \overline{e_1}(x)$, $e_2(x) \leq \overline{e_2}(x)$.[6] Moreover, as one of the conjunctions $C_{g^0}(x) \wedge x_m$, $C_{g^0}(x) \wedge \neg x_m$ is false, a +error occurs in some of $2n$ gates on the path through C_g which is consistent with x. If a +error occurs in $\overline{e_3}$, but not in $\overline{e_1}, \overline{e_2}$, then $x \in \delta_\circ^+(\overline{e_1}, \overline{e_2})$ because

$$\overline{e_1}(x) \circ \overline{e_2}(x) \geq e_1(x) \circ e_2(x) = e_3(x) > \overline{e_3}(x) = \overline{e_1}(x)\overline{\circ}\,\overline{e_2}(x).$$

If x is a '–error' satisfying $C_g(x) < \overline{C_g}(x)$, we can again find gates e_1, e_2, e_3 in C_g such that $e_3 = e_1 \circ e_2$ and $\overline{e_3}$ makes a –error on x but $\overline{e_1}, \overline{e_2}$ do not. Analogously as before, this implies $x \in \delta_\circ^-(\overline{e_1}, \overline{e_2})$. If there is a -error on $C_h(x) \wedge \neg x_m$ which is inconsistent with x, then $\overline{\neg x_m}(x) = \neg x_m = 0$ and we conclude $x \in \delta_\circ^-(\overline{e_1}, \overline{e_2})$, for $e_1 = C_h$ and $e_2 = \neg x_m$, as

$$\overline{e_1}(x)\overline{\wedge}\overline{e_2}(x) = \overline{e_3}(x) > e_3(x) \geq \overline{e_1}(x) \wedge \overline{e_2}(x).$$

Therefore, a –error is covered by an error set corresponding to one of $3n$ positions in C_g: $2n$ positions on the path consistent with x and n positions deviating from the path. (The extra n results from the fact that both disjuncts of C_h can make a -error, but whenever a –error deviates from the path consistent with x it is covered immediately on the first gate on which it deviates.) This proves the claim.

Having Claim 3.1, we can finish the proof of Lemma 1. For each δ in the range of $\boldsymbol{\delta}$, let $T(\delta) := \langle m, t \rangle$, for $m \in [n] \cup \{\oplus\}, t \in \{0, 1, \vee\}$ given by δ, be the 'type' of δ. That is, there are $3(n + 1)$ possible types. Similarly, for fixed x, $g \in \mathcal{F}_n$, and $h = \overline{D_g}$ such that $h(x) \neq f(x)$, let $T(h) := T(\delta_\circ(\overline{e_1}, \overline{e_2}))$ for $\delta_\circ(\overline{e_1}, \overline{e_2})$ corresponding to the first of $3 + 12n$ positions from Claim 3.1 covering x - note that $\delta_\circ(\overline{e_1}, \overline{e_2})$ determines $m \in [n] \cup \{\oplus\}, t \in \{0, 1, \vee\}, g_m \in \mathcal{F}_m$ and the corresponding δ in the range of $\boldsymbol{\delta}$ such that $\delta_\circ(\overline{e_1}, \overline{e_2}) = \delta$. We want to show that

$$\Pr[x \in \boldsymbol{\delta}] \geq \sum_{k = \langle m, t \rangle} \frac{\Pr[x \in \boldsymbol{\delta} \mid T(\boldsymbol{\delta}) = k]}{3(n + 1)}$$

$$\geq \sum_{k = \langle m, t \rangle} \frac{\Pr[\boldsymbol{h}(x) \neq f(x) \wedge T(\boldsymbol{h}) = k]}{12(n + 1)} = \frac{\Pr[\boldsymbol{h}(x) \neq f(x)]}{12(n + 1)}.$$

---

[6]Gate e_i computes the function defined by the corresponding subcircuit of C_g.

It thus remains to prove that $\Pr[x \in \boldsymbol{\delta} \mid T(\boldsymbol{\delta}) = k] \geq \Pr[\boldsymbol{h}(x) \neq f(x) \wedge T(\boldsymbol{h}) = k]/4$. For m given by type k, we have $\Pr[x \in \boldsymbol{\delta} \mid T(\boldsymbol{\delta}) = k] = \Pr_{g_m \in \mathcal{F}_m}[x \in \delta^1]$, where δ^1 is the error set from the range of $\boldsymbol{\delta}$ determined by k and g_m. Further, $\Pr[\boldsymbol{h}(x) \neq f(x) \wedge T(\boldsymbol{h}) = k] = \Pr_{g_n \in \mathcal{F}_n}[\overline{D_{g_n}}(x) \neq f(x) \wedge T(\overline{D_{g_n}}) = k] \leq \Pr_{g_n \in \mathcal{F}_n}[\bigvee_i x \in \delta_i^2]$ where δ_i^2 corresponds to the ith position of type k from $3 + 12n$ positions given by Claim 3.1 (i.e. $1 \leq i \leq 4$). Finally, since all functions $f \oplus \boldsymbol{g}, \neg \boldsymbol{g}, \neg(f \oplus \boldsymbol{g}), \boldsymbol{g}$ are uniformly random, $\Pr_{g_n \in \mathcal{F}_n}[\bigvee_i x \in \delta_i^2] \leq 4 \Pr_{g_m \in \mathcal{F}_m}[x \in \delta^1]$. $\qquad \square$

'Approximation barrier' trick. A way to interpret the proof of Theorem 2 is that it is obtained by simulating the proof of Proposition 1, where instead of a small circuit for f, we want to use a trivial exponential-size circuit C_f. First, we observe that for each x such that $C_f(x) \neq f(x)$ there are $3n$ error sets on the path described in the proof of Claim 3.1 such that one of these $3n$ error sets covers x. We could switch the quantification and obtain $O(n^2)$ error sets covering all errors x, if we constructed a distribution on error sets such that for each x, $\Pr[x \in \boldsymbol{\delta}] \geq 1/n$. The problem is that for each x there are different $3n$ error sets covering x. To fix this, we 'xor' f with a random g and introduce circuit D_g (which we use instead of C_f). Now, for each x, the $3n$ error sets look the same on average. This leads to $\Pr[x \in \boldsymbol{\delta}] \geq \Pr[\boldsymbol{h}(x) \neq f(x)]/O(n)$ and to the final case analysis.

The barrier from Theorem 2 can be strengthened if we assume that approximating circuits $\overline{C}$ depend only on the essential variables of C. We refer to approximation models with this property as 0-projective models. (In Theorem 3 we generalize 0-projectivity to 'full' projectivity.)

Definition 5 (0-projective model). *Let $\mathcal{M}$ be a legitimate model of a circuit class $\mathcal{C}$. We say that $\mathcal{M}$ is 0-projective if the condition $*$ holds:*

* *If C is a $\mathcal{C}$-circuit with inputs x_j, for $j \in J \subseteq [n]$, then for each $z, y \in \{0,1\}^n$,*

$$z|_J = y|_J \quad \Rightarrow \quad \overline{C}(z) = \overline{C}(y).$$

Here, $z|_J \in \{0,1\}^{|J|}$ is a projection of z to bits z_j with $j \in J$.

Lower bounds of Razborov-Smolensky for $\mathsf{AC}^0[p]$ and monotone circuit lower bounds of Razborov are obtained by constructing 0-projective approximation models.

Corollary 1 (A limitation of the 0-projective approximation method). *Let $\mathcal{M}$ be any 0-projective legitimate model of Boolean circuits over $\{\neg, \wedge_2, \vee_2\}$ such that $\overline{\neg} := \neg$. Then, for each $f \in \mathcal{F}_n$ with n_0 essential variables,*

$$\rho(f, \mathcal{M}) \leq O(n_0^2).$$

Proof. We proceed as in the proof of Theorem 2 but note that for 0-projective approximation models, for each $x, y \in \{0,1\}^n$ and error set $\delta_\circ(\overline{C_1}, \overline{C_2})$, where C_1, C_2 have inputs from $\{x_j \mid j \in J \subseteq \{0,1\}^n\}$, we have that

$$x \in \delta_\circ(\overline{C_1}, \overline{C_2}) \wedge x|_J = y|_J \quad \Rightarrow \quad y \in \delta_\circ(\overline{C_1}, \overline{C_2}). \tag{3.1}$$

Further, by 0-projectivity of $\mathcal{M}$, each h in the range of $\boldsymbol{h}$ from Lemma 1 depends just on the essential variables of f. Hence, in Case I it suffices to cover projections of x on the essential variables, which can be done by $O(n_0 \log n_0)$ tuples. Similarly, each δ from the range of $\boldsymbol{\delta}$ in Lemma 1 is of the form $\delta_\circ(\overline{C_1}, \overline{C_2})$, where C_1, C_2 depend only on the essential variables of f. By (3.1), in Case II it thus again suffices to cover projections of x on the essential variables, which can be done by $O(n_0^2)$ tuples. $\square$

Theorem 2 and Corollary 1 leave open the possibility of proving a superpolynomial circuit lower bound by proving a superlinear lower bound using standard or 0-projective approximation method and combining it with hardness magnification. We rule out such attacks for localizable magnification theorems under additional assumptions: 1. There is only a sublinear number of oracles (in the oracle circuit implying the magnification theorem); 2. Assuming 0-projectivity, we allow a superlinear number of oracles, but they have to appear only at the bottom of the circuit; 3. Assuming a strengthening of 0-projectivity, we obtain the ideal localizabilty.

Theorem 3 (Localizability of the approximation method for general circuits)**.** *Let $\mathcal{M}$ be any legitimate model of Boolean circuits over $\{\neg, \wedge_2, \vee_2\}$ such that $\overline{\neg} := \neg$, and $f \in \mathcal{F}_n$.*

1. *[Localizability with few oracles] Suppose that $\rho(f, \mathcal{M}) \geq s$. Then, f is not computable by circuits of size $s - O(knm)$ using connectives $\neg, \wedge_2, \vee_2$ and k arbitrarily powerful, possibly different and asymmetric[7], oracles of arity m.[8]*

2. *[Localizability of 0-projective models] Suppose that $\rho(f, \mathcal{M}) \geq s$ and that $\mathcal{M}$ is 0-projective. Then, f is not computable by circuits of size $s - O(km^2)$ using connectives $\neg, \wedge_2, \vee_2$ and k arbitrarily powerful, possibly different and asymmetric, oracles of arity m whose inputs are among $\{x_1, \neg x_1, \ldots, x_n, \neg x_n\}$.*

---

[7]Allowing asymmetric oracle requires generalizing the notion of circuit so that its underlying graphs specifies the order of inputs.

[8]It is possible to prove Item 1 with Razborov's asymmetric definition of $\rho(f, \mathcal{M})$, if we adjust the resulting lower bound for the oracle circuits to $s - O(kn \cdot max\{m, \log n\})$ and conclude that it holds only for circuits with monotone oracles and negations at the bottom. The term $max\{m, \log n\}$ comes from not using DNFs in the proof, which is done in order to keep negations at the bottom. Further, it is possible to avoid the requirement on the monotonicity of oracles and negations at the bottom by using slice functions (and adjusting the parameters appropriately).

3. [Localizability of projective models] *Suppose that $\rho(f, \mathcal{M}) \geq s$ and that $\mathcal{M}$ is projective. Here, $\mathcal{M}$ is projective if for each Boolean circuit C over $\{\neg, \wedge_2, \vee_2\}$ with m inputs $x_1, \ldots, x_m$, for all $f_1, \ldots, f_m \in \mathcal{M}$ and each $z, y \in \{0, 1\}^n$,*

$$\bigwedge_{i=1,\ldots,m} f_i(z) = f_i(y) \quad \Rightarrow \quad \overline{C}(f_1(z), \ldots, f_m(z)) = \overline{C}(f_1(y), \ldots, f_m(y)).$$

Then, f is not computable by circuits of size $s - O(km^2)$ using connectives $\neg, \wedge_2, \vee_2$ and k arbitrarily powerful, possibly different and asymmetric, oracles of arity m.

While Theorem 3 is stated for general parameters, it is interesting only for $s < Kn^2$, where K is the constant implicit in Theorem 2. Typical parameters of our interest are: $s = n^{1.99}, k = n^{1.88}, m = poly(\log n)$. Item 1 of Theorem 3 is thus not satisfactory because it requires $k < Kn/m$ for a nontrivial conclusion. Item 2 works for the parameters of our interest but it requires oracles at the bottom (which is not the case for the HM frontier in the introduction and many other magnification theorems). By strengthening 0-projectivity to projective models, in Item 3 we can allow oracles to appear anywhere in the circuit. Razborov-Smolensky lower bound method is projective, but Razborov's lower bound for monotone circuits is not. However, for monotone circuit lower bounds already Theorem 2 does not apply.

Proof. We start with the proof of Item 1 and then adapt it to obtain Items 2 and 3.

Item 1: We prove the contrapositive. Assume f is computable by an s-size oracle circuit C^O and that the gates of C^O are ordered so that inputs of the ith gate e^i precede e^i. Following our earlier convention, e^i computes the function defined by the corresponding subcircuit of C^O. We will inductively associate each e^i with a function $f^i \in \mathcal{M}$ so that each x satisfying $e^j(x) \neq f^j(x)$, for some $j \leq i$, is covered by one of $i + Ko_imn$ error sets $\delta \in \Delta$, where o_i is the number of oracles among the first i gates of C^O and K is an absolute constant. In particular, we will have $\rho(e^i, \mathcal{M}) \leq i + Ko_imn$, $e^s = f$ and $o_s = k$.

Inputs x_j and constants 0,1 are associated with functions $x_j, 0, 1 \in \mathcal{M}$. Suppose we already associated the first i gates. If the $(i+1)$'st gate $\circ(e^{i_1}, e^{i_2})$, with $i_1, i_2 \leq i$, computes $\neg, \wedge_2$ or $\vee_2$, we set $f^{i+1} := \overline{\circ}(f^{i_1}, f^{i_2})$. In this case we use one error set $\delta_\circ(f^{i_1}, f^{i_2})$ to cover all x's such that $\overline{\circ}(f^{i_1}, f^{i_2})(x) \neq \circ(f^{i_1}, f^{i_2})(x)$. By the inductive hypothesis, all x's such that $f^j(x) \neq e^j(x)$, for some $j \leq i$, are covered by $i + Ko_imn$ error sets, where $o_i = o_{i+1}$. Hence, all x's such that $f^j(x) \neq e^j(x)$, for some $j \leq i+1$, are covered by $i+1+Ko_{i+1}mn$ error sets.

Assume that the $(i+1)$st gate $\circ(e^{i_1}, \ldots, e^{i_m})$ computes an oracle of arity m. Following the proof of Lemma 1 (with f^{i_j}'s instead of x_j's, with $\overline{f^{i_j}} := f^{i_j}$ and $f = \circ$), we obtain a random variable $\boldsymbol{\delta}$ on Δ and a random variable $\boldsymbol{h}$ on $\mathcal{M}$ such that for each x,

$$\Pr[\boldsymbol{h}(x) \neq \circ(f^{i_1}, \ldots, f^{i_m})(x)] \leq 12(m+1)\Pr[x \in \boldsymbol{\delta}].$$

16

Next, we simulate the case analysis from the proof of Theorem 2.

Case I: Consider x's satisfying $\Pr[\boldsymbol{h}(x) \neq \circ(f^{i_1}, \ldots, f^{i_m})(x)] < 1/3$. Let M be a $K_1 n \log n$-size circuit computing MAJ and $h_1, \ldots, h_{K_2 n}$ be such that on all x's from Case I, $M(h_1(x), \ldots, h_{K_2 n}(x)) = \circ(f^{i_1}, \ldots, f^{i_m})(x)$. Here, K_1, K_2 are absolute constants. That is, setting $f^{i+1} := \overline{M}(h_1, \ldots, h_{K_2 n})$, there are $K_1 n \log n$ error sets in Δ covering all x's from Case I such that $f^{i+1}(x) \neq \circ(f^{i_1}, \ldots, f^{i_m})(x)$. W.l.o.g. $\log n \leq m$, otherwise we can compute an oracle of arity m by a circuit of size $O(n)$ and use its approximator to define f^{i+1} such that all errors of f^{i+1} are covered by $O(n)$ error sets.

Case II: As in the proof of Theorem 2, the remaining x's not considered in Case I are covered by $K_3 mn$ error sets, for an absolute constant K_3.

Cases I and II together with the inductive hypothesis imply that all x's such that $f^j(x) \neq e^j(x)$, for some $j \leq i + 1$, are covered by $Ko_{i+1}mn + (i + 1)$ error sets. This finishes the proof of Item 1.

Item 2: We proceed as in the proof of Item 1 but assume additionally that oracles of C^O are at the bottom. In fact, w.l.o.g. each oracle has the form $\circ(x_{i_1}, \ldots, x_{i_m})$ as otherwise (if some x_j is $\neg x_j$) we can redefine $\circ$. When considering the $(i+1)$st oracle gate $\circ(e^{i_1}, \ldots, e^{i_m})$, by $*$-purity of $\mathcal{M}$, we conclude additionally that each h in the range of $\boldsymbol{h}$ depends only on $x_{i_1}, \ldots, x_{i_m}$. This allows us to use just $K_2 m$ functions h as inputs of M and $K_1 m \log m$ covering error sets in Case I. In Case II, just like in the proof of Corollary 1, we need only $O(m^2)$ error sets as well.

Item 3: Again, we follow the proof of Item 1 observing that for projective models, for each h in the range of $\boldsymbol{h}$, and each $x, y \in \{0, 1\}^n$,

$$\bigwedge_{j=1,\ldots,m} f^{i_j}(x) = f^{i_j}(y) \quad \Rightarrow \quad h(x) = h(y).$$

Therefore, in Cases I and II it suffices to cover projections of x on $f^{i_1}(x), \ldots, f^{i_m}(x)$, which can be done with $O(m^2)$ error sets. $\qquad\square$

Problem 1 (Full localizability of the approximation method for general circuits). *Does Item 3 of Theorem 3 hold without the assumption of projectivity?*

Models preserving the structure of circuits. Proposition 1 holds even for a 'structured' $\rho(f, \mathcal{M})$ defined so that the tuples $\langle \overline{\circ}_i, f_1^i, \ldots, f_{m_i}^i \rangle$ are required to form a circuit. That is, the structured $\rho(f, \mathcal{M})$, denoted $\rho'(f, \mathcal{M})$, is the minimal number of tuples covering $f \oplus g$, for some g, such that we can assign the tuples to nodes of a directed acyclic graph satisfying:

1. If nodes $\left\langle \overline{\circ}_{i_1}, f_1^{i_1}, \ldots, f_{m_{i_1}}^{i_1} \right\rangle, \ldots, \left\langle \overline{\circ}_{i_k}, f_1^{i_k}, \ldots, f_{m_{i_k}}^{i_k} \right\rangle$ are inputs of node $\left\langle \overline{\circ}_i, f_1^i, \ldots, f_k^i \right\rangle$, then $f_j^i = \overline{\circ}_{i_j}(f_1^{i_j}, \ldots, f_{m_{i_j}}^{i_j})$, for $j = 1, \ldots, k$;

17

2. For nodes $\left\langle \overline{\circ}_i, f_1^i, \ldots, f_{m_i}^i \right\rangle$ at the bottom, each f_j^i is $x_k \in \mathcal{M}$ for some k.

For $\rho'(f, \mathcal{M})$, the barrier from Theorem 2 fails. Is it exactly the insensitivity to the structure of circuits what prevents the approximation method from proving strong circuit lower bounds? That is, is there a legitimate model $\mathcal{M}$ such that $\rho'(f, \mathcal{M}) \leq s$ implies $f \in \mathsf{Circuit}[s^{O(1)}]$?

If we required that g in the definition of $\rho(f, \mathcal{M})$ is $\overline{C}$ for a circuit C of size s, the resulting method would become complete for s-size circuit lower bounds. This follows trivially by considering a model with empty error sets.

Formula lower bounds via approximation models. By imposing an additional structure of covering error sets in the definition of $\rho(f, \mathcal{M})$, it is possible to formulate versions of the approximation method which are, in principle, more suitable for proving formula lower bounds. One possibility is to require that the tuples from the definition of the structured $\rho(f, \mathcal{M})$ form not only a circuit but a formula. Is it possible to adapt Theorem 2 to such formula-versions of $\rho(f, \mathcal{M})$? Note that the upper bounds from Theorem 2 are witnessed by $g \in \mathcal{M}$ such that $g = \overline{C}$ for a formula C. In particular, circuits D_g from the proof of Lemma 1, which are used to define the covering error sets, are formulas. Obtaining such versions of Theorem 2 for formulas would imply that known superquadratic formula lower bounds based on the method of random restrictions are not formalizable by the corresponding versions of the approximation method.

3.2 Constant-depth circuits

NC^1 lower bounds can be approached also via the approximation method for constant-depth circuits. This is because p-size formulas are computable by depth $d + 2$ circuits of size $2^{n^{O(1/d)}}$. We use Theorem 2 to impose a limitation of this approach.

Definition 6 ($\rho_d(f, \mathcal{M})$). *The d-depth distance $\rho_d(f, \mathcal{M})$ of a legitimate model $\mathcal{M}$ of a circuit class $\mathcal{C}$ from a Boolean function $f \in \mathcal{F}_n$ is the minimal t such that there are tuples $\left\langle \overline{\circ}_i, f_1^i, \ldots, f_{m_i}^i \right\rangle$, where m_i is the arity of $\circ_i \in \circ(\mathcal{C})$, $i = 1, \ldots, t$ and $f_j^i \in \mathcal{M}$, such that for some $g \in \mathcal{M}$,*

$$f \oplus g \subseteq \bigcup_{i=1,\ldots,t} \delta_{\circ_i}(f_1^i, \ldots, f_{m_i}^i).$$

Moreover, for each f_j^i above there is a $\mathcal{C}$-circuit C of depth at most $d-1$ such that $f_j^i = \overline{C}$ and $g = \overline{D}$ for a $\mathcal{C}$-circuit D of depth at most d. If there is no such t, set $\rho_d(f, \mathcal{M}) := \infty$.

Proposition 1 can be adapted to the case of $\rho_d(f, \mathcal{M})$ showing that lower bounds for $\mathcal{C}$-circuits of depth d can be obtained via lower bounds for $\rho_d(f, \mathcal{M})$. Razborov-Smolensky lower bound for $\mathsf{AC}^0[p]$ can be formulated as a lower bound for $\rho_d(f, \mathcal{M})$. (Originally, Razborov [17] proved a lower bound for constant-depth circuits by measuring the distance

of the target function from a 'stratified' model. In this paper we keep the notion of model intact.)

Theorem 4. *For each legitimate model $\mathcal{M}$ of Boolean circuits over $\bigcup_{a\geq 2}\{\neg, \vee_a, \wedge_a\}$ such that $\overline{} := \neg$, for each $f \in \mathcal{F}_n$ and each $d \geq 8$,*

$$\rho_d(f, \mathcal{M}) \leq Knd(2^{\lceil n/\lfloor d/8 \rfloor \rceil} + 1) + 2^{Kn^{k/d}} + Kn^K,$$

where K, k are absolute constants (independent of d).

Proof. We follow the proof of Theorem 2 and start by adapting Lemma 1.

Lemma 2. *For each legitimate model $\mathcal{M}$ of Boolean circuits over $\bigcup_{a\geq 2}\{\vee_a, \wedge_a, \neg\}$ such that $\overline{} := \neg$, for each $f \in \mathcal{F}_n$ and each $d \geq 1$, there is a random variable $\boldsymbol{\delta}$ on Δ and a random variable $\boldsymbol{h}$ on $\mathcal{M}$ such that for each x,*

$$\Pr[\boldsymbol{h}(x) \neq f(x)] \leq 4(d+1)(2^{\lceil n/d \rceil} + 1)\Pr[x \in \boldsymbol{\delta}].$$

Moreover, for each h in the range of $\boldsymbol{h}$ there is a circuit D of depth $2d+3$ such that $h = \overline{D}$ and for each $\delta_\circ(f_1, \ldots, f_m)$ in the range of $\boldsymbol{\delta}$ there are circuits D_i of depth at most $2d+2$ such that $f_i = \overline{D_i}$.

Lemma 2 implies Theorem 4 by the following case analysis.

Case I: Consider inputs x such that $\Pr[\boldsymbol{h}(x) \neq f(x)] < 1/3$. Since MAJ on m inputs is computable by a $poly(m)$-size formula, it is also computable by a depth $d + 2$ circuit M of size $max\{2^{O(m^{k/d})}, poly(m)\}$, for an absolute constant k. Hence, we can set $g := \overline{M}(h_1, \ldots, h_{O(n)})$ for suitable $h_1, \ldots, h_{O(n)} \in \mathcal{M}$, so that each $x \in g \oplus f$ considered in Case I is covered by one of $2^{O(n^{k/d})} + poly(n)$ error sets. Note that g is computed by an approximator of a circuit of depth $3d + 5$. The error sets thus satisfy the 'moreover' requirement from the definition of ρ_{3d+5}.

Case II: Similarly as in the proof of Theorem 2, the remaining inputs x can be covered by $O((d+1)(2^{\lceil n/d \rceil} + 1)n)$ error sets.

Cases I and II imply that $\rho_{8d}(f, \mathcal{M}) \leq \rho_{3d+5}(f, \mathcal{M}) \leq O(d(2^{\lceil n/d \rceil} + 1)n) + poly(n) + 2^{O(n^{k/d})}$.

It remains to prove Lemma 2. The core change of the construction is the definition of circuits C_h, for $h \in \mathcal{F}_m$, which is given inductively by

$$C_h := \bigvee_{z \in \{0,1\}^u} \left(C_{h^z} \wedge \bigwedge_{i=1,\ldots,u} \overbrace{\neg}^{1-z_i \text{ times}} x_{m-u+i} \right),$$

where $h^z := h(x_1, \ldots, x_{m-u}, z_1, \ldots, z_u) \in \mathcal{F}_{m-u}$, $C_{h(b)} = h(b)$ for $b \in \{0,1\}^m$ and $u := min\{\lceil n/d \rceil, m\}$.

Consequently, $D_g := (C_{f \oplus g} \wedge C_{\neg g}) \vee (C_{\neg(f \oplus g)} \wedge C_g)$ is a circuit of depth $2d + 3$. Recall that $\boldsymbol{h} = \overline{D_{\boldsymbol{g}}}$.

The definition of $\boldsymbol{\delta}$ is modified analogously, following the structure of D_g. Set $v := \lceil n/d \rceil$ and $w := n \bmod \lceil n/d \rceil$ if $n \neq 0 \bmod \lceil n/d \rceil$, $w := \lceil n/d \rceil$ otherwise. Let $m \in [d] \cup \{\oplus\}, t \in \{0,1\}^r \cup \{\vee\}$, with $r := 1$ if $m = \oplus$, $r := v$ if $1 < m \leq d$, $r := w$ if $m = 1$, and $\boldsymbol{g}_{w+(m-1)v} \in \mathcal{F}_{w+(m-1)v}$, with $\boldsymbol{g}_{w+(\oplus-1)v} := \boldsymbol{g}_n \in \mathcal{F}_{w+(\oplus-1)v} := \mathcal{F}_n$, be chosen uniformly at random and independently up to the dependence which is described explicitly. Then,

$$
\boldsymbol{\delta} := \begin{cases}
\delta_\vee(\overline{C_{f \oplus \boldsymbol{g}_n} \wedge C_{\neg \boldsymbol{g}_n}}, \overline{C_{\neg(f \oplus \boldsymbol{g}_n)} \wedge C_{\boldsymbol{g}_n}}) & \text{if } m = \oplus; t = \vee; \\[4pt]
\delta_\wedge(\overline{C_{t \oplus f \oplus \boldsymbol{g}_n}}, \overline{C_{t \oplus \neg \boldsymbol{g}_n}}) & \text{if } m = \oplus; t \in \{0,1\}; \\[4pt]
\delta_\vee(\overline{C_{\boldsymbol{g}_{w+(m-1)v}{}^{1\ldots1}} \wedge \bigwedge_{i=1,\ldots,v} x_{w+(m-2)v+i}}, \ldots \\[2pt]
\quad \ldots, \overline{C_{\boldsymbol{g}_{w+(m-1)v}{}^{0\ldots0}} \wedge \bigwedge_{i=1,\ldots,v} \neg x_{w+(m-2)v+i}}) & \text{if } 1 < m \leq d; t = \vee; \\[4pt]
\delta_\wedge(\overline{C_{\boldsymbol{g}_{w+(m-1)v}{}^t}}, \overline{\underbrace{\neg}_{1-t_1 \text{ times}} x_{w+(m-2)v+1}}, \ldots \\[2pt]
\quad \ldots, \overline{\underbrace{\neg}_{1-t_v \text{ times}} x_{w+(m-1)v}}) & \text{if } 1 < m \leq d; t = \{0,1\}^v; \\[4pt]
\delta_\vee(\overline{C_{\boldsymbol{g}_w{}^{1\ldots1}} \wedge \bigwedge_{i=1,\ldots,w} x_i}, \ldots \\[2pt]
\quad \ldots, \overline{C_{\boldsymbol{g}_w{}^{0\ldots0}} \wedge \bigwedge_{i=1,\ldots,w} \neg x_i}) & \text{if } m = 1; t = \vee; \\[4pt]
\delta_\wedge(\overline{C_{\boldsymbol{g}_w{}^t}}, \overline{\underbrace{\neg}_{1-t_1 \text{ times}} x_1}, \ldots, \overline{\underbrace{\neg}_{1-t_w \text{ times}} x_w}) & \text{if } m = 1; t = \{0,1\}^w.
\end{cases}
$$

Now, for each x, there are $3 + 4(2^v + 1)d$ positions of gates in D_g such that for each $g \in \mathcal{F}_n$ satisfying $\overline{D_g}(x) \neq f(x)$ an error set corresponding to one of these positions covers x. This is because each x is consistent with exactly one disjunct on the path from the output of D_g to its inputs. The final analysis thus differs only in that instead of $3(n+1)$ we have $\leq (2^v + 1)(d+1)$ types. $\qquad\square$

Theorem 4 does not rule out the possibility of obtaining NC^1 lower bounds via the approximation method.

Problem 2 (NC^1 lower bounds via the approximation method). *Let $\mathcal{M}$ be a legitimate model as in Theorem 4 and $f \in \mathcal{F}_n$. Is $\rho_d(f, \mathcal{M}) \leq 2^{n^{O(1/d)}}$, for each sufficiently big d? Does the upper bound hold for all f computable by nondeterministic circuits of p-size?*

Note that a straightforward application of approximation models from the Razborov-Smolensky lower bound does not show that $\rho_d(f, \mathcal{M}) \leq 2^{n^{O(1/d)}}$ fails for some f: Since n-degree polynomials compute all functions $f \in \mathcal{F}_n$, to prove the existence of a hard function for l^d-degree polynomials (approximating depth d circuits), we need $l < n^{1/d}$. As individual error sets in the model consisting of l^d-degree polynomials might cover up

to $1/2^l$ of inputs, the existence of a function f such that $\rho_d(f, \mathcal{M}) > 2^l$ does not follow directly.

Even though Theorem 4 might not be optimal it can be used to derive a nontrivial localization of the approximation method for constant-depth circuits.

Theorem 5 (Localizability of the approximation method for constant-depth circuits). *Let $\mathcal{M}$ be any legitimate model $\mathcal{M}$ of Boolean circuits over $\bigcup_{a \geq 2}\{\neg, \wedge_a, \vee_a\}$ such that $\overline{} := \neg$, and $f \in \mathcal{F}_n$. Suppose that $\rho_{d(2d+6)}(f, \mathcal{M}) \geq s$, for $d \geq 1$.*

Then, f is not computable by depth d circuits of size $s - poly(kdn(2^{m/d} + 1))$ using connectives $\neg, \wedge_a, \vee_a$, for $a \geq 2$, and k arbitrarily powerful, possibly different and asymmetric, oracles of arity m.[9] The constants in the $poly(\cdot)$-notation are universal and, in particular, independent of d.

Theorem 5 does not need any projectivity assumption, but just like in Theorem 4 its conclusion might not be optimal. Note that showing that f is not computable by $poly(n2^m)$-size circuits of depth d implies that f is not computable by p-size circuits of depth $d/3$ with oracles of arity m because we can replace such oracles by DNFs of size $2^{O(m)}$. Theorem 5 shows that even $poly(n2^{m/d})$-size lower bounds localize. Is it possible to localize lower bounds of size $poly(n2^{m^{O(1/d)}})$? A positive answer to Problem 2 could resolve the question.

Proof. We proceed as in the proof of Item 1 in Theorem 3 with some modifications.

We will inductively associate each e^i with a function $f^i \in \mathcal{M}$ so that each x satisfying $e^j(x) \neq f^j(x)$, for some $j \leq i$, is covered by one of $i + Ko_id(2^{Km/d} + 1)n^K$ error sets $\delta \in \Delta$, for an absolute constant K. Additionally, we will ensure that for each f^i there is a circuit D_i of depth $d_i(2d+6)$ such that $f^i = \overline{D_i}$ and each x such that $e^j(x) \neq f^j(x)$, for some $j \leq i$, is covered by an error set whose input functions are approximators of circuits of depth $\leq d_j(2d+6) - 1$. Here, d_i is the depth of the subcircuit corresponding to e^i. In particular, we will have $\rho_{d_i(2d+6)}(e^i, \mathcal{M}) \leq i + Ko_id(2^{Km/d} + 1)n^K$.

The base case and the inductive step when the $(i+1)$st gate is not an oracle are straightforward.

Suppose that the $(i+1)$st gate $\circ(e^{i_1}, \ldots, e^{i_m})$ computes an oracle of arity m. Following the proof of Lemma 2 (with f^{i_j}'s instead of x_j's, with $\overline{f^{i_j}} := f^{i_j}$ and $f = \circ$), we obtain a random variable $\boldsymbol{\delta}$ on Δ and a random variable $\boldsymbol{h}$ on $\mathcal{M}$ such that for each x,

$$\Pr[\boldsymbol{h}(x) \neq \circ(f^{i_1}, \ldots, f^{i_m})(x)] \leq 4(d+1)(2^{\lceil m/d \rceil} + 1)\Pr[x \in \boldsymbol{\delta}].$$

---

[9]Theorem 5 holds also for Razborov's asymmetric definition of $\rho(f, \mathcal{M})$, if we conclude that the resulting lower bound holds only for circuits with monotone oracles and negations at the bottom. Further, the requirement on the monotonicity of oracles and negations at the bottom can be avoided by using slice functions and adjusting the parameters appropriately.

Moreover, for each h in the range of $\boldsymbol{h}$ there is a circuit D of depth $\leq 2d + 3 + (d_{i+1} - 1)(2d+6)$ such that $h = \overline{D}$ and for each $\delta_\circ(f_1, \ldots, f_t)$ in the range of $\boldsymbol{\delta}$ there are circuits D_ℓ of depth at most $2d + 2 + (d_{i+1} - 1)(2d+6)$ such that $f_\ell = \overline{D_\ell}$.

Next, we simulate the case analysis from the proof of Theorem 3.

In Case I we consider x's such that $\Pr[\boldsymbol{h}(x) \neq \circ(f^{i_1}, \ldots, f^{i_m})(x)] < 1/6$ and use approximate majority instead of MAJ. Ajtai [1] constructed (monotone) p-size circuits of depth 3 accepting each input which contains $> 3/4$ 1s and rejecting each input which contains $< 1/4$ 1s. This yields $d_{i+1}(2d+6)$ depth circuits and $K_1 n^{K_1}$ corresponding error sets covering all x's considered in Case I such that $f^{i+1}(x) \neq \circ(f^{i_1}, \ldots, f^{i_m})(x)$. Here, K_1 is an absolute constant.

In Case II we cover the remaining x's by $K_2 d(2^{\lceil m/d \rceil} + 1)n$ error sets, for an absolute constant K_2.

Cases I and II together with the inductive hypothesis imply that all x's such that $f^j(x) \neq e^j(x)$, for some $j \leq i+1$, are covered by $Ko_{i+1}d(2^{\lceil m/d \rceil} + 1)n^K + (i+1)$ error sets. Moreover, for each gate e^j, $j \leq (i+1)$, f^j is computed by an approximator of a circuit of depth $d_j(2d+6)$ and each x such that $e^j(x) \neq f^j(x)$, for some $j \leq i+1$, is covered by an error set whose inputs are approximators of circuits of depth $\leq d_j(2d+6) - 1$. $\qquad\square$

4 Completeness: Fusion

4.1 General circuits

Theorem 2 does not rule out the possibility of deriving superquadratic lower bounds via the approximation method by introducing inessential variables: We might be able to show that $f \notin \mathsf{Circuit}[s]$ by proving $s \leq \rho(f', \mathcal{M})$ for $f'(x_1, \ldots, x_N) := f(x_1, \ldots, x_n) \in \mathcal{F}_N$ and $N = s$.

Razborov [18] showed that this strategy can be adapted so that it becomes complete. We present Karchmer's [9] interpretation of Razborov's construction for which Wigderson [22] coined the term *fusion* method. (The reader might find Wigderson's survey [22] a helpful source of additional intuition.) In the second part of this section we discuss the feasibility of the fusion method.

Let $f \in \mathcal{F}_n$, $V := f^{-1}(1)$, $U := f^{-1}(0)$. We aim to prove $f \notin \mathsf{Circuit}[s]$ by showing that each s-size circuit C computing f yields a rejecting computation of C on an input from V.

The rejecting computions will be 'filtered' from more abstract computations as follows. Given $g \in \mathcal{F}_n$, define $[[g]] := g^{-1}(1) \cap U$. The sets $[[g]]$ can be seen as generalizations of values 0 and 1 with $\emptyset$ standing for 0 and U for 1. Note that $[[f]] = \emptyset$ and for all $g, h \in \mathcal{F}_n$,

$$[[g]] \cap [[h]] = [[g \wedge h]] \quad \text{and} \quad [[g]] \cup [[h]] = [[g \vee h]].$$

We say that $F \subseteq \mathcal{P}(U) := \{U' \mid U' \subseteq U\}$ is a *semi-filter*[10], if

1. (nontriviality) $F(\emptyset) = 0$, $F(U) = 1$;

2. (monotonicity) $A \subseteq B \quad \Rightarrow \quad F(A) \leq F(B)$.

A semi-filter F 'preserves' a pair of sets (A, B), $A, B \subseteq U$, if

$$F(A) = 1 \wedge F(B) = 1 \Rightarrow F(A \cap B) = 1.$$

Otherwise, we say that (A, B) covers F. Suppose that all negation gates in a circuit C computing f are at the bottom layer and that F is a semi-filter with a string $v(F) \in \{0, 1\}^n$ such that

$$v(F)_i = F([[x_i]]) \neq F([[\neg x_i]]), \text{ for } i \in [n].$$

If F preserves pairs $([[g]], [[h]])$, for all gates g, h of C, then F implicitly defines a rejecting computation of C on input $v(F)$.

Let F_0 be an arbitrary set of semi-filters $F \subseteq \mathcal{P}(U)$ with $v(F) \in V$. Define $\rho_{F_0}(f)$ as the minimal number of pairs (A, B), $A, B \subseteq U$ which cover all semi-filters $F \in F_0$. It follows directly from the definition that

$$\rho_{F_0}(f) \leq 2 \cdot Size_C(f),$$

where $\mathcal{C}$ is the set of all Boolean circuits over $\{\neg, \wedge_2, \vee_2\}$. The factor of 2 results from pushing negations to the bottom.

The term 'fusion' refers to the process of 'fusing' rejecting computations of C on U into a rejecting computation of C on an input from V.

Fusion as the approximation method. The fusion method is an instantiation of the approximation method, if we relax the definition of a legitimate model. To see that, we use Razborov's definition of the distance $\rho(f', \mathcal{M})$ where $f' \oplus g \subseteq \bigcup_i \delta_{\circ_i}(f_1^i, \ldots, f_{m_i}^i)$ is replaced by

$$f' \backslash g \subseteq \bigcup_i \delta_{\circ_i}^+(f_1^i, \ldots, f_{m_i}^i) \quad \text{and} \quad g \backslash f' \subseteq \bigcup_i \delta_{\circ_i}^-(f_1^i, \ldots, f_{m_i}^i).$$

Assume that $|V| \geq 2$ and let F_0 be a set of semi-filters $F \subseteq \mathcal{P}(U)$ with $v(F) \in V$ such that there are $F_1, F_2 \in F_0$ with $v(F_1) \neq v(F_2)$. We want to define a legitimate model $\mathcal{M}$ of order $N := n + \lceil \log |F_0| \rceil$. $\mathcal{M}$ will consist of functions $\overline{g} \in \mathcal{F}_N$, for all $g \in \mathcal{F}_n$, given by

$$\overline{g}(x, y) := \begin{cases} g(x) & \text{if } v(y) \neq x, \\ y([[g]]) & \text{if } v(y) = x; \end{cases}$$

---

[10]In order theory, a filter on $\mathcal{P}(U)$ would satisfy also that $F(A) = 1 \wedge F(B) = 1$ implies $F(C) = 1$ for some $C \subseteq A \cap B$.

where $x \in \{0,1\}^n$ and $y \in \{0,1\}^{N-n}$ is identified with the yth semi-filter from F_0. In particular, $\{0, 1, x_i, \neg x_i \mid i \in [n]\} \subseteq \mathcal{M}$. We relax on the requirement that $x_i \in \mathcal{M}$, for $i > n$. The connectives of $\mathcal{M}$ are

$$\overline{\circ}(\overline{g}, \overline{h}) := \overline{g \circ h},$$

for $\circ \in \{\vee_2, \wedge_2\}$. As F_0 contains F_1, F_2 with $v(F_1) \neq v(F_2)$, we have $\overline{g} = \overline{h} \Rightarrow g = h$ and $\overline{\circ}$ is well-defined. Note that $\mathcal{M}$ depends on f.

We want to show that $\rho(f', \mathcal{M}) = \rho_{F_0}(f)$. The function $g \in \mathcal{M}$ from the definition of $\rho(f', \mathcal{M})$ has to be $\overline{f}$ because error sets $\delta_\circ$ do not contain any input (x, y) with $v(y) \neq x$. Further, by definition, $\overline{f}(x, y) \neq f'(x, y) \Leftrightarrow v(y) = x$ and $v(y) = x \Rightarrow \overline{f}(x, y) = 0$. Therefore, $\rho(f', \mathcal{M})$ is the minimal number of pairs g, h such that pairs $([[g]], [[h]])$ cover all semi-filters in F_0, i.e. $\rho(f', \mathcal{M}) = \rho_{F_0}(f)$.

Genesis of fusion. The fusion method arises from the approximation method by a sequence of natural choices. We want to approximate each $g \in \mathcal{F}_n$ by a function $\overline{g} \in \mathcal{F}_N$. Given the correspondence we look for a model with simple approximating connectives $\overline{\circ}(\overline{g}, \overline{h}) = \overline{g \circ h}$. For this to be well-defined we want $\overline{g} = \overline{h} \Rightarrow g = h$, which we guarantee by setting

$$\overline{g}(x, y) := g(x) \quad \text{if } P(x, y)$$

for some predicate P. If $\neg P(x, y)$, we set $\overline{g}(x, y) := F_y(g)$ for a functional $F_y : 2^{2^n} \mapsto \{0, 1\}$. F_y should guarantee that $\overline{x_i} = x_i$ and $\overline{\neg x_i} = \neg x_i$ so $\neg P(x, y)$ is "$F_y(x_i)$'s define x". By imposing the monotonicity of F_y we make $\delta_\vee^+$ empty. Similarly, setting $F_y(f) := 0$ simplifies the situation as it implies that $\overline{f} \backslash f' = \emptyset$ and restricts our attention to U.

Theorem 6 (Completeness of the fusion method [18]). *Let $f \in \mathcal{F}_n$ and assume that $f \notin \mathsf{Circuit}[Ks^3]$, for $s \geq n$. Then $\rho_{F_\forall}(f) \geq s$, where $F_\forall$ is the set of all semi-filters $F \subseteq \mathcal{P}(U)$ with $v(F) \in V$. Moreover, there is $F_0 \subseteq F_\forall$ of size $|F_0| = 2^{O(s|U|)}$ such that $\rho_{F_0}(f) \geq s$. Here, K is an absolute constant.*

Proof. Assuming $\rho_{F_\forall}(f) \leq s$, we want to construct an $O(s^3)$-size circuit computing f. We have s pairs (A_i, B_i), $A_i, B_i \subseteq U, i \in [s]$ covering all semi-filters in $F_\forall$. Let F_z, for $z \in \{0,1\}^n$, be the minimal subset of $\mathcal{P}(U)$ closed on the supersets, preserving all pairs (A_i, B_i), for $i \in [s]$, pairs $([[x_i]], [[\neg x_i]])$, for $i \in [n]$, and such that for all $i \in [n]$,

$$[[\underbrace{\neg}_{1-z_i \text{ times}} x_i]] \in F_z.$$

We claim that

$$f(z) = 1 \quad \Leftrightarrow \quad \emptyset \in F_z.$$

If $f(z) = 0$, then $Z := \{U' \mid z \in U' \subseteq U\}$ witnesses that $\emptyset \notin F_z$. If $f(z) = 1$, then $\emptyset \in F_z$ as otherwise $v(F_z) \in V$ and F_z would be a semi-filter preserving all pairs $(A_i, B_i), i \in [s]$.

It remains to construct an $O(s^3)$-size circuit deciding $\emptyset \in^? F_z$.

Let
$$A := \bigcup_{i \in [s]} \{A_i, B_i, A_i \cap B_i\} \cup \bigcup_{i \in [n]} \{[[x_i]], [[\neg x_i]]\} \cup \{\emptyset\}.$$

For each $B \in A$ and $k \geq 0$, define a function w_B^k inductively by

$$w_B^0 := \begin{cases} 1 & \text{if for some } i \in [n], B = [[x_i]] \wedge z_i = 1 \text{ or } B = [[\neg x_i]] \wedge z_i = 0, \\ 0 & \text{otherwise} \end{cases}$$

$$w_B^{k+1} := \bigvee_{C \subseteq B, C \in A} w_C^k \vee \bigvee_{j \in \{j | A_j \cap B_j = B\}} (w_{A_j}^k \wedge w_{B_j}^k) \vee \bigvee_{i \in [n]} (w_{[[x_i]]}^k \wedge w_{[[\neg x_i]]}^k).$$

Then, $\emptyset \in F_z \Leftrightarrow w_\emptyset^{|A|} = 1$ and the definition of $w_\emptyset^{|A|}$ yields a circuit of size $O(s^3)$ computing f.

The 'moreover' part follows by noting that it suffices to consider 2^n semi-filters $F \in F_\forall$ for each possible set of s pairs $(A_i, B_i), i \in [s]$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Feasibility of fusion. In the approximation method, error sets $\delta_\circ$ consist of inputs of the target function. In the fusion method, the error sets $\delta_\circ$ consist of semi-filters. By Theorem 2, the number of semi-filters cannot be reduced below $2^{\Omega(s)}$, if the fusion method is to prove an s-size lower bound.

Razborov [19] came up with an elegant generalization of the approximation method where the error sets are subsets of a set S which is unrelated to the inputs of the target function and has size just $|S| = O(s^3 \log^2 s)$. A drawback of his construction is that it is nonconstructive and not known to be complete - it proves s-size lower bounds for *almost all* functions, for suitable s.

The motivation for the construction from [19] was the question of finding a version of the approximation method capable of proving strong lower bounds which would be suitable for formalizations in bounded arithmetic S_2^1. Informally, S_2^1 is a theory of polynomial-time reasoning and does not allow to operate with objects which cannot be described by bit-strings of polynomial-size, cf. [4]. The polynomial-size is considered w.r.t. an initial parameter which is typically $s(n)$, where n is the number of inputs of the target function and $s(n)$ is the size of the lower bound we aim to prove, but it is possible to consider also, say, 2^n as the initial parameter, if the lower bound is formulated w.r.t. the truth-table of the target function. The latter formulation corresponds to the setting in the natural proofs and we refer to it as the 'truth-table' formalization, see [11].

If we want to formalize an s-size lower bound for $f \in \mathcal{F}_n$ in S_2^1 using the truth-table formalization, it suffices to show that for each set of s pairs $(A, B), A, B \subseteq U$, there is a suitable semi-filter preserving the set. In the truth-table formalization, S_2^1 can operate only with objects described by $2^{O(n)}$ bits. Nevertheless, semi-filters from F_0 in Theorem

6 can be chosen so that they are predicates computable by circuits of size $2^{O(n)}$. That is, the size of notions appearing in the fusion method does not present an obstacle for proving circuit lower bounds in S^1_2 using the truth-table formalization.

With the more succinct formalization of lower bounds in S^1_2, where the initial parameter is $s(n)$ instead of 2^n, each semi-filter needs to be described by $poly(s)$ bits. This leads to the following adaptation of the fusion method, which we refer to as the *antichecker fusion*.

Lipton and Young [10, Theorem 6] proved that for each sufficiently big n and each function $f \in \mathcal{F}_n$ such that $f \notin \mathsf{Circuit}[s^7]$, $s \geq n$, there is a set $S \subseteq \{0,1\}^n$ of size $|S| = poly(s)$ such that no s^3-size circuit computes f on S. The set S is the set of *anticheckers* of f. Consider the fusion method modified so that instead of sets U, V we use $U \cap S, V \cap S$. If we denote by $\rho^S_{F_0}$ the resulting version of ρ_{F_0}, we have $\rho^S_{F_0}(f_n) \leq 2 \cdot Size_{\mathcal{C}}(f)$, for each set F_0 of semi-filters $F \subseteq \mathcal{P}(U \cap S)$ with $v(F) \in V \cap S$. Moreover, by the proof of Theorem 6, for some F_0 with $|F_0| = 2^{poly(s)}$, $f \notin \mathsf{Circuit}[Ks^7]$ implies $\rho^S_{F_0}(f) \geq s$. This proves the following.

Theorem 7 (Completeness of the antichecker fusion). *Let $f \in \mathcal{F}_n$, for a sufficiently big n, and assume that $f \notin \mathsf{Circuit}[Ks^7]$, for $K^{1/7}s \geq n$. Then there is a set F_0 of semi-filters $F \subseteq \mathcal{P}(U \cap S)$ with $v(F) \in V \cap S$ and a set S of anticheckers of f such that $|F_0| = 2^{poly(s)}$ and $\rho^S_{F_0}(f) \geq s$. Here, K is an absolute constant.*

The approximation model $\mathcal{M}$ corresponding to the antichecker fusion from Theorem 7 uses $N = poly(s)$ and semi-filters from F_0 are predicates computable by $poly(s)$-size circuits. We can consider a more efficient approximation model $\mathcal{M}_S$ consisting of functions $\bar{g}(x, y) \in \mathcal{M}$ restricted to S on x's. That is, $\mathcal{M}_S$ consists of partial Boolean functions - the approximation method straightforwardly generalizes to this setting. Functions $\bar{g} \in \mathcal{M}_S$ are computable by $poly(s)$-size circuits. Further, for each $f_1, f_2 \in \mathcal{M}_S$, the predicate $(x, y) \in^? \delta^+_\wedge(f_1, f_2)$ is computable by a $poly(s)$-size circuit. Therefore, if we ignore the question of generating anticheckers efficiently and consider only the size of the notions involved, S^1_2 is perfectly capable of proving circuit lower bounds by estimating $\rho(f', \mathcal{M}_S)$.

A nonconstructive element of the model $\mathcal{M}_S$ is that we do not have a p-time algorithm which would output the set of anticheckers given a function $f \in \mathcal{F}_n$. For $f = \mathsf{SAT}$, the set of anticheckers w.r.t. $s = poly(n)$ can be generated in p-time given 1^n under the assumption of the existence of a one-way function secure against nonuniform p-size circuits and a function in E hard for subexponential-size circuits [11].

The complication with generating anticheckers disappears if we interpret the antichecker fusion as a *fusion method for partial Boolean functions*: Given $f : S \mapsto \{0,1\}$, $S \subseteq \{0,1\}^n$, which is not computable by any Ks^7-size circuit, we have $s \leq \rho^S_{F_0}(f)$ and $\rho^S_{F_0}(f)/2$ is a lower bound on the size of the smallest circuit computing f.

4.2 Constant-depth circuits

We now adapt the fusion method to the case of constant-depth circuit lower bounds. To make this work we break the monotonicity of semi-filters.

Let $f \in \mathcal{F}_n, V := f^{-1}(1), U := f^{-1}(0)$. We say that $F = (F^0, \ldots, F^d)$, where $F^0 \subseteq \cdots \subseteq F^d \subseteq \mathcal{P}(U)$ is a d-semifilter with $v(F) \in V$, if $F^d(\emptyset) = 0, F^0(U) = 1$ and

1. (initial sets) $v(F)_i = F^0([[x_i]]) \neq F^0([[\neg x_i]])$, for $i \in [n]$;

2. (d-monotonicity) For $1 \leq k \leq d$, $F^{k-1}(B) = 1 \wedge B \subseteq A \Rightarrow F^k(A) = 1$.

A d-semifilter F k-preserves a tuple $(A_1, \ldots, A_t), A_i \subseteq U$, for $1 \leq k \leq d$, if

$$\bigwedge_{i=1,\ldots,t} F^{k-1}(A_i) = 1 \Rightarrow F^k(\bigcap_{i=1,\ldots,t} A_i) = 1.$$

Otherwise, we say that $(A_1, \ldots, A_t)$ k-covers F.

Suppose that C is a $(d+1)$-depth circuit over $\bigcup_{a=2,\ldots,t}\{\neg, \wedge_a, \vee_a\}$ with negation gates at the bottom layer. Further, assume that C computes f and that F is a d-semifilter with $v(F) \in V$. If for each $1 \leq k \leq d$, F k-preserves tuples $([[g_1]], \ldots, [[g_t]])$, for all gates $g_1, \ldots, g_t$ of C computable by k-depth subcircuits of C, then F implicitly defines a rejecting computation of C on input $v(F)$.

Let F_0 be a set of $(d-1)$-semifilters F with $v(F) \in V$, for $d \geq 2$. Define $\rho_{F_0,d,t}(f)$ as the minimal number of tuples $(A_1, \ldots, A_t), A_i \subseteq U$, such that for each $F \in F_0$, there is $1 \leq k < d$ such that one of the tuples k-covers F. If such tuples do not exist, we set $\rho_{F_0,d,t}(f) := \infty$. Then, $\rho_{F_0,d,t}(f)/2$ is a lower bound on the size of a smallest d-depth circuit over $\bigcup_{a=2,\ldots,t}\{\neg, \wedge_a, \vee_a\}$ computing f.

Theorem 8 (Completeness of fusion for constant-depth circuit lower bounds). *Let $f \in \mathcal{F}_n$ and assume that no $(2d+1)$-depth $Kd(st+n)^2$-size circuit over $\bigcup_{a=2,\ldots,s(t+2)+2n+2}\{\neg, \wedge_a, \vee_a\}$, for $d, n, t \geq 2$, computes f. Then, $\rho_{F_\forall,d,t}(f) \geq s$, where $F_\forall$ is the set of all $(d-1)$-semifilters F with $v(F) \in V$. Here, K is an absolute constant.*

Proof. We proceed as in the proof of Theorem 6. Assuming $\rho_{F_\forall,d,t}(f) \leq s$, we want to construct an $O(d(st+n)^2)$-size $(2d+1)$-depth circuit computing f. We have s tuples $(A_1^i, \ldots, A_t^i), i \in [s]$, covering all $F \in F_\forall$.

Define $F_z = (F_z^0, \ldots, F_z^{d-1})$, where $F_z^0 \subseteq \cdots \subseteq F_z^{d-1} \subseteq \mathcal{P}(U)$, for $z \in \{0,1\}^n$ as follows. F_z^0 consists of the set U and sets $[[\underbrace{\neg}_{1-z_i \text{ times}} x_i]]$, for all $i \in [n]$. For $k \geq 1$, define $F_z^{k-1} \subseteq F_z^k \subseteq \mathcal{P}(U)$ by extending F_z^{k-1} by all supersets of sets in F_z^{k-1} and by 'k-preserving' tuples $(A_1^i, \ldots, A_t^i)$, for $i \in [s]$, consisting of sets from F_z^{k-1}.

We claim that $f(z) = 0 \Leftrightarrow \emptyset \notin F_z^{d-1}$. If $f(z) = 0$, then z is included in all sets in F_z^{d-1} so $\emptyset \notin F_z^{d-1}$. If $f(z) = 1$, then $\emptyset \in F_z^{d-1}$ as otherwise $v(F_z) \in V$ and F_z would be a $(d-1)$-semifilter k-preserving all tuples $(A_1^i, \ldots, A_t^i), i \in [s]$, for all $1 \leq k < d$.

It remains to construct a circuit deciding $\emptyset \in^? F_z^{d-1}$. Let

$$A := \bigcup_{i \in [s]} \{A_1^i, \ldots, A_t^i, \bigcap_{j=1,\ldots,t} A_j^i\} \cup \bigcup_{i \in [n]} \{[[x_i]], [[\neg x_i]]\} \cup \{\emptyset, U\}.$$

For each $B \in A$ and $k \geq 0$, define a function w_B^k inductively by

$$w_B^0 := \begin{cases} 1 & \text{if } \exists i \in [n], B = [[x_i]] \wedge z_i = 1 \text{ or } B = [[\neg x_i]] \wedge z_i = 0 \text{ or } B = U, \\ 0 & \text{otherwise} \end{cases}$$

$$w_B^{k+1} := \bigvee_{C \subseteq B, C \in A} w_C^k \vee \bigvee_{i \in \{i | B = \bigcap_{j=1,\ldots,t} A_j^i\}} \left(\bigwedge_{j=1,\ldots,t} w_{A_j^i}^k \right).$$

By definition, $B \in F_z^k \Leftrightarrow w_B^k = 1$, for $k \in \{0, 1\}$ and $B \in A$. For each $C \in F_z^k$, $1 \leq k < d$, with $C \subseteq B \in A$, C was added to F_z^k either as a superset of some $D \in F_z^{k-1}$, in which case $B \in F_z^k \cap A$, or C was added to F_z^k in order to preserve a tuple, in which case $C \in A$. This allows us to prove by induction that the previous equivalence holds for $0 \leq k < d$. In particular, $\emptyset \in F_z^{d-1} \Leftrightarrow w_\emptyset^{d-1} = 1$. Finally, by the construction of w_B^{d-1}, there are $(2d + 1)$-depth $O(d(st + n)^2)$-size circuits over $\bigcup_{a=2,\ldots,s(t+2)+2(n+1)} \{\neg, \vee_a, \wedge_a\}$ computing w_B^{d-1}. $\square$

Random restrictions as the approximation method. By Theorem 8, essentially every lower bound for constant-depth circuits can be turned into a lower bound obtained by the fusion method for constant-depth circuits. In particular, this holds for lower bounds based on the method of random restrictions.

The method of random restrictions proceeds by showing that each small circuit can be trivialized by a partial restriction, which implicitly defines an incorrect computation of the circuit w.r.t. a suitable target function. This is similar to the construction of semi-filters in the fusion method. Moreover, the number of partial restrictions needed to trivialize all $poly(s)$-size constant-depth circuits is just $poly(s)$. This suggests that it might be possible to formalize random restrictions in the framework of the approximation method with just $O(\log s)$ inessential variables. Such formalization would allow us to conclude that random restrictions are subject to the localizability from Theorem 5 (with the asymmetric definition of ρ_d). However, at this point we have not even managed to construct legitimate approximation models corresponding to the fusion method for constant-depth circuits.

It would be interesting to formulate random restrictions also in terms of approximation models with $poly(s)$ inessential variables with an addition property similar to 0-projectivity. Such models could be localizable similarly as in Theorem 5 as well.

Acknowledgement

References

[1] Ajtai M.; Σ_1^1-formulae on finite structures; *Annals of Pure and Applied Logic*, 24(1):1-48, 1983.

[2] Arora S., Barak B.; Computational Complexity: A Modern Approach; *Cambridge University Press*, 2009.

[3] Barrington D., Straubing H.; Superlinear lower bounds for bounded-width branching programs; *Journal of Computer and System Sciences*, 50(3):374-381, 1995.

[4] Buss S.; Bounded arithmetic; *Bibliopolis*, 1986.

[5] Carmosino M., Impagliazzo R., Kabanets V., Kolokolova A.; Learning algorithms from natural proofs; *Proc. of the 31st Computational Complexity Conference*, 1-24, 2016.

[6] Chen L., Hirahara S., Oliveira I.C., Pich J., Rajgopal N., Santhanam R.; Beyond natural proofs: hardness magnification and locality; *Journal of the ACM*, 69(4):1-49, 2022.

[7] Fan Z., Li J., Yang T.; The exact complexity of pseudorandom functions and the black-box natural proof barrier for bootstrapping results in computational complexity; *Proc. of the 54th Symposium on Theory of Computing*, 962-975, 2022.

[8] Hrubeš P., Rao A.; Circuits with medium fan-in; *Proc. of the 30th Computational Complexity Conference*, 381-391, 2015.

[9] Karchmer M.; On proving lower bounds for circuit size; *Proc. of the 8th Structure in Complexity Theory Conference*, 112-119, 1993.

[10] Lipton R.J., Young N.E.; Simple strategies for large zero-sum games with applications to complexity theory; *Proc. of the 26th Symposium on Theory of Computing*, 734-740, 1994.

[11] Müller M., Pich J.; Feasibly constructive proofs of succinct weak circuit lower bounds; *Annals of Pure and Applied Logic*, 171(2), 2020.

[12] Oliveira I.C.; Notes on the method of approximations and the emergence of the fusion method; *Notes available on author's website*, 2018.

[13] Oliveira I.C., Santhanam R.; Hardness magnification for natural problems; *Proc. of the 59th Symposium on Foundations of Computer Science*, 65-76, 2018.

[14] Pich J.; Learning algorithms from circuit lower bounds; *arXiv*, 2020.

[15] Razborov A.; Lower bounds for the monotone complexity of some Boolean functions; *Dokl. Math.*, 31:354-357, 1985.

[16] Razborov A.; Lower bounds on the monotone complexity of the logical permanent; *Mathematical notes of the Academy of Sciences of the USSR*, 37:485-493, 1985.

[17] Razborov A.; Lower bounds on the size of bounded depth networks over a complete basis with logical addition; *Mathematical notes of the Academy of Sciences of the USSR*, 41:333-338, 1987.

[18] Razborov A.; On the method of approximations; *Proc. of the 21st Symposium on Theory of Computing*, 167-176, 1989.

[19] Razborov A.; On small size approximation models; *The Mathematics of Paul Erdős I, Algorithms and Combinatorics*, 13, 1996.

[20] Razborov A., Rudich S.; Natural Proofs; *Journal of Computer and System Sciences*, 55(1):24-35, 1997.

[21] Smolensky R.; Algebraic methods in the theory of lower bounds for Boolean circuit complexity; *Proc. of the 19th Symposium on Theory of Computing*, 77-82, 1987.

[22] Wigderson A.; The fusion method for lower bounds in circuit complexity; *Combinatorics: Paul Erdős is Eighty*, 1993.

[23] Yao A.; Circuits and local computation; *Proc. of the 21st Symposium on Theory of Computing*, 186-196, 1989.