NOTES ON THE METHOD OF APPROXIMATIONS AND THE EMERGENCE OF THE FUSION METHOD

IGOR CARBONI OLIVEIRA

Abstract. This text is mostly a technical exposition of some results from [Raz89] on the power and limitations of the method of approximations. For clarity, we adopt a clear distinction between legitimate models in the "pure" sense, and models with auxiliary variables. We refer to the use of the latter models in circuit lower bound proofs as the "generalized" approximation method.

We carefully explain the origins of the fusion method as a certain instantiation of the generalized approximation method. While fusion can be described as an independent framework (cf. [Kar93, Wig93]), establishing lower bounds using the fusion technique inherits some of the difficulties that apply to the approximation method, as briefly mentioned in the concluding remarks of [Raz89].

One of the purposes of our exposition is to clarify such difficulties for a reader that is not so familiar with these approaches, and to explain the role of adaptivity and probabilistic techniques when applying the fusion method against general boolean circuits.

Contents

Date: August 31, 2018.

1. Formalization and limitations of the pure approximation method

1.1. **Notation.** We adopt most of the notation employed in Razborov's work [Raz89].

- We let B^n denote $\{0,1\}^n$.
- F_n is the set of all functions $f\colon B^n \to \{0,1\}$.
- For $u \in B^n$, u^i denotes the i-th bit of u.
- $X_i^\varepsilon \stackrel{\text{def}}{=} \{u \in B^n \mid u^i = \varepsilon\}$, where $\varepsilon \in \{0,1\}$ and $i \in [n]$.
- We let $x_i^1 \stackrel{\text{def}}{=} x_i$ and $x_i^0 \stackrel{\text{def}}{=} \neg x_i$, and often view these literals as boolean functions in F_n.
- A *circuit* has gates $\{\vee, \wedge\}$ and inputs $0, 1, x_i^\varepsilon$, where $\varepsilon \in \{0,1\}$ and $i \in [n]$.
- We use $[\![C]\!] \in F_n$ to denote the function computed by a circuit C.
- $\mathsf{size}(C)$ denotes the circuit size of C, i.e., the number of $\{\vee, \wedge\}$-gates in C.
- $\mathsf{size}(f)$ is the minimum size of a circuit for f.
- Similarly, $\mathsf{size}^+(g)$ denotes the monotone circuit size of a monotone function g.

Definition 1 (Legitimate model)**.** *A set $\mathcal{M} \subseteq F_n$ supplied with two commutative binary operations*

$$\bar{\vee}, \bar{\wedge}\colon \mathcal{M} \times \mathcal{M} \to \mathcal{M}$$

is a (general)[1] legitimate model *(of order n) if*

$$\{0, 1, x_i^\varepsilon \,(i \in [n], \varepsilon \in \{0,1\})\} \subseteq \mathcal{M}.$$

We describe a legitimate model using the tripe $\langle \mathcal{M}, \bar{\vee}, \bar{\wedge} \rangle$.

- Given C, we use $\bar{C}$ to denote the circuit obtained from C after replacing $\{\vee, \wedge\}$ by $\{\bar{\vee}, \bar{\wedge}\}$. The commutativity of the latter operations guarantees that the function in $\mathcal{M}$ computed by the new circuit is well defined.
- Abusing notation, we use $[\![\bar{C}]\!]$ to denote the function in $\mathcal{M}$ computed by the $\mathcal{M}$-circuit $\bar{C}$.
- We will denote boolean functions in $\mathcal{M}$ by $\bar{f}, \bar{g}$, etc.
- For convenience, we often view a boolean function in F_n as a subset of B^n, and write $f \leq g$ to denote $f \subseteq g$.

Definition 2 (Error sets)**.** *We introduce notation to capture the set of "errors" when operations in $\{\vee, \wedge\}$ are performed over $\bar{h}, \bar{g} \in \mathcal{M}$ instead of the $\mathcal{M}$-operations in $\{\bar{\vee}, \bar{\wedge}\}$:*

$$\begin{aligned}
\delta_\wedge^+(\bar{g}, \bar{h}) &\stackrel{\text{def}}{=} (\bar{g} \wedge \bar{h}) \setminus (\bar{g} \,\bar{\wedge}\, \bar{h}),\\
\delta_\wedge^-(\bar{g}, \bar{h}) &\stackrel{\text{def}}{=} (\bar{g} \,\bar{\wedge}\, \bar{h}) \setminus (\bar{g} \wedge \bar{h}),\\
\delta_\vee^+(\bar{g}, \bar{h}) &\stackrel{\text{def}}{=} (\bar{g} \vee \bar{h}) \setminus (\bar{g} \,\bar{\vee}\, \bar{h}),\\
\delta_\vee^-(\bar{g}, \bar{h}) &\stackrel{\text{def}}{=} (\bar{g} \,\bar{\vee}\, \bar{h}) \setminus (\bar{g} \vee \bar{h}).
\end{aligned}$$

Moreover, we set

$$\begin{aligned}
\Delta^+ &\stackrel{\text{def}}{=} \{\delta_\star^+(\bar{g}, \bar{h}) \mid \bar{g}, \bar{h} \in \mathcal{M}, \star \in \{\wedge, \vee\}\},\\
\Delta^- &\stackrel{\text{def}}{=} \{\delta_\star^-(\bar{g}, \bar{h}) \mid \bar{g}, \bar{h} \in \mathcal{M}, \star \in \{\wedge, \vee\}\}.
\end{aligned}$$

We define the distance between $f \in F_n$ and a function $\bar{g} \in \mathcal{M}$ as follows.

[1] The functions $\neg x_i$ are not necessary if we are interested in lower bounds for monotone models of computation. In this case, we might also refer to $\langle \mathcal{M}, \bar{\vee}, \bar{\wedge} \rangle$ as a *monotone* legitimate model.

Definition 3 (Distance of f to $\bar{g} \in \mathcal{M}$)**.** *Given $f \in F_n$ and $\bar{g} \in \mathcal{M}$, we let $\rho_{\mathcal{M}}(f, \bar{g})$ be the minimal $t \in \mathbb{N} \cup \{\infty\}$ for which there exist t triples $(\star_i, \bar{a}_i, \bar{b}_i)$ with $\star_i \in \{\vee, \wedge\}$ and $\bar{a}_i, \bar{b}_i \in \mathcal{M}$ such that*

$$\begin{aligned} f &\leq \bar{g} \vee \bigvee_{i=1}^{t} \delta^{+}_{\star_i}(\bar{a}_i, \bar{b}_i), \\ \bar{g} &\leq f \vee \bigvee_{i=1}^{t} \delta^{-}_{\star_i}(\bar{a}_i, \bar{b}_i). \end{aligned}$$

Intuitively, we approximate f by $\bar{g}$, and cover the errors/mistakes using the minimal number of sets in Δ^+ and Δ^-. We introduce next a measure of distance from an arbitrary $f \in F_n$ to a legitimate model $\mathcal{M}$.

Definition 4 (Distance to a legitimate model)**.** *Given $f \in F_n$ and a legitimate model $\mathcal{M}$ of order n, we let*

$$\rho(f, \mathcal{M}) \stackrel{\text{def}}{=} \min_{\bar{g} \in \mathcal{M}} \rho_{\mathcal{M}}(f, \bar{g}).$$

1.2. **Obtaining lower bounds via the pure approximation method.** A fundamental property of these definitions is the following connection between circuit complexity and distance to a legitimate model.

Proposition 1 (Lower bounds via the pure approximation method)**.** *For every legitimate model $\mathcal{M}$ of order n and circuit C over n input variables,*

$$\rho_{\mathcal{M}}([\![C]\!], [\![\bar{C}]\!]) \leq \mathsf{size}(C).$$

Consequently, for every $f \in F_n$,

$$\mathsf{size}(f) \geq \max_{\mathcal{M}} \rho(f, \mathcal{M}) = \max_{\mathcal{M}} \min_{\bar{g} \in \mathcal{M}} \rho_{\mathcal{M}}(f, \bar{g}),$$

where $\mathcal{M}$ ranges over all legitimate models of order n.

Proof. We say that a gate location g in C has a *plus-error* on a string $w \in \{0,1\}^n$ if $g(w) = 1$ and $\bar{g}(w) = 0$, where $\bar{g}$ is the corresponding gate of $\bar{C}$. For the i-th $\{\vee, \wedge\}$-gate in C, let g_i and h_i be its predecessors in C, and $\star_i$ be the corresponding boolean operation. We claim that

$$[\![C]\!] \leq [\![\bar{C}]\!] \vee \bigvee_{i} \delta^{+}_{\star_i}(\bar{g}_i, \bar{h}_i).$$

Let v be an input such that $C(v) = 1$ and $\bar{C}(v) = 0$. In other words, v is a plus-error for the output gate of C. Since no plus-error can occur over the input gates of C, there exists an internal gate e in C such that e has a plus-error on v, but its predecessors gate locations g and h in C do not have a plus-error on v. (For later reference, we say that such gate location is a *distinguished gate* w.r.t the input v.) Let $e = g \star h$, where $\star \in \{\vee, \wedge\}$. We claim that $v \in \delta^{+}_{\star}(\bar{g}, \bar{h})$.

Recall that $\delta^{+}_{\star}(\bar{g}, \bar{h}) \stackrel{\text{def}}{=} (\bar{g} \star \bar{h}) \setminus (\bar{g} \bar{\star} \bar{h})$. Thus we need to argue that $(\bar{g} \star \bar{h})(v) = 1$ and $(\bar{g} \bar{\star} \bar{h})(v) = 0$. Notice that the assumptions over the gates e, g, and h on v imply that $e(v) = 1$, $\bar{e}(v) = 0$, $\bar{g}(v) \geq g(v)$, and $\bar{h}(v) \geq h(v)$. Finally, since $e = g \star h$, we have $\bar{e} = \bar{g} \bar{\star} \bar{h}$. These inequalities yield $(\bar{g} \bar{\star} \bar{h})(v) = \bar{e}(v) = 0$ and $(\bar{g} \star \bar{h})(v) \geq (g \star h)(v) = e(v) = 1$, as desired.

The proof of the other direction in Definition 3 is similar, employing the natural analogue notion of *minus-errors* instead of plus-errors. □

Consequently, for every legitimate model $\mathcal{M}$ of order n and $f \in F_n$, $\rho_{\mathcal{M}}(f, \mathcal{M}) \leq \mathsf{size}(f) = O(2^n/n)$. In particular, the distance to a legitimate model as introduced in Definition 4 is always finite.

There is an analogue of Proposition 1 for monotone circuits and monotone legitimate models. Under this restriction, it is possible to construct monotone models $\langle \mathcal{M}_n, \bar{\vee}, \bar{\wedge} \rangle$ for explicit functions f such as k-clique witnessing that $\mathsf{size}^+(f) \geq \rho_{\mathcal{M}_n}(f, \mathcal{M}_n) \geq n^{(1-o(1))k}$ [Raz85], where $k \in \mathbb{N}$.

We describe next the methodology employed in the proof of this lower bound, adapted to the setting of general circuits and general legitimate models. We will refer to it as lower bounds obtained via the approximation method using *probabilistic counting arguments*.

- We use boldface symbols such as $\bar{\boldsymbol{g}}$ and $\boldsymbol{\delta}$ to denote random variables. In some places, we abuse notation and also view such objects as distributions. The support of a random variable $\boldsymbol{x}$ is denoted by $\mathsf{support}(\boldsymbol{x})$.
- If E denotes an inequality or a property involving random variables, we use $\mathbf{1}_{[E]}$ to denote the event that E holds.

Fix a non-constant boolean function $f \in F_n$, and consider distributions $\boldsymbol{v} \sim \mathcal{D}^f_{\mathsf{yes}}$ and $\boldsymbol{u} \sim \mathcal{D}^f_{\mathsf{no}}$ supported over $f^{-1}(1)$ and $f^{-1}(0)$, respectively. Now let

$$d^+ \stackrel{\text{def}}{=} \max_{\delta^+ \in \Delta^+} \Pr[\delta^+(\boldsymbol{v}) = 1] \quad \text{and} \quad d^- \stackrel{\text{def}}{=} \max_{\delta^- \in \Delta^-} \Pr[\delta^-(\boldsymbol{u}) = 1].$$

If both d^+ and d^- are non-zero,[2] we set

$$\rho(f, \mathcal{M}, \boldsymbol{v}, \boldsymbol{u}) \stackrel{\text{def}}{=} \min_{\bar{g} \in \mathcal{M}} \max \left\{ \frac{\Pr[\bar{g}(\boldsymbol{v}) = 0]}{d^+}, \frac{\Pr[\bar{g}(\boldsymbol{u}) = 1]}{d^-} \right\}.$$

This provides a convenient approach to prove lower bounds on $\rho(f, \mathcal{M})$, as described next.

Proposition 2 (Lower bound using a probabilistic counting argument)**.** *For every $\boldsymbol{v}$ and $\boldsymbol{u}$ as above, and $\bar{g} \in \mathcal{M}$, we have*

$$\rho_{\mathcal{M}}(f, \bar{g}) \geq \max\{\lceil \Pr[\bar{g}(\boldsymbol{v}) = 0]/d^+ \rceil, \lceil \Pr[\bar{g}(\boldsymbol{u}) = 1]/d^- \rceil\}.$$

Therefore,

$$\rho(f, \mathcal{M}) \geq \rho(f, \mathcal{M}, \boldsymbol{v}, \boldsymbol{u}).$$

Proof. It is not hard to see that $\rho_{\mathcal{M}}(f, \bar{g}) \geq \lceil \Pr[\bar{g}(\boldsymbol{v}) = 0]/d^+ \rceil$ using the corresponding definitions and the fact that $\boldsymbol{v}$ is supported over $f^{-1}(1)$. The proof that $\rho_{\mathcal{M}}(f, \bar{g}) \geq \lceil \Pr[\bar{g}(\boldsymbol{u}) = 1]/d^- \rceil$ is similar. □

[2] Observe that this might not always be the case. For instance, if $\langle \mathcal{M}, \bar{\vee}, \bar{\wedge} \rangle = \langle F_n, \vee, \wedge \rangle$, the corresponding error sets Δ^+ and Δ^- are trivial, and they contain only the empty set.

1.3. **Limitations of the pure approximation method.** In this section we discuss limitations of the approximation method formalized in the preceding section.

Theorem 1 (Limitations of the pure approximation method [Raz89])**.** *There exists a universal constant $\gamma > 0$ for which the following holds. For every $n \in \mathbb{N}$, $f \in F_n$, and legitimate model $\mathcal{M}$ of order n,*

$$\rho(f, \mathcal{M}) \leq \gamma \cdot n^2.$$

An even stronger limitation can be established in the case of lower bounds obtained by the pure approximation method via probabilistic counting arguments.

Theorem 2 (Probabilistic counting arguments in the pure approximation method [Raz89])**.** *For every $n \in \mathbb{N}$, non-constant boolean function $f \in F_n$, and legitimate model $\mathcal{M}$ of order n, the following holds. If $\boldsymbol{v} \sim \mathcal{D}^f_{\mathsf{yes}}$ and $\boldsymbol{u} \sim \mathcal{D}^f_{\mathsf{no}}$ are supported over $f^{-1}(1)$ and $f^{-1}(0)$, respectively, and the corresponding values $d^+ = d^+(\mathcal{M}, \boldsymbol{v})$ and $d^- = d^-(\mathcal{M}, \boldsymbol{u})$ are non-zero, then*

$$\rho(f, \mathcal{M}, \boldsymbol{v}, \boldsymbol{u}) \leq 24n + 24.$$

The proofs of Theorems 1 and 2 rely on the following lemma.

Lemma 1 (Main technical lemma)**.**
Let $\mathcal{M}$ be a legitimate model of order n, and $f \in F_n$ be a non-constant boolean function. Then there are distributions $\bar{\boldsymbol{h}}$, $\boldsymbol{\delta}^+$, and $\boldsymbol{\delta}^-$, supported over $\mathcal{M}$, Δ^+, and Δ^-, respectively, that satisfy the following properties.

- *If $v \in f^{-1}(1)$, then $\Pr[\bar{\boldsymbol{h}}(v) = 0] \leq (12n + 12) \cdot \Pr[\boldsymbol{\delta}^+(v) = 1]$.*
- *If $u \in f^{-1}(0)$, then $\Pr[\bar{\boldsymbol{h}}(u) = 1] \leq (12n + 12) \cdot \Pr[\boldsymbol{\delta}^-(u) = 1]$.*

Roughly speaking, Lemma 1 says that if a fixed input in $\{0,1\}^n$ is not correctly computed by a typical $\bar{\boldsymbol{h}}$ (with respect to f), then this input is covered with reasonable probability by a random $\boldsymbol{\delta}$.

Assuming Lemma 1, we describe the proof of Theorem 2, which is simpler and of more relevance to the results discussed in Section 2. We sketch the proof of Lemma 1 in Appendix A.

Proof of Theorem 2. We are given f, $\mathcal{M}$, $\boldsymbol{v}$, and $\boldsymbol{u}$. In order to upper bound $\rho(f, \mathcal{M}, \boldsymbol{v}, \boldsymbol{u})$, we prove the existence of a function $\bar{g} \in \mathcal{M}$ such that

$$\frac{\Pr[\bar{g}(\boldsymbol{v}) = 0]}{d^+} \leq 24n + 24 \quad \text{and} \quad \frac{\Pr[\bar{g}(\boldsymbol{u}) = 1]}{d^-} \leq 24n + 24.$$

Recall that $d^+ = \max_{\delta^+ \in \Delta^+} \Pr[\delta^+(\boldsymbol{v}) = 1]$, and $d^- = \max_{\delta^- \in \Delta^-} \Pr[\delta^-(\boldsymbol{u}) = 1]$. Observe that

$$\begin{aligned}
\Pr[\bar{\boldsymbol{h}}(\boldsymbol{v}) = 0] &= \sum_{v \in \mathsf{support}(\boldsymbol{v})} \Pr[\bar{\boldsymbol{h}}(v) = 0] \cdot \Pr[\boldsymbol{v} = v] \\
\text{(by Lemma 1)} \quad &\leq \sum_{v \in \mathsf{support}(\boldsymbol{v})} (12n + 12) \cdot \Pr[\boldsymbol{\delta}^+(v) = 1] \cdot \Pr[\boldsymbol{v} = v] \\
\text{(independence of } \boldsymbol{\delta}^+ \text{ and } \boldsymbol{v}) \quad &= (12n + 12) \cdot \Pr[\boldsymbol{\delta}^+(\boldsymbol{v}) = 1] \\
\text{(definition of } d^+) \quad &\leq (12n + 12) \cdot d^+.
\end{aligned}$$

Equivalently, using that $d^+ > 0$,

$$\mathbb{E}_{\bar{g}\sim\bar{\boldsymbol{h}}}\left[\frac{\Pr[\bar{g}(\boldsymbol{v})=0]}{d^+}\right] = \mathbb{E}\left[\frac{1}{d^+}\cdot \mathbf{1}_{[\bar{\boldsymbol{h}}(\boldsymbol{v})=0]}\right] \leq 12n+12.$$

A similar argument shows that

$$\mathbb{E}_{\bar{g}\sim\bar{\boldsymbol{h}}}\left[\frac{\Pr[\bar{g}(\boldsymbol{u})=1]}{d^-}\right] \leq 12n+12.$$

By linearity of expectation,

$$\mathbb{E}_{\bar{g}\sim\bar{\boldsymbol{h}}}\left[\frac{\Pr[\bar{g}(\boldsymbol{v})=0]}{d^+}+\frac{\Pr[\bar{g}(\boldsymbol{u})=1]}{d^-}\right] \leq 24n+24.$$

This shows the existence of the desired function $\bar{g}\in\mathcal{M}$, which completes the proof. □

To sum up, the probabilistic counting arguments that were used to obtain lower bounds in monotone circuit complexity cannot be use to prove super-linear lower bounds against general circuits.

The proof of Theorem 1 is slightly more complicated. One needs to employ Lemma 1 and an amplification argument using a majority function in order to bound the distance of an arbitrary $f\in F_n$ to $\mathcal{M}$. This extra step weakens the upper bound to $O(n^2)$.[3]

2. The fusion framework and the generalized approximation method

2.1. **Auxiliary variables and extended legitimate models.** The results described in this section use terminology that is slightly different than the one appearing in [Raz89].

In order to go beyond the limitations described in Theorem 1, we will introduce in this section the concept of an *extended legitimate model* for a function $f\in F_n$. Roughly speaking, we will consider legitimate models $\mathcal{M}$ of order $N>n$ and an embedding of f as a projection of a boolean function $f_N\in F_N$. The important points are that $\mathsf{size}(f)=\mathsf{size}(F_N)$ and $\mathsf{size}(f_N)\geq\rho(f_N,\mathcal{M})$. However, since $\mathcal{M}$ is of order N, we have more flexibility when defining its operations $\bar{\vee},\bar{\wedge}\colon \mathcal{M}\times\mathcal{M}\to\mathcal{M}$ (given the additional $N-n$ dimensions). This provides the possibility of proving stronger lower bounds on $\rho(f_N,\mathcal{M})$, and consequently also on $\mathsf{size}(f)$.[4] We refer to lower bounds obtained via this technique as the *generalized approximation method*.

Definition 5 (Extended legitimate models)**.** *Let $f\in F_n$, and $N\geq n$. We say that $\langle\mathcal{M},\bar{\vee},\bar{\wedge}\rangle$ is an* extended legitimate model for f *(of order N) is the following holds*:

- $\mathcal{M}\subseteq F_N$.
- $\bar{\vee},\bar{\wedge}\colon\mathcal{M}\times\mathcal{M}\to\mathcal{M}$ *are commutative operations.*
- $\{0,1,x_i^{\varepsilon}\,(i\in[n],\varepsilon\in\{0,1\})\}\subseteq\mathcal{M}$.

(*Observe that the functions $x_{n+1},\ldots,x_N\in F_N$ are not required to be in $\mathcal{M}$.*)

[3] While the argument in [Raz89] directly produces an $\mathcal{M}$-approximator for each $f\in F_n$, we remark that by the results in [RR97, Section 5] it is enough to show that a *random* function $\boldsymbol{h}\sim F_n$ satisfies $\rho(\boldsymbol{h},\mathcal{M})=O(n^2)$.

[4] The reason we call $\mathcal{M}$ an *extended* legitimate model instead of simply a legitimate model of order N is because (due to our embedding of f in f_N as a projection) there is no need to force the functions $x_{n+1},\ldots,x_N\in F_N$ to be in $\mathcal{M}$ (as in Definition 1). We remark that this distinction is not explicitly made in [Raz89, pg. 170], though it appears to us that the legitimate models investigated there are not required to contain the functions $x_{n+1},\ldots,x_N\in F_N$.

Remark 1. *Observe that in this definition the model $\mathcal{M}$ depends on the parameter n, the number of input arguments of f, but not on the values of f. Since all concrete extended legitimate models discussed later in the text will actually rely on the values of the function f, we maintain the current terminology.*

- Given $f \in F_n$ and $N \geq n$, we use $f_N \in F_N$ to denote the boolean function defined as follows. For every $x \in \{0,1\}^n$ and $x' \in \{0,1\}^{N-n}$, $f_N(xx') \stackrel{\text{def}}{=} f(x)$.
- For $\mathcal{M}$ an extended legitimate model for $f \in F_n$ of order N, $h \in F_N$, and $\bar{g} \in \mathcal{M}$, the definitions of $\rho_{\mathcal{M}}(h, \bar{g})$ and $\rho(h, \mathcal{M})$ remain unchanged.

Claim 1. *For $f \in F_n$, $f_N \in F_N$, and $\mathcal{M}$ as above,*

$$\mathsf{size}(f) \;=\; \mathsf{size}(f_N) \;\geq\; \rho(f_N, \mathcal{M}).$$

Proof. First, $\mathsf{size}(f) \leq \mathsf{size}(F_N)$ since a circuit $C(x_1, \ldots, x_n)$ for f can also be viewed as a circuit $C(x_1, \ldots, x_N)$ for f_N. On the other hand, from any circuit $C(x_1, \ldots, x_N)$ for f_N, the circuit $C'(x_1, \ldots, x_n) \stackrel{\text{def}}{=} C(x_1, \ldots, x_n, \vec{0})$ computes f_n, and has no larger size that C.

In order to see that $\rho(f_N, \mathcal{M}) \leq \mathsf{size}(f_N)$, let $C(x_1, \ldots, x_N)$ be an optimal circuit for f_N, the embedding of f in F_N. This time we view the circuit $C'(x_1, \ldots, x_n) \stackrel{\text{def}}{=} C(x_1, \ldots, x_n, \vec{0})$ as a circuit computing a function in F_N. Clearly, $f_N = [\![C']\!]$, since f_N does not depend on the input variables $x_{n+1}, \ldots, x_N$. Now an argument entirely analogous to the proof of Proposition 1 establishes that $\rho(f_N, \mathcal{M}) \leq \mathsf{size}(C') \leq \mathsf{size}(f_N)$. □

- From now on, let $f \in F_n$ be a *non-constant* boolean function.
- Let $U \stackrel{\text{def}}{=} f^{-1}(0)$, and $V \stackrel{\text{def}}{=} f^{-1}(1)$.
- We use $\mathcal{P}(U)$ to denote the set of subsets of U. For a family $\mathcal{F} \subseteq \mathcal{P}(U)$, we use $\mathcal{F}(A) \in \{0,1\}$ to denote whether $A \in \mathcal{F}$.

Definition 6 (Semi-filter consistent with v)**.** *We say that a family $\mathcal{F} \subseteq \mathcal{P}(U)$ is a* semi-filter *if the following conditions hold*:

- (*non-trivial*) $\mathcal{F}(\emptyset) = 0$ *and* $\mathcal{F}(U) = 1$.
- (*monotonicity*) *For* $A \subseteq B \subseteq U$, $\mathcal{F}(A) \leq \mathcal{F}(B)$.

A semi-filter $\mathcal{F}$ is said to be consistent with $v \in V$ *if*

- *For each* $i \in [n]$, $\mathcal{F}(U \cap X_i^{\varepsilon}) = v^i \oplus \varepsilon \oplus 1$.

Let $\mathfrak{F}_v$ denote the class of all semi-filters consistent with $v \in V$, and set $\mathfrak{F} \stackrel{\text{def}}{=} \bigcup_{v \in V} \mathfrak{F}_v$.

- Observe that by definition if a semi-filter $\mathcal{F}$ is consistent with v_1 and with v_2, then $v_1 = v_2$. We let $v(\mathcal{F})$ be the unique element $v \in V$ such that $\mathcal{F} \in \mathfrak{F}_v$.

The extended legitimate models $\mathcal{M}(\mathfrak{F}')$.

- Let $\mathfrak{F}' \subseteq \mathfrak{F}$ be a non-empty collection of semi-filters, and $N \stackrel{\text{def}}{=} n + \lceil \log |\mathfrak{F}'| \rceil$.
- In order to avoid trivial considerations, we assume that there exist $v_1, v_2 \in V$ with $v_1 \neq v_2$ such that $\mathfrak{F}_{v_1} \cap \mathfrak{F}' \neq \emptyset$ and $\mathfrak{F}_{v_2} \cap \mathfrak{F}' \neq \emptyset$. We say that a collection of semi-filters with this property is a *non-trivial* collection.
- Fix a surjective function $\gamma \colon \{0,1\}^{N-n} \to \mathfrak{F}'$. From now on, we identify $y \in \{0,1\}^{N-n}$ with the semi-filter $\mathcal{F} = \gamma(y)$.

Given $\mathfrak{F}'$ as above, we define an extended legitimate model of order N for the boolean function $f \in F_n$. In order to do that, to each $g \in F_n$ we associate a function $\bar{g} \in F_N$, defined as follows:

$$\bar{g}(x,y) \;=\; \bar{g}(x,\mathcal{F}) \;=\; \begin{cases} g(x) & \text{if } \; x \neq v(\mathcal{F}), \\ \mathcal{F}(U_g) & \text{if } \; x = v(\mathcal{F}), \end{cases}$$

where $U_g \stackrel{\text{def}}{=} U \cap g^{-1}(1)$. We let $\mathcal{M}(\mathfrak{F}') \stackrel{\text{def}}{=} \{\bar{g} \mid g \in F_n\}$. Finally, we define its corresponding binary operations $\bar{\vee}$ and $\bar{\wedge}$ by:

$$\bar{g} \,\bar{\vee}\, \bar{h} \;\stackrel{\text{def}}{=}\; \overline{g \vee h} \qquad \text{and} \qquad \bar{g} \,\bar{\wedge}\, \bar{h} \;\stackrel{\text{def}}{=}\; \overline{g \wedge h}.$$

Proposition 3. *The operations $\bar{\vee}, \bar{\wedge} \colon \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ are well-defined, and $\langle \mathcal{M}(\mathfrak{F}'), \bar{\vee}, \bar{\wedge} \rangle$ is an extended legitimate model for f of order N.*

Proof. It is clear that $\mathcal{M}(\mathfrak{F}') \subseteq F_N$, and that $\bar{\vee}, \bar{\wedge} \colon \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ are commutative operations. It also follows easily from the definitions that $\{0, 1, x_i^{\varepsilon} \, (i \in [n], \varepsilon \in \{0,1\})\} \subseteq \mathcal{M}(\mathfrak{F}')$ as functions in F_N, since the definitions of $\bar{g}(x, \mathcal{F})$ and of semi-filter consistent with v imply that $\bar{0} = 0$, $\bar{1} = 1$, and that $\overline{x_i^{\varepsilon}} = x_i^{\varepsilon}$. Finally, to prove that the operations are well-defined, observe that if $\overline{g_1} = \overline{g_2}$ for $g_1, g_2 \in F_n$, then $g_1 = g_2$. This follows immediately from the definition of $\bar{g}$, our assumptions on $\mathfrak{F}'$, and the fact that the map γ is surjective. □

2.2. **Extended legitimate models and the fusion method.** In this section we discuss special features of the models $\mathcal{M}(\mathfrak{F}')$. Such properties considerably simplify the generalized approximation method, and give rise to a certain combinatorial characterization of $\rho(f_N, \mathcal{M}(\mathfrak{F}'))$ as a cover problem. We refer to the corresponding combinatorial cover problem and its investigation as the *fusion method*.

Let $\delta_{\wedge}^{+}$, $\delta_{\wedge}^{-}$, $\delta_{\vee}^{+}$ and $\delta_{\vee}^{-}$ be the error functions in $\mathcal{M}(\mathfrak{F}') \times \mathcal{M}(\mathfrak{F}') \to \mathcal{P}(\{0,1\}^N)$ associated to $\mathcal{M}(\mathfrak{F}')$ (Definition 2). Recall that $f \in F_n$ embeds into $f_N \in F_N$.

Theorem 3 (The fusion method for lower bounds)**.** *Let $\mathcal{M} = \mathcal{M}(\mathfrak{F}')$ be the extended legitimate model for $f \in F_n$ of order N under consideration. Then,*

(*i*) *For any $\bar{g}, \bar{h} \in \mathcal{M}$,*

$$\delta_{\wedge}^{-}(\bar{g}, \bar{h}) = \delta_{\vee}^{+}(\bar{g}, \bar{h}) = \emptyset.$$

Consequently, the mistakes appearing during the approximation can only be covered by sets in $\Delta_{\wedge}^{+} \stackrel{\text{def}}{=} \{\delta_{\wedge}^{+}(\bar{g}, \bar{h}) \mid \bar{g}, \bar{h} \in \mathcal{M}\}$ and $\Delta_{\vee}^{-} \stackrel{\text{def}}{=} \{\delta_{\vee}^{-}(\bar{g}, \bar{h}) \mid \bar{g}, \bar{h} \in \mathcal{M}\}$.

(*ii*) *The only element in $\mathcal{M}$ that can be used to approximate $f_N \in F_N$ as given by Definition 3 is the function $\bar{f}$. In other words,*

$$\rho(f_N, \mathcal{M}) = \rho_{\mathcal{M}}(f_N, \bar{f}) \in \mathbb{N} \quad \text{and} \quad \rho_{\mathcal{M}}(f_N, \bar{g}) = \infty \text{ for every } \bar{g} \neq \bar{f} \text{ in } \mathcal{M}.$$

Moreover, $\bar{f} \leq f_N$, and no sets from $\Delta_{\vee}^{-}$ are required for the approximation.

(*iii*) *For $A, B \subseteq U$, we say that the pair (A, B)* covers *a semi-filter $\mathcal{F} \in \mathfrak{F}'$ if*

$$\mathcal{F}(A) = 1, \quad \mathcal{F}(B) = 1, \quad \mathcal{F}(A \cap B) = 0.$$

Then $\rho_{\mathcal{M}}(f_N, \bar{f})$ is the minimum number of pairs (A, B) that cover all semi-filters in $\mathfrak{F}'$. Furthermore, a mistake occurs on an input $(x, y) \in \{0,1\}^N$ if and only if it is of the form $(v(\mathcal{F}), \mathcal{F})$.

Proof. In order to prove (i), let $\bar{g}, \bar{h} \in \mathcal{M}$. As in the proof of Proposition 3, there are unique functions $g, h \in F_n$ mapping to $\bar{g}$ and $\bar{h}$, respectively. Observe that:

$$\begin{aligned}\delta^+_\vee(\bar{g},\bar{h}) &= (\bar{g} \vee \bar{h}) \setminus (\bar{g} \barwedge \bar{h}) = (\bar{g} \vee \bar{h}) \setminus (\overline{g \vee h}),\\ \delta^-_\wedge(\bar{g},\bar{h}) &= (\bar{g} \barwedge \bar{h}) \setminus (\bar{g} \wedge \bar{h}) = (\overline{g \wedge h}) \setminus (\bar{g} \wedge \bar{h}).\end{aligned}$$

Let $(x, y) = (x, \mathcal{F}) \in \{0,1\}^N$ be an input. If $x \neq v(\mathcal{F})$ then $\bar{g}(x, \mathcal{F}) = g(x)$, $\bar{h}(x, \mathcal{F}) = h(x)$, $(\overline{g \vee h})(x, \mathcal{F}) = (g \vee h)(x)$, etc., which imply that the previous equations evaluate to 0 on $(x, \mathcal{F})$. On the other hand, if $x \neq v(\mathcal{F})$, the expressions evaluate on $(x, \mathcal{F})$ to

$$(\mathcal{F}(U_g) \vee \mathcal{F}(U_h)) \setminus \mathcal{F}(U_{g\vee h}) \quad \text{and} \quad \mathcal{F}(U_{g\wedge h}) \setminus (\mathcal{F}(U_g) \wedge \mathcal{F}(U_h)),$$

respectively. Since $\mathcal{F}$ is a monotone semi-filter, it follows that both expressions evaluate to 0 as well. This completes the proof that the sets $\delta^+_\vee$ and $\delta^-_\wedge$ are trivial.

In (ii), it is clear that $\rho(f_N, \mathcal{M})$ is finite, since $\rho(f_N, \mathcal{M}) \leq \mathsf{size}(f_N) \leq 2^n$ according to Claim 1.

On the other hand, suppose $\bar{g} \neq \bar{f}$. Then $g \neq f$, as observed before. Let $w \in \{0,1\}^n$ be an input such that $f(w) \neq g(w)$, and $\mathcal{F} \in \mathfrak{F}'$ be a semi-filter such that $v(\mathcal{F}) \neq w$ (such semi-filter exists since $\mathfrak{F}'$ is assumed to be a non-trivial collection). By definition, $\bar{f}(w, \mathcal{F}) = f(w) = f_N(w, \mathcal{F})$ and $\bar{g}(w, \mathcal{F}) = g(w)$, which gives $f_N(w, \mathcal{F}) \neq \bar{g}(w, \mathcal{F})$. Since $w \neq v(\mathcal{F})$, it is easy to see that the input $(w, \mathcal{F})$ cannot be covered by an error set. This shows that $\rho_{\mathcal{M}}(f_N, \bar{g}) = \infty$.

For the moreover part of the claim, observe that on an input $(x, \mathcal{F})$ with $x \neq v(\mathcal{F})$ we get $\bar{f}(x, \mathcal{F}) = f(x) = f_N(x)$, while if $x = v(\mathcal{F})$, we obtain $\bar{f}(x, \mathcal{F}) = \mathcal{F}(U_f) = \mathcal{F}(\emptyset) = 0$. In particular, by inspecting Definition 3 it is easy to see that sets in $\Delta^-_\vee$ do not play a role in the approximation.

We proceed with the proof of (iii). By the previously established claims, $\rho_{\mathcal{M}}(f_N, \bar{f})$ is the minimum number $t \in \mathbb{N}$ of sets $\delta^+_\wedge(\bar{g}_i, \bar{h}_i) \in \Delta^+_\wedge$ such that

$$f_N \;\leq\; \bar{f} \vee \bigvee_{i=1}^{t} \delta^+_\wedge(\bar{g}_i, \bar{h}_i).$$

Let $T \stackrel{\text{def}}{=} \{(x, \mathcal{F}) \in \{0,1\}^N \mid f_N(x, \mathcal{F}) = 1 \text{ and } \bar{f}(x, \mathcal{F}) = 0\}$ be the set of inputs that need to be covered. We claim that $(x, \mathcal{F}) \in T$ if and only if $x = v(\mathcal{F})$. Indeed, if $f_N(x, \mathcal{F}) = 1$ and $\bar{f}(x, \mathcal{F}) = 0$ we must have $x = v(\mathcal{F})$, as otherwise $f_N(x, \mathcal{F}) = \bar{f}(x, \mathcal{F}) = f(x)$. On the other hand, if $x = v(\mathcal{F})$ note that $f_N(x, \mathcal{F}) = f(x) = 1$, using that $v(\mathcal{F}) \in V = f^{-1}(1)$. In addition, $\bar{f}(x, \mathcal{F}) = \mathcal{F}(U_f) = \mathcal{F}(\emptyset) = 0$, which is a property of semi-filters.

Now to cover an input $(x, \mathcal{F})$ satisfying $x = v(\mathcal{F})$ by a set in $\Delta^+_\wedge$, it is necessary and sufficient that

$$(\bar{g} \wedge \bar{h})(x, \mathcal{F}) = 1 \quad \text{and} \quad (\bar{g} \barwedge \bar{h})(x, \mathcal{F}) = (\overline{g \wedge h})(x, \mathcal{F}) = 0.$$

Using $x = v(\mathcal{F})$, this is equivalent to

$$\mathcal{F}(U_g) = \mathcal{F}(U_h) = 1 \quad \text{and} \quad \mathcal{F}(U_{g\wedge h}) = 0.$$

Finally, during the construction of the cover the sets $U_g = A$ and $U_h = B$ can be arbitrary subsets of U, by the fact that $\mathcal{M}$ is in one-to-one correspondence with F_n. This proves that

$\rho_{\mathcal{M}}(f_N, \bar{f})$ is precisely the minimum number of pairs (A, B) of subsets of U that cover all semi-filters in $\mathfrak{F}'$. □

Observe that in the proof we have used that the second condition in the definition of $\bar{g}(x, y)$ enforces that $x = v(\mathcal{F}) \in V$ is a *positive* input of f. However, it is not necessary in the proof of Theorem 3 to use the assumption that the semi-filter $\mathcal{F}$ is *consistent with* v.

Consider now the entire collection $\mathfrak{F} = \bigcup_{v \in V} \mathfrak{F}_v$ as introduced in Definition 6, and let $\mathcal{M}_{\mathsf{max}} \stackrel{\text{def}}{=} \mathcal{M}(\mathfrak{F}) \subseteq F_{N_{\mathsf{max}}}$ be the corresponding extended legitimate model of f (over some choice of the surjective map γ).

Theorem 4 (Completeness of the fusion method)**.** *There exists a universal constant $\zeta > 0$ independent of n, $f \in F_n$, and $\mathcal{M}_{\mathsf{max}}$ for which the following holds*:

$$\mathsf{size}(f) \leq \zeta \cdot (\rho(f_{N_{\mathsf{max}}}, \mathcal{M}_{\mathsf{max}}) + n)^3.$$

In other words, it follows from Claim 1, Theorem 3, and Theorem 4 that the circuit complexity of a boolean function $f\colon \{0,1\}^n \to \{0,1\}$ can be *characterized* up to a polynomial by a purely combinatorial cover problem. A proof of Theorem 4 can be found in [Raz89]. We remark that the proof of this result makes crucial use that each semi-filter is *consistent* with a (positive) input of f.

2.3. **Probabilistic counting arguments in the generalized approximation method.** Let $f \in F_n$, and $\mathcal{M}$ be an arbitrary extended legitimate model of order N for f, where $N \geq n$. Following our convention, $f_N \in F_N$ denotes the extension of f to an N-bit function that depends only on $x_1, \ldots, x_n$.

Similarly to our treatment of probabilistic counting arguments in Section 1.2, consider random variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ supported over $f_N^{-1}(1)$ and $f_N^{-1}(0)$, respectively. Define d^+ and d^- as before. If these values are non-zero, $\rho(f_N, \mathcal{M}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is well-defined, and it is not hard to show that

$$\mathsf{size}(f) \;=\; \mathsf{size}(f_N) \;\geq\; \rho(f_N, \mathcal{M}) \;\geq\; \rho(f_N, \mathcal{M}, \boldsymbol{\alpha}, \boldsymbol{\beta}).$$

It thus make sense to consider possibly stronger bounds that one can obtain on $\rho(f_n, \mathcal{M}, \boldsymbol{\alpha}, \boldsymbol{\beta})$, i.e., the power of *probabilistic counting arguments* in the *generalized approximation method*.

Unfortunately, an analogue of Lemma 1 still holds in this context, as stated next.

Lemma 2. *Let $\mathcal{M}$ be an extended legitimate model of order N for $f \in F_n$, where f is non-constant and $N \geq n$. There are distributions $\bar{\boldsymbol{h}}$, $\boldsymbol{\delta}^+$, and $\boldsymbol{\delta}^-$, supported over $\mathcal{M}$, Δ^+, and Δ^-, respectively, that satisfy the following properties.*

- *If $\alpha \in f_N^{-1}(1)$, then $\Pr[\bar{\boldsymbol{h}}(\alpha) = 0] \leq (12n + 12) \cdot \Pr[\boldsymbol{\delta}^+(\alpha) = 1]$.*
- *If $\beta \in f_N^{-1}(0)$, then $\Pr[\bar{\boldsymbol{h}}(\beta) = 1] \leq (12n + 12) \cdot \Pr[\boldsymbol{\delta}^-(\beta) = 1]$.*

Crucially, in the previous inequalities we have the parameter n instead of N. We discuss the proof of this lemma in Section A (it is immediate from the proof of Lemma 1, and in [Raz89] these lemmas are treated together).

Proceeding as in the proof of Theorem 2, we get the following consequence.

Theorem 5. *Let f, n, $\mathcal{M}$, f_N, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ be as described above. If $d^+ = d^+(\mathcal{M}, \boldsymbol{\alpha})$ and $d^- = d^-(\mathcal{M}, \boldsymbol{\beta})$ are non-zero, then*

$$\rho(f, \mathcal{M}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \leq 24n + 24.$$

In particular, one cannot use an extended model $\mathcal{M}$ of order $N \gg n$ and probabilistic counting arguments to a prove super-linear circuit lower bound for f.

However, we stress that the analogue of Theorem 1 obtained via Lemma 2 only provides an upper bound of order $N \cdot n$. Indeed, as we explained in Section 2.2, circuit lower bounds for $f\colon \{0,1\}^n \to \{0,1\}$ are in fact polynomially equivalent to establishing lower bounds on $\rho(f_N, \mathcal{M})$, for an appropriate choice of the extended legitimate model $\mathcal{M}$ (such as the model $\mathcal{M}_{\mathsf{max}}$). Thus non-trivial lower bounds can be established within the generalized approximation method framework via techniques that go beyond probabilistic counting arguments.

2.4. **On proving non-monotone lower bounds using the fusion framework.** We discuss in this section the role of probabilistic arguments in the context of the fusion framework, i.e., with respect to the extended models $\mathcal{M} = \mathcal{M}(\mathfrak{F}')$, where $\mathfrak{F}' \subseteq \bigcup_{v \in f^{-1}(1)} \mathfrak{F}_v$ (Definition 6) is non-trivial.

Recall that $\rho(f_N, \mathcal{M}) = \rho_{\mathcal{M}}(f_N, \bar{f})$ and $\bar{f} \leq f_N$ (Theorem 3). Consequently, for any distribution $\boldsymbol{\beta}$ supported over $f_N^{-1}(0)$, we have $\Pr[\bar{f}(\boldsymbol{\beta}) = 1] = 0$. This implies that $\rho(f_N, \mathcal{M}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = 0$, i.e., the approach described in Section 2.3 is trivial in the fusion context.

In order to remedy this situation, it is natural to consider a one-sided version of this approach. We crucially use Theorem 3 to simplify our notation. Fix a single distribution $\boldsymbol{\alpha}$ supported over $T \stackrel{\text{def}}{=} \{(v(\mathcal{F}), \mathcal{F}) \mid \mathcal{F} \in \mathfrak{F}'\} \subseteq f_N^{-1}(1)$ (recall that $f_N(v, \mathcal{F}) = f(v) = 1$, since v is by assumption a positive input of f). Here we focused on this particular subset of $f_N^{-1}(1)$ because by Theorem 3 (*iii*) errors cannot occur over $f_N^{-1}(1) \setminus T$. Observe that $\boldsymbol{\alpha}$ is simply a distribution over $\mathfrak{F}'$, given that $v(\mathcal{F})$ is determined by $\mathcal{F}$.

In addition, we let $d^+ \stackrel{\text{def}}{=} \max_{\delta_\wedge^+ \in \Delta_\wedge^+} \Pr[\delta_\wedge^+(\boldsymbol{\alpha}) = 1]$. We take the maximum over $\Delta_\wedge^+$ instead of Δ^+ thanks to Theorem 3 (*i*). Observe that d^+ is simply the maximum measure (with respect to $\boldsymbol{\alpha}$) of the number of semi-filters in $\mathfrak{F}'$ that are covered by a single pair (A, B) with $A, B \subseteq U$, according to Theorem 3 (*iii*). If $d^+ \neq 0$, we set

$$\rho(f_N, \mathcal{M}, \boldsymbol{\alpha}) \stackrel{\text{def}}{=} \frac{\Pr[\bar{f}(\boldsymbol{\alpha}) = 0]}{d^+} = \frac{1}{d^+}.$$

We consider only $\bar{f}$ thanks to Theorem 3 (*ii*). It should be clear that

$$\rho(f_N, \mathcal{M}) = \rho(f_N, \bar{f}) \geq \rho(f_N, \mathcal{M}, \boldsymbol{\alpha}).$$

We proceed to describe an important difficulty when one tries to apply the fusion method to non-monotone circuits. The one-sided variant of Lemmas 1 and 2 can be stated in the language of the fusion method as follows.

Lemma 3 (Following the concluding remarks in [Raz89])**.** *There exists a distribution $(\boldsymbol{A}, \boldsymbol{B})$ over pairs (A, B) with $A, B \subseteq U = f^{-1}(0)$ such that for every semi-filter $\mathcal{F} \in \mathfrak{F}'$,*

$$\Pr[\,(\boldsymbol{A}, \boldsymbol{B}) \textit{ covers } \mathcal{F}\,] \geq \frac{1}{8n+8}.$$

Lemma 3 is obtained by specializing the proof of Lemma 2 to the extended legitimate models employed in the fusion method. While we will discuss the argument in Section A, let us emphasize that so far we have only adapted the terminology of Section 2.3 to fusion (thanks to Theorem 3). In order to establish Lemma 3, it is enough to argue that the analogue of $\Pr[\bar{\boldsymbol{h}}(\alpha) = 0]$ from Lemma 2 occurs with probability 1 due to our choice of

$\mathcal{M}(\mathfrak{F}')$, and to use properties of this class of extended legitimate models to obtain slightly better bounds.

A simpler averaging argument (compared to the proofs of Theorems 2 and 5) implies the following result.

Theorem 6. *Under the notation and assumptions made above,*

$$\rho(f_N, \mathcal{M}, \boldsymbol{\alpha}) \;\leq\; 8n + 8.$$

In contrast to the monotone case (in the context of the pure approximation method), this result says that the power of the fusion method can only be harnessed via the use of techniques that either go beyond or adapt the aforementioned probabilistic counting arguments. (Recall that, by Theorem 4, fusion can be used to characterize circuit size.)

Can probabilistic arguments still be useful? The next discussion assumes basic familiarity with [Kar93] or [Raz89, Section 4], where it is explained how the pairs (A, B) that appear in Theorem 3 are connected to the AND-gates of a boolean circuit.

The right way to proceed in a lower bound proof against general circuits in the fusion framework is by picking the distribution over semi-filters after a candidate small circuit C has been exposed. If C could correctly compute f, its set Γ_C of AND-gates (A, B) (relative to $U = f^{-1}(0)$) would cover all semi-filters in $\mathfrak{F}$ (in the sense of Theorem 3). But if f is hard and C is small (say, of size s), we can inspect the collection Γ_C of pairs (A, B) and produce a distribution $\boldsymbol{\alpha}_C$ supported over $\mathfrak{F}$ such that for every $(A, B) \in \Gamma_C$,

$$\Pr_{(v(\boldsymbol{\mathcal{F}}),\boldsymbol{\mathcal{F}})\sim\boldsymbol{\alpha}_C}[\,(A, B) \text{ covers } \boldsymbol{\mathcal{F}}\,] < 1/s.$$

If this can be done for every circuit C of size $\leq s$, then $\mathsf{size}(f) > s$.

Clearly, if f is not computable by size-s circuits, then there is some distribution $\boldsymbol{\alpha}_C$ with this property for each circuit C of size $\leq s$. The challenge of course is to unconditionally establish the existence of such distribution, which is equivalent to proving the corresponding circuit lower bound. (We remark that the strategy employed in [Kar93] for *monotone* circuit lower bounds for 3-clique can be naturally formulated in this context.)

We conclude that for *non-monotone* boolean circuits the generalized approximation method (which includes the fusion method) needs to be applied in an *adaptive* way if the lower bound proof is based solely on probabilistic counting arguments (as opposed to the simpler case of *monotone* circuits). One cannot hope to consider a fixed distribution of input instances (pure approximation) or semi-filters (fusion) and to *count* the number of errors in each approximation step. However, we stress that a successful proof might still employ probabilistic arguments in a crucial way, by *adapting* the distribution over instances *given the circuit* that must be defeated.

References

[AB87] Noga Alon and Ravi B. Boppana. The monotone circuit complexity of boolean functions. *Combinatorica*, 7(1):1–22, 1987.

[BS90] Ravi B. Boppana and Michael Sipser. The complexity of finite functions. In *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity*, pages 757–804. 1990.

[COS17] Xi Chen, Igor C. Oliveira, and Rocco A. Servedio. Addition is exponentially harder than counting for shallow monotone circuits. In *Symposium on Theory of Computing* (STOC), pages 1232–1245, 2017.

[dOOP17] Mateus de Oliveira Oliveira and Pavel Pudlák. Representations of monotone boolean functions by linear programs. *Electronic Colloquium on Computational Complexity* (ECCC), 24:106, 2017.

[Juk01] S. Jukna. *Extremal combinatorics with applications in computer science*. Springer, 2001.

[Kar93] Mauricio Karchmer. On proving lower bounds for circuit size. In *Structure in Complexity Theory Conference* (CCC), pages 112–118, 1993. 1, 12

[KO17] Jan Krajíček and Igor C. Oliveira. On monotone circuits with local oracles and clique lower bounds. Available at arXiv:1704.06241, 2017.

[Kra97] Jan Krajíček. Interpolation theorems, lower bounds for proof systems, and independence results for bounded arithmetic. *J. Symbolic Logic*, 62(2):457–486, 1997.

[Kra16] Jan Krajíček. Randomized feasible interpolation and monotone circuits with a local oracle. Available at arXiv:1611.08680, 2016.

[PR17] Toniann Pitassi and Robert Robere. Strongly exponential lower bounds for monotone computation. In *Symposium on Theory of Computing* (STOC), pages 1246–1255, 2017.

[PS96] Pavel Pudlák and Jirí Sgall. Algebraic models of computation and interpolation for algebraic proof systems. In *DIMACS Workshop on Proof Complexity and Feasible Arithmetics*, pages 279–296, 1996.

[Pud97] Pavel Pudlák. Lower bounds for resolution and cutting plane proofs and monotone computations. *J. Symbolic Logic*, 62(3):981–998, 1997.

[Raz85] Alexander A. Razborov. Lower bounds for the monotone complexity of some boolean functions. *Soviet Math. Doklady*, 31:354–357, 1985. 4

[Raz89] Alexander A. Razborov. On the method of approximations. In *Symposium on Theory of Computing* (STOC), pages 167–176, 1989. 1, 2, 5, 6, 10, 11, 12

[Raz90] Alexander A. Razborov. Lower bounds of the complexity of symmetric boolean functions of contact-rectifier circuits. *Mathematical Notes*, 48(6):1226–1234, 1990.

[Ros14] Benjamin Rossman. The monotone complexity of k-clique on random graphs. *SIAM J. Comput.*, 43(1):256–279, 2014.

[Ros15] Benjamin Rossman. Correlation bounds against monotone NC^1. In *Conference on Computational Complexity* (CCC), pages 392–411, 2015.

[RR97] Alexander A. Razborov and Steven Rudich. Natural proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, 1997. 6

[ST00] Janos Simon and Shi-Chun Tsai. On the bottleneck counting argument. *Theor. Comput. Sci.*, 237(1-2):429–437, 2000.

[Tar88] Éva Tardos. The gap between monotone and non-monotone circuit complexity is exponential. *Combinatorica*, 8(1):141–142, 1988.

[Wig93] Avi Wigderson. The fusion method for lower bounds in circuit complexity. In *Combinatorics, Paul Erdos is Eighty, Bolyai Math. Society*, pages 453–467, 1993. 1

Appendix A. The proof of Lemma 1 and its variants

In this section we sketch the proofs of Lemmas 1, 2, and 3.

Proof of Lemma 1. Let $\mathcal{M}$ be a legitimate model of order n, and $f \in F_n$ be a non-constant function. Our goal is to define distributions $\bar{\boldsymbol{h}}$, $\boldsymbol{\delta}^+$, and $\boldsymbol{\delta}^-$ with the desired properties. For simplicity, we will focus here only on $\bar{\boldsymbol{h}}$ and $\boldsymbol{\delta}^+$ (the construction and argument for $\boldsymbol{\delta}^-$ is not very different). Recall the proof of Proposition 1, and the definition of distinguished gate location. There one constructs an $\mathcal{M}$-circuit $\overline{C}$ from a circuit C that correctly computes f. This $\mathcal{M}$-circuit and its gate locations are used to define the corresponding function in $\mathcal{M}$ and the pairs associated with the sets in Δ^+, respectively.

The argument here will borrow part of this idea. The main difficulty is that we don't have a circuit for f of small size. Informally, the proof consists of three mains steps:

(i) Define a "brute-force" circuit $\boldsymbol{C}$ of size exponential in n that correctly computes f, and consider its corresponding $\mathcal{M}$-circuit $\overline{\boldsymbol{C}}$ and associated function $\bar{\boldsymbol{h}}$.

(ii) For each $v \in f^{-1}(1)$ that is a plus-error for the pair $(\boldsymbol{C}, \overline{\boldsymbol{C}})$, we isolate a set S_v of $O(n)$ gate locations in $\boldsymbol{C}$ that will always contain a *distinguished gate location* for v (in particular, it covers this mistake, as in the proof of Proposition 1). $\boldsymbol{C}$ is actually a random circuit, and while different inputs v will be associated to different gate locations S_v, all circuits in the support of $\boldsymbol{C}$ share the same structure (i.e., directed acyclic graph), and for those that make a plus-error on v, we can take the same (small) set S_v of gate locations.

(iii) Crucially, these properties and the randomized construction of $\boldsymbol{C}$ will allow us to define $\boldsymbol{\delta}^+$ in a way that is independent of v, while we will still able to prove that for each $v \in f^{-1}(1)$,

$$\Pr[\bar{\boldsymbol{h}}(v) = 0] \leq (12n + 12) \cdot \Pr[\boldsymbol{\delta}^+(v) = 1]. \tag{1}$$

In other words, distinguished gates (for each fixed v) will be sampled by $\boldsymbol{\delta}^+$ with the desired probability, although the definition of $\boldsymbol{\delta}^+$ does not depend on a particular v.

We provide more details now. For $g \in F_n$ over input variables $x_1, \ldots, x_n$, we recursively define a circuit C_g using binary operations in $\{\vee, \wedge\}$ and inputs in $\{x_1, \ldots, x_d, \overline{x_1}, \ldots, \overline{x_d}, 0, 1\}$ such that $[\![C_g]\!] = g$. For $\varepsilon \in \{0, 1\}$, set $g^\varepsilon \stackrel{\text{def}}{=} g(x_1, \ldots, x_{d-1}, \varepsilon) \in F_{d-1}$. We let

$$C_g \stackrel{\text{def}}{=} (C_{g^0} \wedge x_d^0) \vee (C_{g^1} \wedge x_d^1), \tag{2}$$

where in the base case consisting of a boolean function over no input variables (i.e. a constant), the corresponding sub-circuit is replaced by the appropriate constant (no syntactic simplifications take place over the circuits produced by this recursive procedure). Roughly speaking, C_g can be viewed as a complete decision tree for g.

First, we define $\bar{\boldsymbol{h}}$. Let $\boldsymbol{g_n}$ be a random function in F_n. Clearly,

$$f = \boldsymbol{g_n} \oplus (\boldsymbol{g_n} \oplus f) = (\boldsymbol{g_n} \wedge (\boldsymbol{g_n} \oplus f \oplus 1)) \vee ((\boldsymbol{g_n} \oplus 1) \wedge (\boldsymbol{g_n} \oplus f)),$$

for any function in the support of $\boldsymbol{g_n}$. We define the random circuit

$$\boldsymbol{C} \stackrel{\text{def}}{=} (C_{\boldsymbol{g_n}} \wedge C_{\boldsymbol{g_n} \oplus f \oplus 1}) \vee (C_{\boldsymbol{g_n} \oplus 1} \wedge C_{\boldsymbol{g_n} \oplus f}). \tag{3}$$

Since each circuit C_h computes the function h, it follows that $[\![\boldsymbol{C}]\!]$ computes f with probability 1. We let $\bar{\boldsymbol{h}} \stackrel{\text{def}}{=} [\![\bar{\boldsymbol{C}}]\!]$ be the function in $\mathcal{M}$ computed by the $\mathcal{M}$-circuit $\bar{\boldsymbol{C}}$. Observe that all circuits in the support of $\boldsymbol{C}$ share the same directed acyclic graph, with variations only on the labels corresponding to the constants 0 and 1 (i.e., the base case). This completes our discussion of Part (i).

Now fix an input $v \in f^{-1}(1)$ and the circuit D_g in the support of $\boldsymbol{C}$ that is obtained from a function g in the support of $\boldsymbol{g_n}$. Following the proof of Proposition 1, we say that a gate e of D_g has a plus-error on v if $e(v) = 1$ and $\bar{e}(v) = 0$, where $\bar{e}$ is the function computed at the corresponding gate of the $\mathcal{M}$-circuit $\overline{D_g}$. Recall that D_g computes f, which implies that $D_g(v) = 1$, and that errors do not occur at locations corresponding to the input literals of $\overline{D_g}$. Therefore, if $\overline{D_g}(v) = 0$ there is some internal $\{\vee, \wedge\}$-gate e where a plus-error must occur but no plus-error occur at the gates feeding e. In other words, e is a distinguished gate location (with respect to v). The importance of distinguished gates is that they can fix a mistake on v, as in the proof of Proposition 1.

Recall that $D_g = (C_g \wedge C_{g \oplus f \oplus 1}) \vee (C_{g \oplus 1} \wedge C_{g \oplus f})$. Consider for each such sub-circuit the leaf addressed by v in the natural way, and marked by a constant in $\{0, 1\}$ (it corresponds to C_{g^v} in Equation 2, where we go left or right according to the bits of v). Each such leaf induces a unique path to the root node of D_g. We say that a gate location of D_g is in S_v if it is an $\vee$-gate or $\wedge$-gate appearing in one of the four paths from these leaves to the root of D_g. Consequently, there are $3 + 4 \cdot 3n$ gate locations in S_v. Note that the definition of S_v does not depend on g or on D_g, only on the underlying graph of the circuit, which we denote by $G = (V, E)$.

Claim 2. *If $\overline{D_g}(v) = 0$ then there is some gate location in S_v corresponding to operation $\overline{e_i} \star \overline{e_2}$ over sub-circuits $\overline{e_1}$ and $\overline{e_2}$ (with respect to the circuits D_g and $\overline{D_g}$) such that $v \in \delta^+_\star(\overline{e_1}, \overline{e_2})$.*

We argue as follows. If one of the three gates appearing in Equation 3 are distinguished gates for D_g and $\overline{D_g}$ we are done, since these are in S_v by definition. Otherwise, there is a plus-error for v on the top gate of one of the four sub-circuits, say, in C_g. The crucial observation is that on the path from the leaf to the top gate of C_g marking the corresponding gate locations in S_v, there must be a distinguished gate location somewhere along the path that produces a plus-error (with respect to C_g and $\overline{C_g}$), but whose preceding input gates cause no plus-error on v. This is because there is no plus-error occurring on a leaf of D_g, and no plus-error can occur on the $\wedge$-gate of the sub-circuits coming from Equation 2 when we take the wrong path with respect to v: the literal x_d^ε evaluates to 0 on v when we deviate from the correct path, and so does the $\wedge$-gate above this literal. This completes our discussion of Part (ii).

We proceed with Part (iii), starting with the definition of $\boldsymbol{\delta}^+$ (the random variables appearing below are independent of the construction of $\bar{\boldsymbol{h}}$). We then explain why this definition satisfies Equation 1 for each choice of $v \in f^{-1}(1)$.

Let $\boldsymbol{\ell}_\gamma$ be uniformly and independently sampled from F_γ, for each choice of $\gamma \in \{1,\ldots,n\}$. Pick independently and uniformly at random $\boldsymbol{\gamma} \in \{1,\ldots,n,\oplus\}$ and $\boldsymbol{t} \in \{0,1,\vee\}$. Now let

$$\boldsymbol{\delta}^+ \stackrel{\text{def}}{=} \begin{cases} \delta^+_\wedge([\![\overline{C_{\ell^{\boldsymbol{t}}_{\boldsymbol{\gamma}}}}]\!], x^{\boldsymbol{t}}_{\boldsymbol{\gamma}}) & \text{if } \boldsymbol{\gamma} \in \{1,\ldots,n\} \text{ and } \boldsymbol{t} \in \{0,1\}, \\ \delta^+_\vee([\![\overline{C_{\ell^0_\gamma} \wedge x^0_{\boldsymbol{\gamma}}}]\!], [\![\overline{C_{\ell^1_\gamma} \wedge x^1_{\boldsymbol{\gamma}}}]\!]) & \text{if } \boldsymbol{\gamma} \in \{1,\ldots,n\} \text{ and } \boldsymbol{t} = \vee, \\ \delta^+_\wedge([\![\overline{C_{\boldsymbol{\ell_n} \oplus \boldsymbol{t}}}]\!], [\![\overline{C_{\boldsymbol{\ell_n} \oplus \boldsymbol{t} \oplus f \oplus 1}}]\!]) & \text{if } \boldsymbol{\gamma} = \oplus \text{ and } \boldsymbol{t} \in \{0,1\}, \\ \delta^+_\vee([\![\overline{C_{\boldsymbol{\ell_n}} \wedge C_{\boldsymbol{\ell_n} \oplus f \oplus 1}}]\!], [\![\overline{C_{\boldsymbol{\ell_n} \oplus 1} \wedge C_{\boldsymbol{\ell_n} \oplus f}}]\!]) & \text{if } \boldsymbol{\gamma} = \oplus \text{ and } \boldsymbol{t} = \vee. \end{cases}$$

We say that the pair $(\boldsymbol{\gamma}, \boldsymbol{t})$ defines the $\boldsymbol{type}$ of $\boldsymbol{\delta}^+$. Note that the random variable $\boldsymbol{type}$ has support size $3n+3$. For convenience, let $W \stackrel{\text{def}}{=} \mathsf{Support}(\boldsymbol{type})$.

We consider a fixed map $\psi \colon V(G) \setminus \{\text{leaves of } G\} \to W$, where G is the underlying directed acyclic graph of $\boldsymbol{C}$. This is done in the natural way, and won't be formalized here (observe that $\boldsymbol{\delta}^+$ closely follows the structure of $\boldsymbol{C}$). In particular, the map ψ labels by a type each gate location in $S_v \subseteq V(G)$, for an arbitrary $v \in f^{-1}(1)$.

Fix $v \in f^{-1}(1)$. Given a function $\boldsymbol{g_n} \in F_n$ which induces circuits $\boldsymbol{C}$ and $\overline{\boldsymbol{C}}$ and a function $\bar{\boldsymbol{h}}$ such that $\bar{\boldsymbol{h}}(v) = 0$, we consider a decomposition $\Pr[\bar{\boldsymbol{h}}(v) = 0] = \sum_{w \in W} p^v_w$ where each value p^v_w is defined as follows. Say that $\boldsymbol{g_n} = g$. We fix for this choice of v and g a canonical distinguished gate location in $S_v \subseteq V(G)$, which is guaranteed to exist by Claim 2. This provides a corresponding type given by the map ψ. Then each value p^v_w collects the probability mass of the type w over those functions g in the support of $\boldsymbol{C}$ that yield a function $\bar{h}$ such that $\Pr[\bar{h}(v) = 0]$.

Under these definitions, it is possible to show that:

$$\Pr[\boldsymbol{\delta}^+(v) = 1] = \sum_{w \in W} \frac{1}{3n+3} \cdot \Pr[\boldsymbol{\delta}^+(v) = 1 \mid \boldsymbol{type} = w] \geq \frac{1}{3n+3} \sum_w \frac{p^v_w}{4} = \frac{\Pr[\bar{\boldsymbol{h}}(v) = 0]}{12n+12},$$

where we leave the formal verification that $\Pr[\boldsymbol{\delta}^+(v) = 1 \mid \boldsymbol{type} = w] \geq p^v_w/4$ for each type $w \in W$ to the reader. The crucial point is that each sub-circuit of the four main sub-circuits in Equation 3 is a random circuit, in the sense that each function $\boldsymbol{g_n}$, $\boldsymbol{g_n} \oplus f \oplus 1$, $\boldsymbol{g_n} \oplus 1$, and $\boldsymbol{g_n} \oplus f$ appearing in that equation is uniformly random over F_n, and so are the 0/1-values labeling the leaves of the corresponding subgraphs of G. This justifies the definition of $\boldsymbol{\delta}^+$, and completes our description of Part (iii). □

Proof of Lemma 2. The proof is the same, except for the following change of perspective. We view each circuit $\boldsymbol{C}$ as a circuit that computes the function $f_N \in F_N$, instead of a function in F_n. Nevertheless, $\boldsymbol{C}$ as a (random) syntactical object will be labelled only by literals produced from $x_1,\ldots,x_n$ (recall that f_N only depends on the initial n inputs). The size of $\boldsymbol{C}$ is still exponential in n (instead of N), and $|S_v| = O(n)$ as before. The rest of the argument is unaffected, with f_N and $\alpha \in f_N^{-1}(1)$ substituted for f and v (respectively) whenever necessary. □

Proof of Lemma 3. The statement of this lemma uses a different terminology when compared to the statement of Lemma 2, but the argument is essentially the same. This is because the fusion method is a particular case of the generalized approximation method, as explained in Section 2.2. The change of terminology is motivated by Theorem 3. The distribution $(\boldsymbol{A}, \boldsymbol{B})$ is simply the distribution induced by $\boldsymbol{\delta}^+$, which we discuss next.

First, observe that the second case of Lemma 2 is not necessary here, thanks to Theorem 3 (ii) and the one-sided formulation described in Section 2.4. Thus in this proof we only need

to consider $\bar{\boldsymbol{h}}$ and $\boldsymbol{\delta}^{+}$, as covered in the proof of Lemma 1 and its generalization described above (Lemma 2).

We proceed as in the proof of Lemma 2, defining a random circuit $\boldsymbol{C}$, the sets S_{α}, etc. There are only two modifications in the argument. First, observe that $\Pr[\bar{\boldsymbol{h}}(\alpha) = 0] = 1$, where $\bar{\boldsymbol{h}}$ is the function computed by the $\mathcal{M}(\mathfrak{F}')$-circuit $\overline{\boldsymbol{C}}$. Indeed, it follows that $\bar{\boldsymbol{h}} = \bar{\boldsymbol{f}}$ using the definitions of $\boldsymbol{C}$ and of the operations $\bar{\wedge}$ and $\bar{\vee}$ in $\mathcal{M}(\mathfrak{F}')$. Furthermore, the only inputs $\alpha \in f_N^{-1}(1)$ that need to be considered in Lemma 3 are of the form $(v(\mathcal{F}), \mathcal{F})$, following the discussion in Section 2.4. On such inputs, by definition, we have $\bar{f}(v(\mathcal{F}), \mathcal{F}) = \mathcal{F}(U_f) = \mathcal{F}(\emptyset) = 0$.

Finally, observe that in Lemma 3 we use $8n+8$ instead of the term $12n+12$ which appears in Lemma 2. This is because only $\wedge$-gates are needed to cover the errors in the extended legitimate models employed in the fusion method (Theorem 3), and the number of types of gates appearing in the definition of $\boldsymbol{\delta}^{+}$ can be reduced accordingly. □

Department of Computer Science, University of Oxford, Oxford OX1 3QD
E-mail: igor.carboni.oliveira@cs.ox.ac.uk