Shrinkage of Decision Lists and DNF Formulas

Benjamin Rossman
Duke University, Durham, NC, USA
benjamin.rossman@duke.edu

Abstract

We establish nearly tight bounds on the expected shrinkage of decision lists and DNF formulas under the p-random restriction $\boldsymbol{R}_p$ for all values of $p \in [0,1]$. For a function f with domain $\{0,1\}^n$, let $\mathsf{DL}(f)$ denote the minimum size of a decision list that computes f. We show that

$$\mathbb{E}[\ \mathsf{DL}(f{\restriction}\boldsymbol{R}_p)\ ] \leq \mathsf{DL}(f)^{\log_{2/(1-p)}(\frac{1+p}{1-p})}.$$

For example, this bound is $\sqrt{\mathsf{DL}(f)}$ when $p = \sqrt{5}-2 \approx 0.24$. For Boolean functions f, we obtain the same shrinkage bound with respect to DNF formula size plus 1 (i.e., replacing $\mathsf{DL}(\cdot)$ with $\mathsf{DNF}(\cdot)+1$ on both sides of the inequality).

2012 ACM Subject Classification Theory of computation → Circuit complexity

Keywords and phrases shrinkage, decision lists, DNF formulas

Digital Object Identifier 10.4230/LIPIcs.ITCS.2021.70

Acknowledgements I am grateful to the anonymous referees of ITCS 2021 for their valuable comments and to the authors of [17], Shachar Lovett, Kewen Wu and Jiapeng Zhang, for stimulating conversations related to this work.

1 Introduction

Random restrictions are a powerful tool in circuit complexity and the analysis of Boolean functions. A *restriction* is a partial assignment to the input bits of a function f on the hypercube $\{0,1\}^n$. For a parameter $p \in [0,1]$, the *p-random restriction* $\boldsymbol{R}_p$ independently leaves each input bit free with probability p and otherwise assigns it to 0 or 1 with equal probability. We denote by $f{\restriction}\boldsymbol{R}_p$ the function obtained from f by restricting its inputs to the subcube of $\{0,1\}^n$ that correspond to $\boldsymbol{R}_p$.

Random restrictions are known to reduce the complexity of functions in simple models of computations, such as decision trees (DT), decision lists (DL), DNF formulas (DNF), and DeMorgan formulas ($\mathcal{L}$); the symbols in parentheses are notation for the corresponding size measures (see Section 2 for definitions). With respect to DeMorgan formula leaf-size $\mathcal{L}$, it is easy to see that $\mathcal{L}(f{\restriction}\boldsymbol{R}_p)$ has expectation at most $p \cdot \mathcal{L}(f)$. (This follows by linearity of expectation from the observation that each input literal in a minimal formula for f is eliminated by $\boldsymbol{R}_p$ with probability p.) Subbotovskaya [25] was the first to show that the expected shrinkage factor is in fact significantly smaller than p (she showed an upper bound $O(p^{3/2})$ for $p \geq 1/\mathcal{L}(f)^{2/3}$). A subsequent line of results [1, 14, 19, 11, 26], culminating in an $p^{2-o(1)}$ bound of Håstad [11] and a low-order improvement by Tal [26], eventually established an asymptotically tight bound:

▶ **Theorem 1** (Shrinkage of DeMorgan formulas [26])**.** *For all Boolean functions f,*

$$\mathbb{E}[\ \mathcal{L}(f{\restriction}\boldsymbol{R}_p)\ ] = O(\,p^2\mathcal{L}(f) + p\sqrt{\mathcal{L}(f)}\,).$$

The constant 2 in the exponent p in Theorem 1 is known as the "shrinkage exponent" of DeMorgan formulas. Shrinkage under $\boldsymbol{R}_p$ has also been studied for restricted types of formulas, namely read-once, monotone, and bounded-depth (AC^0). It was shown in [5, 13] that read-once formulas have shrinkage exponent $\log_{\sqrt{5}-1}(2) \approx 3.27$. The shrinkage exponent of monotone formulas is between 2 and $\log_{\sqrt{5}-1}(2)$ and conjectured to equal the latter; determining the exact constant is a longstanding question (Open Problem 4). In the AC^0 setting (bounded-depth formulas with unbounded AND and OR gates), it is known that depth-d formulas with fan-in m shrink to expected size $O(1)$ under $\boldsymbol{R}_p$ when p is $O(1/\log m)^{d-1}$ [22]. However, it is open to determine the shrinkage rate for larger p, particularly in the "mild random restriction" regime where p is $\Omega(1)$ or $1 - o(1)$ (Open Question 2).

The results of this paper give nearly tight bounds on the shrinkage under $\boldsymbol{R}_p$ of depth-2 formulas (also known as DNF and CNF formulas), as well as the more general computational model of *decision lists*. Before stating our main result, it is instructive to first consider shrinkage in the simpler model of *decision trees*. For a function f on the hypercube (with domain $\{0,1\}^n$ and arbitrary range), we denote by $\mathsf{DT}(f)$ the minimum number of leaves (i.e., output nodes) in a decision tree that computes f. The following bound is shown by straightforward induction on $\mathsf{DT}(f)$. (I believe this bound is probably folklore, but could not find a reference so have included the short proof in Section 3.1.)

▶ **Theorem 2** (Shrinkage of decision trees). *For all functions f on the hypercube,*

$$\mathbb{E}[\ \mathsf{DT}(f{\restriction}\boldsymbol{R}_p)\ ] \leq \mathsf{DT}(f)^{\log_2(1+p)}.$$

This bound holds with equality when f is a parity function.

Decision lists are a natural computational model that has been studied in many contexts [3, 4, 16, 9, 21]. A *decision list of size m* is a sequence $L = ((C_1, b_1), \ldots, (C_m, b_m))$ where $b_1, \ldots, b_m$ are arbitrary output values and $C_1, \ldots, C_m$ are conjunctive clauses (ANDs of literals) such that $C_1 \vee \cdots \vee C_m$ is a tautology.[1] L computes a function on the hypercube as follows: on input $x \in \{0,1\}^n$, the output is b_i for the first index $i \in [m]$ such that $C_i(x)$ is satisfied. We denote by $\mathsf{DL}(f)$ the minimum size of a decision list that computes f.

Decision lists are a generalization decision trees: every decision tree is equivalent to a decision list of the same size, and thus $\mathsf{DL}(f) \leq \mathsf{DT}(f)$ for all functions f on the hypercube.[2] Boolean decision lists, in which $b_1, \ldots, b_m \in \{0,1\}$, are moreover a generalization of both DNF and CNF formulas. In particular, *DNF formulas* are the special case where $b_1 = \cdots = b_{m-1} = 1$ and $b_m = 0$. Following custom, we count the *size* of a DNF formula as $m-1$ instead of m, and thus $\mathsf{DL}(f) \leq \mathsf{DNF}(f) + 1$ for all Boolean functions f.

Despite decision lists and DNF/CNF formulas being more complex computational models than decision trees, our main result shows that they shrink at a similar rate under $\boldsymbol{R}_p$.

---

[1] In other words, every input $x \in \{0,1\}^n$ satisfies at least one of $C_1, \ldots, C_m$. Without loss of generality, C_m may be chosen as the empty (always true) conjunctive clause $\top$. We allow $C_1 \vee \cdots \vee C_m$ to be an arbitrary tautology in order to more naturally define the class of *orthogonal* decision lists later on in Section 3.3.

[2] The name "decision list" elsewhere commonly refers to (what we call) width-1 decision trees, in which each clause is a single literal (i.e., an input variable x_i or its negation $\overline{x_i}$). Whereas unbounded-width decision lists are a generalization decision trees, width-1 decision lists are instead a special case.

▶ **Theorem 3** (Shrinkage of decision lists and DNF formulas). *For all functions f on the hypercube,*

$$\mathbb{E}[\ \mathsf{DL}(f{\restriction}\boldsymbol{R}_p)\ ] \le \mathsf{DL}(f)^{\gamma(p)} \qquad \textit{where} \qquad \gamma(p) := \log_{\frac{2}{1-p}}(\tfrac{1+p}{1-p}).$$

If f is Boolean, then also $\mathbb{E}[\ \mathsf{DNF}(f{\restriction}\boldsymbol{R}_p) + 1\ ] \le (\mathsf{DNF}(f) + 1)^{\gamma(p)}$ (and similarly for $\mathsf{CNF}(\cdot) + 1$).

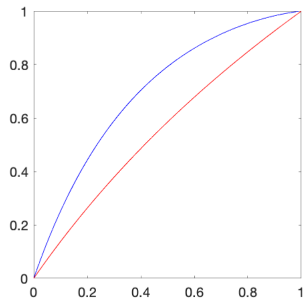

Figure 1 Plots of $\gamma(p) := \log_{\frac{2}{1-p}}(\frac{1+p}{1-p})$ (blue) and $\log_2(1+p)$ (red).

Note that $\gamma : [0,1] \to [0,1]$ is an increasing function with $\gamma(0) = 0$ and $\gamma(1) = 1$ (see Figure 1). The bound of Theorem 3 is thus nontrivial for all values of $p \in (0,1)$. This bound is moreover close to optimal: $\log_2(1+p)$ is a lower bound on the best possible function $\gamma(p)$ (Section 3.4). As corollaries, we obtain additional bounds $\mathsf{ODL}(f)^{\gamma(p)}$ and $\mathsf{wODL}(f)^{\gamma(p)}$ on the shrinkage of orthogonal and weakly orthogonal decision lists (Corollary 14), as well as $(\mathcal{L}_2(f) + 1)^{\gamma(2p)}$ for depth-2 formula leaf-size (Corollary 18).

Theorem 3 yields the following bounds for particular settings of p in terms of $m = \mathsf{DL}(f)$:

$$\mathbb{E}[\ \mathsf{DL}(f{\restriction}\boldsymbol{R}_p)\ ] \le \begin{cases} 2 & \text{for } p = O(\frac{1}{\log m}), \\ \sqrt{m} & \text{for } p = \sqrt{5} - 2 \approx 0.24, \\ m/2 & \text{for } p = 1 - O(\frac{\log\log m}{\log m}), \\ m-1 & \text{for } p = 1 - O(\frac{\log m}{m}). \end{cases}$$

For small $p = O(1/\log m)$, a variant of Håstad's Switching Lemma (discussed below) actually implies a stronger inequality $\mathbb{E}[\ \mathsf{DT}(f{\restriction}\boldsymbol{R}_p)\ ] \le 2$ with DT in place of DL (Corollary 6). Theorem 3 is mainly interesting for larger values of p. In particular, the "mild random restriction" regime when p is $\Omega(1)$ or $1 - o(1)$ has important applications in pseudorandomness [8, 20], DNF sparsification [7, 17] and hypercontractivity [18].

1.1 Switching lemmas and size measures vs. width/depth measures

We have so far discussed the shrinkage of various complexity measures under the p-random restriction $\boldsymbol{R}_p$. The switching lemmas stated below can be viewed as apples-to-oranges shrinkage results that bound one complexity measure on $f{\restriction}\boldsymbol{R}_p$ in terms of another complexity measure on f. Here there is a useful distinction between "size measures" DT, DL, DNF and their corresponding "width/depth measures", denoted by $\mathsf{DT}_{\mathrm{depth}}$, $\mathsf{DL}_{\mathrm{width}}$, $\mathsf{DNF}_{\mathrm{width}}$. Width/depth measures are typically related to the logarithm of size measures: functions with size complexity m are approximable by (or in some cases equivalent to) functions with width/depth complexity $O(\log m)$. Håstad's Switching Lemma [10] gives a tail bound on the decision tree size of $f{\restriction}\boldsymbol{R}_p$ in terms of the decision list width of f.[3]

▶ **Theorem 4** (Switching Lemma [10]). *For all functions f on the hypercube and $t \in \mathbb{N}$,*

$$\mathbb{P}[\ \mathsf{DT}_{\text{depth}}(f{\restriction}\boldsymbol{R}_p) \geq t\ ] \leq O(p \cdot \mathsf{DL}_{\text{width}}(f))^t.$$

A variant of the Switching Lemma with $\log \mathsf{DL}(f)$ in place of $\mathsf{DL}_{\text{width}}(f)$ was proved in [22].

▶ **Theorem 5** (Switching Lemma in terms of decision list size [22]). *For every function f on the hypercube and $t \in \mathbb{N}$,*

$$\mathbb{P}[\ \mathsf{DT}_{\text{depth}}(f{\restriction}\boldsymbol{R}_p) \geq t\ ] \leq O(p \cdot \log \mathsf{DL}(f))^t.$$

We remark that Theorem 5 follows directly from Theorem 4 for $t \leq O(\log \mathsf{DL}(f))$ (by the standard width reduction argument), but not for larger t. Obtaining a tail bound for all $t \in \mathbb{N}$ is essentially to the following:

▶ **Corollary 6** (Decision tree size of decision lists). *For all functions f on $\{0,1\}^n$,*

$$\mathbb{E}[\ \mathsf{DT}(f{\restriction}\boldsymbol{R}_p)\ ] \leq 2 \quad \textit{and} \quad \mathsf{DT}(f) \leq O(2^{(1-p)n}) \quad \textit{where} \quad p = O(1/\log \mathsf{DL}(f)).$$

As previously mentioned, Corollary 6 strengthen the bound $\mathbb{E}[\ \mathsf{DL}(f{\restriction}\boldsymbol{R}_p)\ ] \leq 2$ for $p = O(1/\log \mathsf{DL}(f))$ that follows from Theorem 3 (albeit for p that is a constant factor smaller). However, note that Corollary 6 is trivial for p above $\Omega(1/\log \mathsf{DL}(f))$. A different switching lemma for large p (even $1 - o(1)$) in terms of $\mathsf{DNF}_{\text{width}}(f)$ was introduced by Segerlind, Buss and Impagliazzo [24] and quantitatively improved by Razborov [20]. It is unclear if these switching lemmas for "mild random restriction" have analogues in terms of $\log \mathsf{DL}(f)$; if so, that might entail a shrinkage bound for DL that is nontrivial for all $p \in (0,1)$, although potentially weaker than Theorem 3.

Our proof of Theorem 3 involves an application of Jensen's inequality with respect to a certain carefully defined probability distribution on the set of clauses in a decision list L. This distribution is related to (but not identical to) the distribution of the first satisfied clause of L under a uniform random input. A similar convexity argument appears in the proof of Theorem 5 in [22]. A second key idea, the notion of "useful indices" of L under a restriction ϱ, comes from a recent paper of Lovett, Wu and Zhang [17] who proved the following result as the main lemma in establishing tight bound on the sparsification of bounded-width decision lists.

▶ **Theorem 7** (Decision list shrinkage in terms of width [17]). *For every function f on the hypercube,*

$$\mathbb{E}[\ \mathsf{DL}(f{\restriction}\boldsymbol{R}_p)\ ] \leq \left(\frac{4}{1-p}\right)^{\mathsf{DL}_{\text{width}}(f)}.$$

Note that our main result, Theorem 3, stands in relation to Theorem 7 just as Theorem 5 does to Theorem 4: in both cases we are essentially replacing $\mathsf{DL}_{\text{width}}(f)$ with $\log \mathsf{DL}(f)$.

3 In its application to AC^0 circuit lower bounds, Theorem 4 is usually stated (more narrowly) in the form

$$\mathbb{P}[\ \mathsf{CNF}_{\text{width}}(f{\restriction}\boldsymbol{R}_p) \geq t\ ] \leq O(p \cdot \mathsf{DNF}_{\text{width}}(f))^t$$

for Boolean functions f. The name "Switching Lemma" refers to the conversion of a DNF formula to a CNF formula. The more general bound stated in Theorem 4 is implicit in proofs of [10].

1.2 Other related work

There are different ways to quantify the effect of random restrictions on complexity measures. Instead of bounding expectation, one may show that shrinkage occurs with high probability. For DeMorgan formulas, high probability shrinkage results were shown in [23, 15]. Shrinkage results and switching lemmas have also been studied for random restrictions other than $\boldsymbol{R}_p$ (see [2]). Very interesting recent work of Filmus, Meir and Tal [6] extends the technique of Håstad [11] to obtain $p^{2-o(1)}$ factor shrinkage bounds for DeMorgan formulas under a family of pseudorandom projections that generalize $\boldsymbol{R}_p$.

2 Preliminaries

Throughout this paper, p is an arbitrary parameter in $[0, 1]$. All inequalities involving p hold for all values in $[0, 1]$. We often use the special case of Jensen's inequality $\mathbb{E}[\, X^c \,] \leq \mathbb{E}[\, X \,]^c$ where X is a nonnegative random variable and $c \in [0, 1]$ (in particular, when c is $\log_2(1 + p)$ or $\gamma(p)$). We write $\mathbb{N}$ for the natural numbers $\{0, 1, 2, \ldots\}$, and for $m \in \mathbb{N}$, we write $[m]$ for $\{1, \ldots, m\}$.

2.1 Functions and restrictions on the hypercube

Function on the hypercube refers to any function with domain $\{0, 1\}^n$ where n is a positive integer. A *Boolean function* is a function on the hypercube with codomain $\{0, 1\}$. (The parameter n plays no role in most results in this paper, so we suppress its mention whenever possible.)

A *restriction* is a partial assignment of Boolean variables $x_1, \ldots, x_n$ to values 0 and 1; this is formally defined as a function $\varrho : \{1, \ldots, n\} \to \{0, 1, *\}$ where $\varrho(i) = *$ signifies that x_i is left free by ϱ. We denote by $\mathrm{Stars}(\varrho) \subseteq [n]$ the set of free variables under ϱ. For a function f on the hypercube $\{0, 1\}^n$ and a restriction ϱ, we denote by $f{\restriction}\varrho$ the restricted function on the subcube $\{0, 1\}^{\mathrm{Stars}(\varrho)}$ defined in the obvious way: $(f{\restriction}\varrho)(y) = f(x)$ where $x \in \{0, 1\}^n$ is the input with $x_i = y_i$ if $i \in \mathrm{Stars}(\varrho)$ and $x_i = \varrho(i)$ otherwise.

For $p \in [0, 1]$, the *p-random restriction* $\boldsymbol{R}_p$ is the random restriction that independently leaves each variable x_i free with probability p and otherwise sets x_i to 0 or 1 with equal probability. Thus, for any particular restriction ϱ, we have $\mathbb{P}[\, \boldsymbol{R}_p = \varrho \,] = p^{|\mathrm{Stars}(\varrho)|}((1 - p)/2)^{n-|\mathrm{Stars}(\varrho)|}$.

2.2 Complexity measures DL, DT, DNF, CNF and their width/depth versions

▶ **Definition 8** (DNF formulas)**.** We first define literals, conjunctive clauses, and DNF formulas over n variables.

- A *literal* is a Boolean variable x_i or negated Boolean variable $\overline{x_i}$ where $i \in \{1, \ldots, n\}$.
- A *conjunctive clause* (a.k.a. *term*) is an expression C of the form $\ell_1 \wedge \cdots \wedge \ell_w$ where $\ell_1, \ldots, \ell_w$ are literals on disjoint variables. The parameter w is the *width* of C; this may be any nonnegative integer. The conjunctive clause of width zero is denoted by $\top$.
- A *DNF formula* is an expression F of the form $C_1 \vee \cdots \vee C_m$ where $C_1, \ldots, C_m$ are conjunctive clauses. The parameter m is the *size* of F; this may be any nonnegative integer. The DNF formula of size 0 is denoted by $\perp$. The *width* of F is defined as the maximum width of any C_i.
- *CNF formulas* are defined dually (with the roles of $\vee$ and $\wedge$ exchanged).

Every literal, conjunctive clause, and DNF formula computes a Boolean function $\{0,1\}^n \to \{0,1\}$ in the usual way.

- A DNF formula F is a *tautology* if it computes the identically 1 function. Note that any DNF formula that includes the empty conjunctive clause $\top$ is a tautology.

▶ **Definition 9** (Decision lists).

- A *decision list* is an expression L of the form $((C_1, b_1), \dots, (C_m, b_m))$ where $b_1, \dots, b_m$ are arbitrary output values (not necessarily Boolean) and $C_1, \dots, C_m$ are conjunctive clauses such that $C_1 \vee \cdots \vee C_m$ is a tautology. The parameter m is the *size* of L; this may be any positive integer. The *width* of C is defined as the maximum width of any C_i.

A decision list L computes a function $\{0,1\}^n \to \{b_1, \dots, b_m\}$ as follows: on input x, the output is b_ℓ where $i \in [m]$ is the minimum index such that $C_i(x) = 1$. (Note that the final clause C_m may be replaced by $\top$ without changing the function computed by L.)

▶ **Definition 10** (Decision trees).

- A *decision tree* is a rooted binary tree T in which each leaf is labeled by an output value (not necessarily Boolean) and each non-leaf node is labeled by a variable x_i, with the edges to its two children labeled "$x_i = 0$" and "$x_i = 1$". The *size* of T is the number of leaves; this may be any positive integer. The *depth* of T is the maximum number of non-leaf nodes on any root-to-leaf branch; this may be any nonnegative integer.

▶ **Definition 11** (Associated complexity measures). For a function f with domain $\{0,1\}^n$ (and arbitrary codomain), let

$$\begin{aligned}\mathsf{DT}(f) &:= \text{minimum size of a decision tree that computes } f,\\ \mathsf{DL}(f) &:= \text{minimum size of a decision list that computes } f,\end{aligned}$$

When f is Boolean, we additionally define

$$\begin{aligned}\mathsf{DNF}(f) &:= \text{minimum size of a DNF formula that computes } f,\\ \mathsf{CNF}(f) &:= \text{minimum size of a CNF formula that computes } f.\end{aligned}$$

For constant functions $\underline{0}$ and $\underline{1}$, note that $\mathsf{DNF}(\underline{0}) = 0$ and $\mathsf{DNF}(\underline{1}) = 1$ according to our definition, since $\underline{0}$ is computed by the empty DNF formula, while $\underline{1}$ is computed by the DNF formula with a single empty clause. Also note that $\mathsf{CNF}(f) = \mathsf{DNF}(\neg f)$.

Each of the above size measures has a corresponding width/depth measure. These are denoted by

$$\mathsf{DT}_{\text{depth}}(f), \qquad \mathsf{DL}_{\text{width}}(f), \qquad \mathsf{DNF}_{\text{width}}(f), \qquad \mathsf{CNF}_{\text{width}}(f).$$

▶ **Proposition 12** (see [3, 16]). *These size measures satisfy the following inequalities for all Boolean functions:*

$$1 \le \mathsf{DL} \le \left\{ \begin{array}{c} \mathsf{DNF} + 1 \\ \mathsf{CNF} + 1 \end{array} \right\} \le \mathsf{DNF} + \mathsf{CNF} \le \mathsf{DT}.$$

The corresponding width/depth measures satisfy:

$$0 \le \mathsf{DL}_{\text{width}} \le \left\{ \begin{array}{c} \mathsf{DNF}_{\text{width}} \\ \mathsf{CNF}_{\text{width}} \\ \lceil \log_2(\mathsf{DT}) \rceil \end{array} \right\} \le \mathsf{DT}_{\text{depth}} \le \mathsf{DNF}_{\text{width}} \cdot \mathsf{CNF}_{\text{width}}.$$

The above inequalities that involve decision trees and decision lists also apply to non-Boolean functions on the hypercube.

We introduce additional computational models later on: (weakly) orthogonal decision lists in Section 3.3 and AC^0 formulas in Section 4.

3 Shrinkage of decision trees and decision lists

We prove Theorems 2 and 3 in Sections 3.1 and 3. We then discuss extensions of our shrinkage bound to (weakly) orthogonal decision lists in Section 3.3 and tightness of the bounds Section 3.4.

3.1 Shrinkage of decision trees

Proof of Theorem 2. Let T be a decision tree (with arbitrary output values). We must show that

$$\mathbb{E}[\ \mathrm{size}(T{\restriction}\boldsymbol{R}_p)\ ] \le \mathrm{size}(T)^{\log_2(1+p)}.$$

We argue by induction of the size of T. The inequality is trivial in the base case that T has size 1.

Assume T has size $m \ge 2$. Then T has the form "If $x_i = 0$ then T_0 else T_1" where T_0, T_1 are decision trees of size $m_0, m_1 \ge 1$ with $m_0 + m_1 = m$. Without loss of generality, T_0 and T_1 never query x_i. We have

$$\begin{aligned}
\mathbb{E}[\ \mathrm{size}(T{\restriction}\boldsymbol{R}_p)\ ] &= p\,\mathbb{E}\left[\ \mathrm{size}(T{\restriction}\boldsymbol{R}_p)\ \middle|\ \boldsymbol{R}_p(x_i) = *\ \right] \\
&\quad + \frac{1-p}{2}\Big(\ \mathbb{E}\left[\ \mathrm{size}(T_0{\restriction}\boldsymbol{R}_p)\ \middle|\ \boldsymbol{R}_p(x_i) = 0\ \right] + \mathbb{E}\left[\ \mathrm{size}(T_1{\restriction}\boldsymbol{R}_p)\ \middle|\ \boldsymbol{R}_p(x_i) = 1\ \right]\Big) \\
&= \frac{1+p}{2}\Big(\ \mathbb{E}[\ \mathrm{size}(T_0{\restriction}\boldsymbol{R}_p)\ ] + \mathbb{E}[\ \mathrm{size}(T_1{\restriction}\boldsymbol{R}_p)\ ]\Big) \\
&\le \frac{1+p}{2}\Big((m_0)^{\log_2(1+p)} + (m_1)^{\log_2(1+p)}\Big) && \text{(induction hypothesis)} \\
&\le (1+p)\Big(\frac{m}{2}\Big)^{\log_2(1+p)} && \text{(Jensen's inequality)} \\
&= m^{\log_2(1+p)}.
\end{aligned}$$

As for tightness of the bound: If f is a parity function $f(x_1, \dots, x_k) = x_1 \oplus \dots \oplus x_k$, then we have $\mathsf{DT}(f) = 2^k$ and

$$\begin{aligned}
\mathbb{E}[\ \mathsf{DT}(f{\restriction}\boldsymbol{R}_p)\ ] = \mathbb{E}[\ 2^{\mathrm{Bin}(k,p)}\ ] &= \sum_{i=0}^{k} 2^i\,\mathbb{P}[\ \mathrm{Bin}(k,p) = i\ ] \\
&= \sum_{i=0}^{k} \binom{k}{i}(2p)^i(1-p)^{k-i} = (1+p)^k = \mathsf{DT}(f)^{\log_2(1+p)}. \blacktriangleleft
\end{aligned}$$

3.2 Shrinkage of decision lists

We now prove our main result on the shrinkage of decision lists and DNF formulas.

Proof of Theorem 3. Let f be any function on the hypercube and let $p \in [0, 1]$. (Note: Neither the hypercube dimension n nor the nature of output values of f play no role in our analysis.)

Let $L = ((C_1, b_1), \ldots, (C_m, b_m))$ be a decision list of minimum size that computes f, that is, with $m = \mathsf{DL}(f)$. For $\ell \in [m]$, let $|C_\ell|$ denote the width of the clause C_ℓ (i.e., the number of literals in C_ℓ). Without loss of generality, we have $|C_1|, \ldots, |C_{m-1}| \geq 1$ and $|C_m| = 0$ (i.e., C_m is the empty clause $\top$).

Following Lovett, Wu and Zhang [17], for a restriction ϱ, we define the set $U(\varrho) \subseteq [m]$ of *useful indices of L under ϱ* by

$$U(\varrho) := \{\ell \in [m] : \exists\, x \text{ consistent with } \varrho \text{ s.t. } C_\ell(x) = 1 \text{ and } C_1(x) = \cdots = C_{\ell-1}(x) = 0\}.$$

If $U(\varrho) = \{\ell_1, \ldots, \ell_t\}$ where $1 \leq \ell_1 < \cdots < \ell_t \leq m$, then the restricted function $f{\restriction}\varrho$ is computed by the decision list $L{\restriction}\varrho$ defined by

$$L{\restriction}\varrho := ((C_{\ell_1}{\restriction}\varrho, b_{\ell_1}), \ldots, (C_{\ell_t}{\restriction}\varrho, b_{\ell_t}))$$

where $C_{\ell_i}{\restriction}\varrho$ is the sub-clause of C_{ℓ_i} on the variables left unrestricted by ϱ. (Note that $C_{\ell_1} \vee \cdots \vee C_{\ell_t}$ is a tautology, so $L{\restriction}\varrho$ is indeed a decision list.) Thus, we have

$$\mathsf{DL}(f{\restriction}\varrho) \leq |U(\varrho)|. \tag{1}$$

For example, suppose $m = 4$ and

$$C_1 = x_1 \wedge x_3, \quad C_2 = \overline{x_1} \wedge x_4, \quad C_3 = x_2 \wedge \overline{x_3}, \quad C_4 = \top.$$

For $\varrho_1 := \{x_1 \mapsto 1\}$ (the restriction fixing x_1 to 1 and leaving other variables free), we have

$$U(\varrho_1) = \{1, 3, 4\}, \qquad L{\restriction}\varrho_1 = ((x_3, b_1), (x_2 \wedge \overline{x_3}, b_3), (\top, b_4)).$$

For $\varrho_2 := \{x_1 \mapsto 1,\, x_2 \mapsto 1\}$, we have

$$U(\varrho_2) = \{1, 3\}, \qquad L{\restriction}\varrho_2 = ((x_3, b_1), (\overline{x_3}, b_3)).$$

In particular, the final clause C_4 is not useful under ϱ_2 (since any input consistent with ϱ_2 satisfies C_1 or C_3).

Now comes a key definition: let $\mu = (\mu_1, \ldots, \mu_m)$ be the probability density vector (defining a probability distribution on $[m]$)

$$\mu_\ell := \mathop{\mathbb{P}}_{\varrho \sim \boldsymbol{R}_p}[\, \max(U(\varrho)) = \ell \text{ and } C_\ell{\restriction}\varrho \equiv 1 \,] \quad \text{for } \ell \in [m-1],$$

$$\mu_m := \mathop{\mathbb{P}}_{\varrho \sim \boldsymbol{R}_p}[\, \max(U(\varrho)) = m \text{ or } C_{\max(U(\varrho))}{\restriction}\varrho \not\equiv 1 \,].$$

Since events $\max(U(\varrho)) = \ell$ are mutually exclusive, clearly we have $\mu_1 + \cdots + \mu_m = 1$.

Note that $\max(U(\varrho)) = \ell$ does not imply $C_\ell{\restriction}\varrho \equiv 1$, that is, μ_ℓ does not necessarily equal $\mathbb{P}_{\varrho \sim \boldsymbol{R}_p}[\, \max(U(\varrho)) = \ell \,]$. This is illustrated by the restriction ϱ_2 in the above example, for which we have $\max(U(\varrho_2)) = 3$, yet $C_3{\restriction}\varrho_2 = \overline{x}_3 \not\equiv 1$. Restrictions ϱ_1 and ϱ_2 both contribute to probability mass μ_4: in the case of ϱ_1, this is because $\max(U(\varrho_1)) = 4$, and in the case of ϱ_2, this is because $C_{\max(U(\varrho_2))}{\restriction}\varrho_2 \not\equiv 1$.

For each $\ell \in [m]$, we have $\mu_\ell \leq \mathbb{P}[\, C_\ell{\restriction}\varrho \equiv 1 \,] = ((1-p)/2)^{|C_\ell|}$ and therefore

$$|C_\ell| \leq \log_{2/(1-p)}(1/\mu_\ell). \tag{2}$$

We require one more definition. For a restriction ϱ and a useful index $\ell \in U(\varrho)$, let $\varrho^{(\ell)}$ be the restriction obtained by augmenting ϱ by the unique satisfying assignment for the clause C_ℓ. That is, $\varrho^{(\ell)}$ fixes a variable x_i to $a \in \{0,1\}$ if, and only if, ϱ fixes x_i to a or $x_i = a$ in the satisfying assignment to C_ℓ.

As in proofs of the Switching Lemma, we will use the fact that

$$\frac{\mathbb{P}[\ \boldsymbol{R}_p = \varrho\ ]}{\mathbb{P}[\ \boldsymbol{R}_p = \varrho^{(\ell)}\ ]} = \left(\frac{2p}{1-p}\right)^{|\mathrm{Stars}(\varrho)\cap\mathrm{Vars}(C_\ell)|} \tag{3}$$

since $\varrho^{(\ell)}$ has exactly $|\mathrm{Stars}(\varrho) \cap \mathrm{Vars}(C_\ell)|$ fewer unrestricted variables ("stars") than ϱ.

As observed in [17], for every $\ell \in U(\varrho)$, we have $U(\varrho^{(\ell)}) = U(\varrho) \cap [\ell]$ and therefore

$$\max(U(\varrho^{(\ell)})) = \ell \quad \text{ and } \quad C_\ell{\restriction}\varrho^{(\ell)} \equiv 1. \tag{4}$$

Thus, $\varrho^{(\ell)}$ contributes to the probability mass μ_ℓ.

As a consequence of (3) and (4), we claim that for all $\ell \in [m]$,

$$\mathop{\mathbb{P}}_{\varrho\sim\boldsymbol{R}_p}[\ \ell \in U(\varrho)\ ] \le \mu_\ell \left(\frac{1+p}{1-p}\right)^{|C_\ell|}. \tag{5}$$

In the case $\ell = m$, this follows from $m \in U(\varrho) \Rightarrow \max(U(\varrho)) = m$. For $\ell \in [m-1]$, this is shown as follows:

$$\begin{aligned}
&\mathop{\mathbb{P}}_{\varrho\sim\boldsymbol{R}_p}[\ \ell \in U(\varrho)\ ]\\
&= \sum_{S\subseteq\mathrm{Vars}(C_\ell)} \mathop{\mathbb{P}}_{\varrho\sim\boldsymbol{R}_p}[\ \ell \in U(\varrho) \text{ and } \mathrm{Stars}(\varrho)\cap\mathrm{Vars}(C_\ell) = S\ ]\\
&\overset{(4)}{\le} \sum_{S\subseteq\mathrm{Vars}(C_\ell)} \mathop{\mathbb{P}}_{\varrho\sim\boldsymbol{R}_p}[\ \ell = \max(U(\varrho^{(\ell)})) \text{ and } C_\ell{\restriction}\varrho^{(\ell)} \equiv 1 \text{ and } \mathrm{Stars}(\varrho)\cap\mathrm{Vars}(C_\ell) = S\ ]\\
&= \sum_{S\subseteq\mathrm{Vars}(C_\ell)} \sum_{\varrho\,:\,\ell=\max(U(\varrho^{(\ell)})) \text{ and } C_\ell\restriction\varrho^{(\ell)}\equiv 1 \text{ and } \mathrm{Stars}(\varrho)\cap\mathrm{Vars}(C_\ell)=S} \mathbb{P}[\ \boldsymbol{R}_p = \varrho\ ]\\
&= \sum_{S\subseteq\mathrm{Vars}(C_\ell)} \sum_{\sigma\,:\,\ell=\max(U(\sigma)) \text{ and } C_\ell\restriction\sigma\equiv 1} \sum_{\varrho\,:\,\varrho^{(\ell)}=\sigma \text{ and } \mathrm{Stars}(\varrho)\cap\mathrm{Vars}(C_\ell)=S} \mathbb{P}[\ \boldsymbol{R}_p = \varrho\ ]\\
&\overset{(3)}{=} \sum_{S\subseteq\mathrm{Vars}(C_\ell)} \sum_{\sigma\,:\,\ell=\max(U(\sigma)) \text{ and } C_\ell\restriction\sigma\equiv 1} \sum_{\varrho\,:\,\varrho^{(\ell)}=\sigma,\ \mathrm{Stars}(\varrho)\cap\mathrm{Vars}(C_\ell)=S} \left(\frac{2p}{1-p}\right)^{|S|} \mathbb{P}[\ \boldsymbol{R}_p = \sigma\ ]\\
&= \sum_{S\subseteq\mathrm{Vars}(C_\ell)} \left(\frac{2p}{1-p}\right)^{|S|} \sum_{\sigma\,:\,\ell=\max(U(\sigma)),\ C_\ell\restriction\sigma\equiv 1} \mathbb{P}[\ \boldsymbol{R}_p = \sigma\ ] \qquad (\varrho \text{ is determined by } \sigma \text{ and } S)\\
&= \mu_\ell \sum_{S\subseteq\mathrm{Vars}(C_\ell)} \left(\frac{2p}{1-p}\right)^{|S|} \qquad (\text{definition of } \mu_\ell)\\
&= \mu_\ell \left(\frac{1+p}{1-p}\right)^{|C_\ell|} \qquad (\text{binomial expansion of } (1+\tfrac{2p}{1-p})^{|C_\ell|}).
\end{aligned}$$

Finally, we obtain the shrinkage bound of Theorem 3 by the following calculation, which uses Jensen's inequality in addition to the above observations:

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\varrho\sim\boldsymbol{R}_p}[\ \mathsf{DL}(f\restriction\varrho)\ ] \overset{(1)}{\leq} \mathop{\mathbb{E}}_{\varrho\sim\boldsymbol{R}_p}[\ |U(\varrho)|\ ] &= \sum_{\ell\in[m]} \mathop{\mathbb{P}}_{\varrho\sim\boldsymbol{R}_p}[\ \ell \in U(\varrho)\ ] \\
&\overset{(5)}{=} \sum_{\ell\in[m]} \mu_\ell \left(\frac{1+p}{1-p}\right)^{|C_\ell|} \\
&\overset{(2)}{\leq} \sum_{\ell\in[m]} \mu_\ell \left(\frac{1+p}{1-p}\right)^{\log_{2/(1-p)}(1/\mu_\ell)} \\
&= \mathop{\mathbb{E}}_{\ell\sim\mu}\left[\ \left(\frac{1}{\mu_\ell}\right)^{\gamma(p)}\ \right] \qquad \text{(def. of } \gamma(p) = \log_{\frac{2}{1-p}}(\tfrac{1+p}{1-p})) \\
&\leq \left(\mathop{\mathbb{E}}_{\ell\sim\mu}\left[\ \frac{1}{\mu_\ell}\ \right]\right)^{\gamma(p)} \qquad \text{(Jensen's inequality)} \\
&= m^{\gamma(p)}.
\end{aligned}
$$

Since $m = \mathsf{DL}(f)$, this complete the proof of our bound on decision list shrinkage.

We shall now assume that f is Boolean and $C_1 \vee \cdots \vee C_m$ is a minimum size DNF formula computing f. Let L be the equivalent decision list $((C_1, 1), \ldots, (C_m, 1), (\top, 0))$ of size $m + 1$. The shrinkage bound

$$\mathbb{E}[\ \mathsf{DNF}(f\restriction\boldsymbol{R}_p) + 1\ ] \leq (\mathsf{DNF}(f) + 1)^{\gamma(p)}$$

now follows from the above analysis, noting that $\mathsf{DNF}(f\restriction\varrho)+1 \leq \mathrm{size}(L\restriction\varrho)$ for all restrictions ϱ. ◀

3.3 Shrinkage of (weakly) orthogonal decision lists

▶ **Definition 13.** Let $L = ((C_1, b_1), \ldots, (C_m, b_m))$ be a decision list. We say that L is

- *orthogonal* if each input x satisfies exactly one of the conjunctive clauses $C_1, \ldots, C_m$,
- *weakly orthogonal* if each input x satisfies at most one of $C_1, \ldots, C_{m-1}$.

(Note that if L is weakly orthogonal, then it remains so after replacing C_m with $\top$. In contrast, an orthogonal decision list has $C_m = \top$ if and only if $m = 1$.)

For a function f on the hypercube, we denote by $\mathsf{(w)ODL}(f)$ the minimum size of a (weakly) orthogonal decision list that computes f. These complexity measures lies in-between DL and DT:

$$\mathsf{DL} \leq \mathsf{wODL} \leq \mathsf{ODL} \leq \mathsf{DT}.$$

Our proof of Theorem 3 implies a shrinkage bound for ODL and wODL in the same way as for $\mathsf{DNF} + 1$.

▶ **Corollary 14.** *For every function f on the hypercube,*

$$\mathbb{E}[\ \mathsf{ODL}(f\restriction\boldsymbol{R}_p)\ ] \leq \mathsf{ODL}(f)^{\gamma(p)} \quad \textit{and} \quad \mathbb{E}[\ \mathsf{wODL}(f\restriction\boldsymbol{R}_p)\ ] \leq \mathsf{wODL}(f)^{\gamma(p)}.$$

This follows from the observation that if L is orthogonal, then so is $L\restriction\varrho$ for any restriction ϱ, and if L is semi-orthogonal, then $L\restriction\varrho$ is semi-orthogonal after replacing the final conjunctive clause with $\top$.

3.4 Lower bound on the optimal $\gamma(p)$

What is the optimal function $\gamma(p)$ that may be chosen in the bound on decision list shrinkage of Theorem 3? We observe that $\gamma(p)$ cannot be improved beyond $\log_2(1+p)$. The lower bound is given by a (non-Boolean) function f computed by a read-once decision tree of depth k and size 2^k, in which each internal node queries a distinct variable and each leaf returns a distinct output value. For this f, we have $\mathsf{DL}(f) = 2^k$ and $\mathbb{E}[\ \mathsf{DL}(f{\restriction}\boldsymbol{R}_p)\ ] = (1+p)^k = \mathsf{DL}(f)^{\log_2(1+p)}$. The same function also shows that $\gamma(p)$ in Corollary 14 cannot improved beyond $\log_2(1+p)$. Since this function is not Boolean, it does not imply a lower bound on DNF shrinkage; however, a similar bound can be shown asymptotically by considering parity functions.

4 Shrinkage of AC^0 formulas

Our bound the shrinkage DNF and CNF formulas implies an (only slightly weaker) bound on the shrinkage of depth-2 formula leaf-size. We also discuss the relationship between leaf-size and a related size measure on AC^0 formulas, the number of depth-1 gates.

▶ **Definition 15.** An *AC^0 formula* is a formula composed unbounded fan-in AND and OR gates with inputs labeled by literals. We measure *depth* by the maximum number of gates on an input-to-output path; the expression "depth-d formula" refers to an AC^0 formula of depth at most d. As with DeMorgan formulas, the *leaf-size* of an AC^0 formula is the number of leaves labeled by literals. An alternative size measure is the number of depth-1 gates (that have only literals as inputs). This number is at least half the total number of gates in any formula with no (useless) gates of fan-in 1.

For a Boolean function f and $d \geq 2$, we denote by $\mathcal{L}_d(f)$ the minimum leaf-size of depth-d formula that computes f, and we denote by $\mathcal{F}_d(f)$ the minimum number of depth-1 gates in a depth-d formula that computes f. Note that $\mathcal{L}_d(f) = 1$ iff f is a literal, and $\mathcal{F}_d(f) = 1$ iff f is a nonempty conjunctive or disjunctive clause, and $\mathcal{L}_d(f) = \mathcal{F}_d(f) = 0$ iff f is constant (hence computed by a single AND or OR gate with fan-in zero, which as a formula has no inputs and no depth-1 gates).

Finally, we denote by $\mathcal{F}(f)$ the minimum number of depth-1 gates in an (unbounded depth, unbounded fan-in) formula that computes f.

Note that $\mathcal{F}_2 = \min\{\mathsf{DNF}, \mathsf{CNF}\}$. Theorem 3 therefore implies:

▶ **Corollary 16.** *For all Boolean functions f,*

$$\mathbb{E}[\ \mathcal{F}_2(f{\restriction}\boldsymbol{R}_p) + 1\ ] \leq (\mathcal{F}_2(f) + 1)^{\gamma(p)}.$$

Over n-variable Boolean functions, clearly $\mathcal{F}_d \leq \mathcal{L}_d \leq n \cdot \mathcal{F}_d$ and $\mathcal{F} \leq \mathcal{L} \leq n \cdot \mathcal{F}$. The next lemma shows that, under a 1/2-random restriction, $\mathcal{F}_d$ shrinks below $\mathcal{L}_d$ and $\mathcal{F}$ shrinks below $\mathcal{L}$ (independent of n).

▶ **Lemma 17.** *For all Boolean functions f and $d \geq 2$,*

$$\mathbb{E}[\ \mathcal{L}_d(f{\restriction}\boldsymbol{R}_{1/2})\ ] \leq \mathcal{F}_d(f) \quad \textit{and} \quad \mathbb{E}[\ \mathcal{L}(f{\restriction}\boldsymbol{R}_{1/2})\ ] \leq \mathcal{F}(f).$$

Proof. Let F be a [depth-d] AC^0 formula that computes f using the minimum number of depth-1 gates. By linearity of expectation, it suffices to show that each depth-1 subformula of F (i.e., conjunctive or disjunctive clause) has expected leaf-size at most 1 under $\boldsymbol{R}_{1/2}$. Indeed, for any $k \geq 1$ and $p \in [0, 1]$,

$$\mathbb{E}[\ \mathcal{L}(\mathrm{AND}_k{\restriction}\boldsymbol{R}_p)\ ] = \mathbb{E}[\ \mathcal{L}(\mathrm{OR}_k{\restriction}\boldsymbol{R}_p)\ ] = \sum_{j=0}^{k} j\binom{k}{j}p^j\left(\frac{1-p}{2}\right)^{k-j} = kp\left(\frac{1-p}{2}\right)^{k-1}.$$

When $p = \frac{1}{2}$, we have $\frac{k}{2}\left(\frac{3}{4}\right)^{k-1} < 1$ for all $k \geq 1$. ◀

Using Lemma 17, we obtain the following bound on the shrinkage of depth-2 formula leaf-size $\mathcal{L}_2$, which has a slightly worse exponent $\gamma(2p)$ compared to $\gamma(p)$ for $\mathcal{F}_2$ in Corollary 16.

▶ **Corollary 18** (Shrinkage of depth-2 formula leaf-size). *For all Boolean functions f,*

$$\mathbb{E}[\ \mathcal{L}_2(f{\restriction}\boldsymbol{R}_p) + 1\ ] \leq (\mathcal{L}_2(f) + 1)^{\gamma(2p)}.$$

Proof. Viewing $\boldsymbol{R}_p$ as a composition of $\boldsymbol{R}_{1/2}$ (first) and $\boldsymbol{R}_{2p}$ (second), we have

$$\begin{aligned}
\mathbb{E}[\ \mathcal{L}_2(f{\restriction}\boldsymbol{R}_p) + 1\ ] &= \mathop{\mathbb{E}}_{\varrho\sim\boldsymbol{R}_{2p}}\left[\ \mathop{\mathbb{E}}_{\sigma\sim\boldsymbol{R}_{1/2}}[\ \mathcal{L}_2((f{\restriction}\varrho){\restriction}\sigma) + 1\ ]\ \right] \\
&\leq \mathop{\mathbb{E}}_{\varrho\sim\boldsymbol{R}_{2p}}[\ \mathcal{F}_2(f{\restriction}\varrho) + 1\ ] && \text{(Lemma 17)} \\
&= (\mathcal{F}_2(f) + 1)^{\gamma(2p)} && \text{(Corollary 16)} \\
&\leq (\mathcal{L}_2(f) + 1)^{\gamma(2p)} && (\mathcal{F}_2 \leq \mathcal{L}_2).
\end{aligned}$$ ◀

As an additional consequence of Lemma 17, we observe that $\mathcal{F}$ has the same expected shrinkage factor (up to a constant factor) as DeMorgan leaf-size $\mathcal{L}$.

▶ **Corollary 19** (Shrinkage of unbounded fan-in, unbounded depth formulas). *For all Boolean functions f,*

$$\mathbb{E}[\ \mathcal{F}(f{\restriction}\boldsymbol{R}_p)\ ] = O(\ p^2\mathcal{F}(f) + p\sqrt{\mathcal{F}(f)}\ ).$$

Proof. Assume $p \leq 1/2$, since the bound is trivial otherwise. Viewing $\boldsymbol{R}_p$ as a composition of $\boldsymbol{R}_{2p}$ (first) and $\boldsymbol{R}_{1/2}$ (second), we have

$$\begin{aligned}
\mathbb{E}[\ \mathcal{F}(f{\restriction}\boldsymbol{R}_p)\ ] &= \mathop{\mathbb{E}}_{\sigma\sim\boldsymbol{R}_{1/2}}\left[\ \mathop{\mathbb{E}}_{\varrho\sim\boldsymbol{R}_{2p}}[\ \mathcal{F}((f{\restriction}\sigma){\restriction}\varrho)\ ]\ \right] \\
&\leq \mathop{\mathbb{E}}_{\sigma\sim\boldsymbol{R}_{1/2}}\left[\ \mathop{\mathbb{E}}_{\varrho\sim\boldsymbol{R}_{2p}}[\ \mathcal{L}((f{\restriction}\sigma){\restriction}\varrho)\ ]\ \right] && (\mathcal{F} \leq \mathcal{L}) \\
&= \mathop{\mathbb{E}}_{\sigma\sim\boldsymbol{R}_{1/2}}\left[\ O\big(4p^2\mathcal{L}(f{\restriction}\sigma) + 2p\sqrt{\mathcal{L}(f{\restriction}\sigma)}\big)\ \right] && \text{(Theorem 1)} \\
&= O\Big(p^2\mathop{\mathbb{E}}_{\sigma\sim\boldsymbol{R}_{1/2}}[\ \mathcal{L}(f{\restriction}\sigma)\ ] + p\sqrt{\mathop{\mathbb{E}}_{\sigma\sim\boldsymbol{R}_{1/2}}[\ \mathcal{L}(f{\restriction}\sigma)\ ]}\Big) && \text{(Jensen's inequality)} \\
&= O(\ p^2\mathcal{F}(f) + p\sqrt{\mathcal{F}(f)}\ ) && \text{(Lemma 17)}.
\end{aligned}$$ ◀

5 Open problems

We conclude by mentioning some questions raised by this work.

▶ **Open Problem 1.** *Determine the optimal function $\gamma_{\mathsf{DL}}(p)$ in Theorem 3. We have shown that*

$$\log_2(1+p) = \gamma_{\mathsf{DT}}(p) \leq \gamma_{\mathsf{DL}}(p) \leq \log_{\frac{2}{1-p}}(\tfrac{1+p}{1-p}).$$

A simpler problem is to determine the least constant C_{DL} such that $\mathbb{E}[\ \mathsf{DL}(f{\restriction}\boldsymbol{R}_p)\ ] \leq O(\mathsf{DL}(f)^{C_{\mathsf{DL}} \cdot p})$. It follows from our bounds that $\frac{1}{\ln 2} = C_{\mathsf{DT}} \leq C_{\mathsf{DL}} \leq \frac{2}{\ln 2}$. The same questions may be asked with respect to complexity measures ODL*,* wODL *and* DNF*.*

▶ **Open Problem 2.** *Determine the shrinkage rate of depth-d AC^0 formulas for $d \geq 3$. We expect that*

$$\mathbb{E}[\ \mathcal{L}_d(f{\restriction}\boldsymbol{R}_p)\ ] \leq \mathcal{L}_d(f)^{O(p^{1/(d-1)})}. \tag{6}$$

Ideally the constant in this big-O should not depend on d.

We remark that inequality (6) is known to hold for small $p = O(1/\log \mathcal{L}_d(f))^{d-1}$, when the bound is $O(1)$. This can be shown using the (Multi-)Switching Lemma of Håstad [12]. It is also a direct consequence of the following result of the author [22], which generalizes Corollary 6 (on the decision tree size of decision lists) to AC^0 formulas of any depth.

▶ **Theorem 20** (Decision tree size of AC^0 formulas [22])**.** *For all functions $f : \{0,1\}^n \to \{0,1\}$ computable by depth-d AC^0 formulas with fan-in m (and leaf-size at most nm^{d-1}),*

$$\mathbb{E}[\ \mathsf{DT}(f{\restriction}\boldsymbol{R}_p)\ ] \leq 2 \quad \text{and} \quad \mathsf{DT}(f) \leq O(2^{(1-p)n}) \quad \text{where} \quad p = O(1/\log m)^{d-1}.$$

A related question:

▶ **Open Problem 3.** *Prove a stronger version of Theorem 20 for depth-d AC^0 formulas with $m = \mathcal{F}_d(f)^{1/(d-1)}$ (instead of fan-in, which is larger for unbalanced formulas). Such a result could be helpful in proving the shrinkage bound (6).*

Finally, we repeat the longstanding question concerning shrinkage of monotone formulas:

▶ **Open Problem 4.** *Determine the shrinkage exponent of monotone formulas. That is, find the maximum constant Γ_{m} such that*

$$\mathbb{E}[\ \mathcal{L}_{\mathrm{m}}(f{\restriction}\boldsymbol{R}_p)\ ] \leq O(p^{\Gamma_{\mathrm{m}}-o(1)}\mathcal{L}_{\mathrm{m}}(f)+1)$$

for all monotone Boolean functions f, where $\mathcal{L}_{\mathrm{m}}$ is monotone formula leaf-size. It is known that $2 = \Gamma_{\mathrm{DeMorgan}} \leq \Gamma_{\mathrm{m}} \leq \Gamma_{\text{read-once}} = \log_{\sqrt{5}-1}(2) \approx 3.27$, and the second inequality is believed to be tight [5, 13].

References

1 Alexander E Andreev. On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes. *Moscow Univ. Math. Bull.*, 42(1):63–66, 1987.

2 Paul Beame. A switching lemma primer. Technical report, Technical Report UW-CSE-95-07-01, Department of Computer Science and Engineering, University of Washington, 1994.

3 Avrim Blum. Rank-r decision trees are a subclass of r-decision lists. *Information Processing Letters*, 42(4):183–185, 1992.
4 Nader H Bshouty. A subexponential exact learning algorithm for dnf using equivalence queries. *Information Processing Letters*, 59(1):37–39, 1996.
5 Moshe Dubiner and Uri Zwick. How do read-once formulae shrink? *parity*, 2(1):63, 1993.
6 Yuval Filmus, Or Meir, and Avishay Tal. Shrinkage under random projections and cubic formula lower bounds for AC^0. In *12th Innovations in Theoretical Computer Science Conference (ITCS)*, 2021.
7 Parikshit Gopalan, Raghu Meka, and Omer Reingold. DNF sparsification and a faster deterministic counting algorithm. *Computational Complexity*, 22(2):275–310, 2013.
8 Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 120–129. IEEE, 2012.
9 Thomas Hancock, Tao Jiang, Ming Li, and John Tromp. Lower bounds on learning decision lists and trees. *Information and Computation*, 126(2):114–122, 1996.
10 Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proc. 18th ACM Symposium on Theory of Computing*, pages 6–20, 1986.
11 Johan Håstad. The shrinkage exponent of de Morgan formulas is 2. *SIAM Journal on Computing*, 27(1):48–64, 1998.
12 Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM Journal on Computing*, 43(5):1699–1708, 2014.
13 Johan Håstad, Alexander Razborov, and Andrew Yao. On the shrinkage exponent for read-once formulae. *Theoretical Computer Science*, 141(1-2):269–282, 1995.
14 Russell Impagliazzo and Noam Nisan. The effect of random restrictions on formula size. *Random Structures & Algorithms*, 4(2):121–133, 1993.
15 Ilan Komargodski and Ran Raz. Average-case lower bounds for formula size. In *Proceedings of the 45th ACM Symposium on Theory of Computing*, pages 171–180, 2013.
16 Matthias Krause. On the computational power of boolean decision lists. *Computational Complexity*, 14(4):362–375, 2006.
17 Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Decision list compression by mild random restrictions. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 247–254, 2020.
18 Ryan O'Donnell. *Analysis of boolean functions.* Cambridge University Press, 2014.
19 Michael S Paterson and Uri Zwick. Shrinkage of de Morgan formulae under restriction. *Random Structures & Algorithms*, 4(2):135–150, 1993.
20 Alexander A Razborov. Pseudorandom generators hard for k-DNF resolution and polynomial calculus resolution. *Annals of Mathematics*, pages 415–472, 2015.
21 Ronald L Rivest. Learning decision lists. *Machine learning*, 2(3):229–246, 1987.
22 Benjamin Rossman. Criticality of Regular Formulas. In *34th Computational Complexity Conference (CCC 2019)*, volume 137 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 1:1–28, 2019. doi:10.4230/LIPIcs.CCC.2019.1.
23 Rahul Santhanam. Fighting perebor: New and improved algorithms for formula and QBF satisfiability. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 183–192. IEEE, 2010.
24 Nathan Segerlind, Sam Buss, and Russell Impagliazzo. A switching lemma for small restrictions and lower bounds for k-DNF resolution. *SIAM Journal on Computing*, 33(5):1171–1200, 2004.
25 Bella A. Subbotovskaya. Realizations of linear functions by formulas using $+,\cdot,-$. *Doklady Akademii Nauk SSSR*, 136(3):553–555, 1961.
26 Avishay Tal. Shrinkage of De Morgan formulae by spectral techniques. In *55th Annual IEEE Symposium on Foundations of Computer Science*, pages 551–560, 2014.