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 $R_p$ for all values of $p \in [0,1]$. For a function $f$ with domain $\{0,1\}^n$, let $\text{DL}(f)$ denote the minimum size of a decision list that computes $f$. We show that
\[ \mathbb{E}[\text{DL}(f|R_p)] \leq \text{DL}(f) \log_2 \frac{1}{(1-p)(1+p)} + \frac{1}{2} \cdot p. \]
For example, this bound is $\sqrt{\text{DL}(f)}$ when $p = \sqrt{3} - 2 \approx 0.24$. For Boolean functions $f$, we obtain the same shrinkage bound with respect to DNF formula size plus 1 (i.e., replacing $\text{DL}(\cdot)$ with $\text{DNF}(\cdot) + 1$ 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 $R_p$ independently leaves each input bit free with probability $p$ and otherwise assigns it to 0 or 1 with equal probability. We denote by $f|R_p$ the function obtained from $f$ by restricting its inputs to the subcube of $\{0,1\}^n$ that correspond to $R_p$.

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 $\mathcal{L}(f|R_p)$ 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 $R_p$ 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 $p \geq 1/\mathcal{L}(f)^{1/3}$). 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:

\[ \mathbb{E}[\mathcal{L}(f|R_p)] = O(p^2 \mathcal{L}(f) + p \sqrt{\mathcal{L}(f)}). \]

© Benjamin Rossman; licensed under Creative Commons License CC-BY
12th Innovations in Theoretical Computer Science Conference (ITCS 2021).
Editor: James R. Lee; Article No. 70; pp. 70:1–70:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
70:2 Shrinkage of Decision Lists and DNF Formulas

The constant 2 in the exponent \( p \) in Theorem 1 is known as the “shrinkage exponent” of DeMorgan formulas. Shrinkage under \( R_p \) 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 \( \log \sqrt{\pi-1} (2) \approx 3.27 \). The shrinkage exponent of monotone formulas is between 2 and \( \log \sqrt{\pi-1} (2) \) 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 \( R_p \) 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 \( R_p \) 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 \( DT(f) \) 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 \( DT(f) \). (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,

\[
E[ DT(f|R_p) ] \leq DT(f)^{\log_d(1+p)}.
\]

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 \( C_1 \lor \cdots \lor C_m \) 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 \( DL(f) \) 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 \( DL(f) \leq DT(f) \) 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 \( DL(f) \leq DT(f) + 1 \) 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 \( R_p \).

---

\(^1\) 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 \( C_1 \lor \cdots \lor C_m \) 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.
\begin{enumerate}
  \item \textbf{Theorem 3} (Shrinkage of decision lists and DNF formulas). For all functions \( f \) on the hypercube,
  \[
  \mathbb{E}[\ DL(f|\mathcal{R}_p) ] \leq DL(f)^{\gamma(p)} \quad \text{where} \quad \gamma(p) := \log_{1+\frac{1}{1-p}}(\frac{1+\frac{1}{1-p}}{1-p}) \ .
  \]
  If \( f \) is Boolean, then also \( \mathbb{E}[\ DNF(f|\mathcal{R}_p) + 1 ] \leq (DNF(f) + 1)^{\gamma(p)} \) (and similarly for CNF\)(\cdot + 1)).
\end{enumerate}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{gamma_plot.png}
\caption{Plots of \( \gamma(p) := \log_{1+\frac{1}{1-p}}(\frac{1+\frac{1}{1-p}}{1-p}) \) (blue) and \( \log_{2}(1 + p) \) (red).}
\end{figure}

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 \( ODL(f)^{\gamma(p)} \) and \( wODL(f)^{\gamma(p)} \) on the shrinkage of orthogonal and weakly orthogonal decision lists (Corollary 14), as well as \( (L_2(f) + 1)^{\gamma(2p)} \) for depth-2 formula leaf-size (Corollary 18).

Theorem 3 yields the following bounds for particular settings of \( p \) in terms of \( m = DL(f) \):
\[
\mathbb{E}[\ DL(f|\mathcal{R}_p) ] \leq \\
\begin{cases} 
2 & \text{for } p = O(\frac{1}{\log m}), \\
\sqrt{m} & \text{for } p = \sqrt{5} - 2 \approx 0.24, \\
m/2 & \text{for } p = 1 - O(\frac{\log \log m}{\log m}), \\
m - 1 & \text{for } p = 1 - O(\frac{\log \log m}{\log m}).
\end{cases}
\]

For small \( p = O(1/\log m) \), a variant of Håstad’s Switching Lemma (discussed below) actually implies a stronger inequality \( \mathbb{E}[\ DT(f|\mathcal{R}_p) ] \leq 2 \) 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 \( p = \Omega(1) \) or \( 1 - o(1) \) has important applications in pseudorandomness [8, 20], DNF sparsification [7, 17] and hypercontractivity [18].

\subsection{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 \( \mathcal{R}_p \). The switching lemmas stated below can be viewed as apples-to-oranges shrinkage results that bound one complexity measure on \( f|\mathcal{R}_p \) 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 \( DT_{\text{depth}}, DL_{\text{width}}, DNF_{\text{width}} \). 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 $O(\log m)$. Hästad’s Switching Lemma [10] gives a tail bound on the decision tree size of $f\lceil R_p$ in terms of the decision list width of $f$.\footnote{In its application to $AC^0$ circuit lower bounds, Theorem 4 is usually stated (more narrowly) in the form

$$E[\text{CNF}_{\text{width}}(f \lceil R_p) \geq t] \leq O(p \cdot \text{DNF}_{\text{width}}(f))^t$$

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].}

\begin{itemize}
  \item \textbf{Theorem 4 (Switching Lemma [10])}. For all functions $f$ on the hypercube and $t \in \mathbb{N}$,
  \[P[ DT_{\text{depth}}(f \lceil R_p) \geq t] \leq O(p \cdot \text{DL}_{\text{width}}(f))^t.\]
  A variant of the Switching Lemma with log DL$(f)$ in place of DL$_{\text{width}}(f)$ was proved in [22].
  \item \textbf{Theorem 5 (Switching Lemma in terms of decision list size [22])}. For every function $f$ on the hypercube and $t \in \mathbb{N}$,
  \[P[ DT_{\text{depth}}(f \lceil R_p) \geq t] \leq O(p \cdot \log \text{DL}(f))^t.\]
  
  We remark that Theorem 5 follows directly from Theorem 4 for $t \leq O(\log \text{DL}(f))$ (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:
  \item \textbf{Corollary 6 (Decision tree size of decision lists)}. For all functions $f$ on $\{0,1\}^n$,
  \[E[ DT(f \lceil R_p)] \leq 2 \quad \text{and} \quad DT(f) \leq O(2(1-p)n) \quad \text{where} \quad p = O(1/\log \text{DL}(f)).\]
  As previously mentioned, Corollary 6 strengthen the bound $E[\text{DL}(f \lceil R_p)] \leq 2$ for $p = O(1/\log \text{DL}(f))$ that follows from Theorem 3 (albeit for $p$ that is a constant factor smaller). However, note that Corollary 6 is trivial for $p$ above $\Omega(1/\log \text{DL}(f))$. A different switching lemma for large $p$ (even $1-o(1)$) in terms of 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 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 clause of $L$ and $R$ 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.
  \item \textbf{Theorem 7 (Decision list shrinkage in terms of width [17])}. For every function $f$ on the hypercube,
  \[E[\text{DL}(f \lceil R_p)] \leq \left(\frac{4}{1-p}\right)^{\text{DL}_{\text{width}}(f)}.\]
  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 DL$_{\text{width}}(f)$ with log DL$(f)$.
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 $R_p$ (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 $R_p$.

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 $\rho : \{1, \ldots, n\} \rightarrow \{0, 1, *\}$ where $\rho(i) = *$ signifies that $x_i$ is left free by $\rho$. We denote by $\text{Stars}(\rho) \subseteq [n]$ the set of free variables under $\rho$. For a function $f$ on the hypercube $\{0, 1\}^n$ and a restriction $\rho$, we denote by $f|_{\rho}$ the restricted function on the subcube $\{0, 1\}^{\text{Stars}(\rho)}$ defined in the obvious way: $(f|_{\rho})(y) = f(x)$ where $x \in \{0, 1\}^n$ is the input with $x_i = y_i$ if $i \in \text{Stars}(\rho)$ and $x_i = \rho(i)$ otherwise.

For $p \in [0, 1]$, the $p$-random restriction $R_p$ 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 $\rho$, we have $\mathbb{P}[R_p = \rho] = p^{\text{Stars}(\rho)}((1 - p)/2)^{n - \text{Stars}(\rho)}$.

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 $\ell_1 \land \cdots \land \ell_w$ 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 $C_1 \lor \cdots \lor C_m$ 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 $\bot$. The width of $F$ is defined as the maximum width of any $C_i$.
- CNF formulas are defined dually (with the roles of $\lor$ and $\land$ exchanged).
Every literal, conjunctive clause, and DNF formula computes a Boolean function \( \{0, 1\}^n \rightarrow \{0, 1\} \) 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 \((C_1, b_1), \ldots, (C_m, b_m)\) where \( b_1, \ldots, b_m \) are arbitrary output values (not necessarily Boolean) and \( C_1, \ldots, C_m \) are conjunctive clauses such that \( C_1 \lor \cdots \lor C_m \) 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 \( \{0, 1\}^n \rightarrow \{b_1, \ldots, b_m\} \) as follows: on input \( x \), the output is \( b_i \) 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

\[
DT(f) := \text{minimum size of a decision tree that computes } f,
\]

\[
DL(f) := \text{minimum size of a decision list that computes } f,
\]

When \( f \) is Boolean, we additionally define

\[
\text{DNF}(f) := \text{minimum size of a DNF formula that computes } f,
\]

\[
\text{CNF}(f) := \text{minimum size of a CNF formula that computes } f.
\]

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

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

\[
\text{DT}_{\text{width}}(f), \quad \text{DL}_{\text{width}}(f), \quad \text{DNF}_{\text{width}}(f), \quad \text{CNF}_{\text{width}}(f).
\]

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

\[
1 \leq DL \leq \{ \text{DNF} + 1, \text{CNF} + 1 \} \leq \text{DNF} + \text{CNF} \leq DT.
\]

The corresponding width/depth measures satisfy:

\[
0 \leq DL_{\text{width}} \leq \left\{ \frac{\text{DNF}_{\text{width}}}{\text{CNF}_{\text{width}}}, \left\lceil \log_2(DT) \right\rceil \right\} \leq DT_{\text{depth}} \leq \text{DNF}_{\text{width}} \cdot \text{CNF}_{\text{width}}.
\]

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

$$E[\text{size}(T | R_p)] \leq \text{size}(T)^{\log_2(1+p)}.$$

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 $m \geq 2$. 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 $m_0, m_1 \geq 1$ with $m_0 + m_1 = m$. Without loss of generality, $T_0$ and $T_1$ never query $x_i$. We have

$$E[\text{size}(T | R_p)] = pE[\text{size}(T | R_p) | R_p(x_i) = \ast]$$

$$+ \frac{1-p}{2} \left( E[\text{size}(T_0 | R_p) | R_p(x_i) = 0] + E[\text{size}(T_1 | R_p)] \right)$$

$$\leq \frac{1+p}{2} \left( (m_0)^{\log_2(1+p)} + (m_1)^{\log_2(1+p)} \right) \quad \text{(induction hypothesis)}$$

$$\leq (1+p) \left( \frac{m}{2} \right)^{\log_2(1+p)} \quad \text{(Jensen’s inequality)}$$

$$= m^{\log_2(1+p)}.$$

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

$$E[DT(f | R_p)] = E[2^{\text{Bin}(k,p)}] = \sum_{i=0}^{k} 2^i P[\text{Bin}(k,p) = i]$$

$$= \sum_{i=0}^{k} \binom{k}{i} (2p)^i (1-p)^{k-i} = (1+p)^k = DT(f)^{\log_2(1+p)}.$$ 

### 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.)
Shrinkage of Decision Lists and DNF Formulas

Let $L = ((C_1, b_1), \ldots, (C_m, b_m))$ be a decision list of minimum size that computes $f$, that is, with $m = DL(f)$. For $\ell \in [m]$, let $|C_\ell|$ denote the width of the clause $C_\ell$ (i.e., the number of literals in $C_\ell$). 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 $\rho$, we define the set $U(\rho) \subseteq [m]$ of useful indices of $L$ under $\rho$ by

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

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

$$L|_{\rho} := ((C_{\ell_1}|_{\rho}, b_{\ell_1}), \ldots, (C_{\ell_t}|_{\rho}, b_{\ell_t}))$$

where $C_{\ell_i}|_{\rho}$ is the sub-clause of $C_{\ell_i}$ on the variables left unrestricted by $\rho$. (Note that $C_{\ell_i} \lor \cdots \lor C_{\ell_t}$ is a tautology, so $L|_{\rho}$ is indeed a decision list.) Thus, we have

$$\text{DL}(f|_{\rho}) \leq |U(\rho)|.$$  

(1)

For example, suppose $m = 4$ and

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

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

$$U(\varrho_1) = \{1, 3, 4\}, \quad L|_{\varrho_1} = ((x_3, b_1), (x_2 \land \overline{x_3}, b_3), (\top, b_4)).$$

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

$$U(\varrho_2) = \{1, 3\}, \quad L|_{\varrho_2} = ((x_3, b_1), (\overline{x_3}, b_3)).$$

In particular, the final clause $C_4$ is not useful under $\varrho_2$ (since any input consistent with $\varrho_2$ 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 $[m]$)

$$\mu_\ell := \Pr_{\varrho \sim R_p}[\max(U(\varrho)) = \ell \text{ and } C_\ell|_{\varrho} \equiv 1] \quad \text{for } \ell \in [m-1],$$

$$\mu_m := \Pr_{\varrho \sim R_p}[\max(U(\varrho)) = m \text{ or } C_{\max(U(\varrho))}|_{\varrho} \not\equiv 1].$$

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 $C_\ell|_{\varrho} \equiv 1$, that is, $\mu_\ell$ does not necessarily equal $\Pr_{\varrho \sim R_p}[\max(U(\varrho)) = \ell]$. This is illustrated by the restriction $\varrho_2$ in the above example, for which we have $\max(U(\varrho_2)) = 3$, yet $C_3|_{\varrho_2} = \overline{x_3} \not\equiv 1$. Restrictions $\varrho_1$ and $\varrho_2$ both contribute to probability mass $\mu_4$: in the case of $\varrho_1$, this is because $\max(U(\varrho_1)) = 4$, and in the case of $\varrho_2$, this is because $C_{\max(U(\varrho_2))}|_{\varrho_2} \not\equiv 1$.

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

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

(2)
We require one more definition. For a restriction \( g \) and a useful index \( \ell \in U(g) \), let \( g^{(\ell)} \) be the restriction obtained by augmenting \( g \) by the unique satisfying assignment for the clause \( C_\ell \). That is, \( g^{(\ell)} \) fixes a variable \( x_i \) to \( a \in \{0, 1\} \) if, and only if, \( g \) fixes \( x_i \) to \( a \) or \( x_i = a \) in the satisfying assignment to \( C_\ell \).

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

\[
\Pr[R_p = g^{(\ell)}] = \left( \frac{2p}{1 - p} \right)^{\text{Stars}(g) \cap \text{Vars}(C_\ell)} \tag{3}
\]

since \( g^{(\ell)} \) has exactly \( |\text{Stars}(g) \cap \text{Vars}(C_\ell)| \) fewer unrestricted variables ("stars") than \( g \).

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

\[
\max(U(g^{(\ell)})) = \ell \quad \text{and} \quad C_\ell | g^{(\ell)} \equiv 1. \tag{4}
\]

Thus, \( g^{(\ell)} \) contributes to the probability mass \( \mu_\ell \).

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

\[
\Pr_{g \sim R_p} [ \ell \in U(g) ] \leq \mu_\ell \left( \frac{1 + p}{1 - p} \right)^{|C_\ell|}. \tag{5}
\]

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

\[
\Pr_{g \sim R_p} [ \ell \in U(g) ] \\
= \sum_{S \subseteq \text{Vars}(C_\ell)} \Pr_{g \sim R_p} [ \ell \in U(g) \text{ and } \text{Stars}(g) \cap \text{Vars}(C_\ell) = S ] \\
\overset{(4)}{=} \sum_{S \subseteq \text{Vars}(C_\ell)} \Pr_{g \sim R_p} [ \ell = \max(U(g^{(\ell)})) \text{ and } C_\ell | g^{(\ell)} \equiv 1 \text{ and } \text{Stars}(g) \cap \text{Vars}(C_\ell) = S ] \\
= \sum_{S \subseteq \text{Vars}(C_\ell)} \left( \sum_{g : \ell = \max(U(g^{(\ell)})) \text{ and } C_\ell | g^{(\ell)} \equiv 1 \text{ and } \text{Stars}(g) \cap \text{Vars}(C_\ell) = S} \Pr[R_p = g] \right) \\
= \sum_{S \subseteq \text{Vars}(C_\ell)} \left( \frac{2p}{1 - p} \right)^{|S|} \left( \sum_{g : \ell = \max(U(g)) \text{ and } C_\ell | g \equiv 1 \text{ and } \text{Stars}(g) \cap \text{Vars}(C_\ell) = S} \Pr[R_p = g] \right) \\
= \mu_\ell \left( \frac{2p}{1 - p} \right)^{|C_\ell|} \left( \sum_{S \subseteq \text{Vars}(C_\ell)} \left( \frac{2p}{1 - p} \right)^{|S|} \right) \\
= \mu_\ell \left( \frac{1 + p}{1 - p} \right)^{|C_\ell|} \text{ (definition of } \mu_\ell \text{)}.
\]

Finally, we obtain the shrinkage bound of Theorem 3 by the following calculation, which uses Jensen’s inequality in addition to the above observations:
Since \( m = DL(f) \), this completes the proof of our bound on decision list shrinkage.

We shall now assume that \( f \) is Boolean and \( C_1 \lor \cdots \lor C_m \) 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

\[
E[\frac{DNF(f|\varrho) + 1}{m}] \leq (DNF(f) + 1)^{\gamma(p)}
\]

now follows from the above analysis, noting that \( DNF(f|\varrho) + 1 \leq size(L|\varrho) \) for all restrictions \( \varrho \).

### 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 \((w)ODL(f)\) the minimum size of a (weakly) orthogonal decision list that computes \( f \). These complexity measures lies in-between DL and DT:

\[
DL \leq wODL \leq ODL \leq DT.
\]

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

**Corollary 14.** For every function \( f \) on the hypercube,

\[
E[ODL(f|R_p)] \leq ODL(f)^{\gamma(p)} \quad \text{and} \quad E[wODL(f|R_p)] \leq wODL(f)^{\gamma(p)}.
\]

This follows from the observation that if \( L \) is orthogonal, then so is \( L|\varrho \) for any restriction \( \varrho \), and if \( L \) is semi-orthogonal, then \( L|\varrho \) 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 $ DL(f) = 2^k $ and $ E[ DL(f | R_p) ] = (1+p)^k = 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 $\text{AC}^0$ 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 $\text{AC}^0$ formulas, the number of depth-1 gates.

Definition 15. An $\text{AC}^0$ 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 $\text{AC}^0$ formula of depth at most $d$. As with DeMorgan formulas, the leaf-size of an $\text{AC}^0$ 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 $\mathcal{F}_2 = \min\{ \text{DNF, CNF} \}$. Theorem 3 therefore implies:

Corollary 16. For all Boolean functions $f$,

$$ E[ \mathcal{F}_2(f | R_p) + 1 ] \leq (\mathcal{F}_2(f) + 1)^{\gamma(p)}. $$

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 $d \geq 2$,

$$ E[ \mathcal{L}_d(f | R_{1/2}) ] \leq \mathcal{F}_d(f) \quad \text{and} \quad E[ \mathcal{L}(f | R_{1/2}) ] \leq \mathcal{F}(f). $$

Proof. Let $F$ be a [depth-$d$] $\text{AC}^0$ 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 $R_{1/2}$. Indeed, for any $k \geq 1$ and $p \in [0, 1]$,
\[ E[\mathcal{L}(\text{AND}_k|R_p)] = E[\mathcal{L}(\text{OR}_k|R_p)] = \sum_{j=0}^{k} \binom{k}{j}p^j\left(\frac{1-p}{2}\right)^{k-j} = kp\left(\frac{1-p}{2}\right)^{k-1}. \]

When \( p = \frac{1}{2} \), we have \( \frac{k}{2} \left(\frac{1}{4}\right)^{k-1} < 1 \) 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.

\[ \mathbf{Corollary 18 (Shrinkage of depth-2 formula leaf-size).} \quad \text{For all Boolean functions } f, \]
\[ E[\mathcal{L}_2(f|R_p) + 1] \leq (\mathcal{L}_2(f) + 1)^{\gamma(2p)}. \]

\textbf{Proof.} Viewing \( R_p \) as a composition of \( R_{1/2} \) (first) and \( R_{2p} \) (second), we have
\[ E[\mathcal{L}_2(f|R_p) + 1] = E_{\sigma \sim R_{1/2}} \left[ E_{\varrho \sim R_{2p}} [\mathcal{L}_2((f|\varrho)|\sigma) + 1] \right] \]
\[ \leq E_{\sigma \sim R_{1/2}} \left[ E_{\varrho \sim R_{2p}} [\mathcal{F}_2(f|\varrho) + 1] \right] \quad \text{(Lemma 17)} \]
\[ = (\mathcal{F}_2(f) + 1)^{\gamma(2p)} \quad \text{(Corollary 16)} \]
\[ \leq (\mathcal{L}_2(f) + 1)^{\gamma(2p)} \quad \text{(\( \mathcal{F}_2 \leq \mathcal{L}_2 \)).} \]

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 \( \mathcal{L} \).

\[ \mathbf{Corollary 19 (Shrinkage of unbounded fan-in, unbounded depth formulas).} \quad \text{For all Boolean functions } f, \]
\[ E[\mathcal{F}(f|R_p)] = O(p^2\mathcal{F}(f) + p\sqrt{\mathcal{F}(f)}). \]

\textbf{Proof.} Assume \( p \leq 1/2 \), since the bound is trivial otherwise. Viewing \( R_p \) as a composition of \( R_{2p} \) (first) and \( R_{1/2} \) (second), we have
\[ E[\mathcal{F}(f|R_p)] = E_{\sigma \sim R_{1/2}} \left[ E_{\varrho \sim R_{2p}} [\mathcal{F}(f|\sigma)|\varrho)] \right] \]
\[ \leq E_{\sigma \sim R_{1/2}} \left[ E_{\varrho \sim R_{2p}} [\mathcal{L}(f|\sigma)|\varrho)] \right] \quad \text{(\( \mathcal{F} \leq \mathcal{L} \))} \]
\[ = E_{\sigma \sim R_{1/2}} \left[ O(4p^2\mathcal{L}(f|\sigma) + 2p\sqrt{\mathcal{L}(f|\sigma)}) \right] \quad \text{(Theorem 1)} \]
\[ = O\left( p^2 E_{\sigma \sim R_{1/2}} [\mathcal{L}(f|\sigma)] + p\sqrt{E_{\sigma \sim R_{1/2}} [\mathcal{L}(f|\sigma)]} \right) \quad \text{(Jensen’s inequality)} \]
\[ = O(p^2\mathcal{F}(f) + p\sqrt{\mathcal{F}(f)}) \quad \text{(Lemma 17).} \]
5 Open problems

We conclude by mentioning some questions raised by this work.

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

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

A simpler problem is to determine the least constant $C_{DL}$ such that

$$E[DL(f|m)] \leq O(p^{\Omega/(d-1)}).$$  

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

We remark that inequality (6) is known to hold for small $p = O(1/\log L_d(f))^{d-1}$, 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 $f : \{0, 1\}^n \rightarrow \{0, 1\}$ computable by depth-$d$ $AC^0$ formulas with fan-in $m$ (and leaf-size at most $nm^{d-1}$),

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

A related question:

▶ Open Problem 3. Prove a stronger version of Theorem 20 for depth-$d$ $AC^0$ formulas with $m = F_d(f)^{1/(d-1)}$ (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 $\Gamma_m$ such that

$$E[L_m(f|m)] \leq O(p^{\Gamma_m-o(1)}L_m(f) + 1)$$  

for all monotone Boolean functions $f$, where $L_m$ is monotone formula leaf-size. It is known that $2 = \Gamma_{DeMorgan} \leq \Gamma_m \leq \Gamma_{read-once} = \log_5\sqrt{5-1}(2) \approx 3.27$, and the second inequality is believed to be tight [5, 13].

References


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.


