Randomized Query Composition and Product Distributions

Swagato Sanyal
Department of Computer Science and Engineering,
Indian Institute of Technology Kharagpur, India

Abstract

Let R_ϵ denote randomized query complexity for error probability ϵ, and $R := \mathsf{R}_{1/3}$. In this work we investigate whether a perfect composition theorem $\mathsf{R}(f \circ g^n) = \Omega(\mathsf{R}(f) \cdot \mathsf{R}(g))$ holds for a relation $f \subseteq \{0,1\}^n \times \mathcal{S}$ and a total inner function $g : \{0,1\}^m \to \{0,1\}$.

Composition theorems of the form $\mathsf{R}(f \circ g^n) = \Omega(\mathsf{R}(f) \cdot \mathsf{M}(g))$ are known for various measures M. Such measures include the sabotage complexity RS defined by Ben-David and Kothari (ICALP 2015), the max-conflict complexity defined by Gavinsky, Lee, Santha and Sanyal (ICALP 2019), and the linearized complexity measure defined by Ben-David, Blais, Göös and Maystre (FOCS 2022). The above measures are asymptotically non-decreasing in the above order. However, for total Boolean functions no asymptotic separation is known between any two of them.

Let $\mathsf{D}^{\mathsf{prod}}$ denote the maximum distributional query complexity with respect to any product (over variables) distribution . In this work we show that for any total Boolean function g, the sabotage complexity $\mathsf{RS}(g) = \widetilde{\Omega}(\mathsf{D}^{\mathsf{prod}}(g))$. This gives the composition theorem $\mathsf{R}(f \circ g^n) = \widetilde{\Omega}(\mathsf{R}(f) \cdot \mathsf{D}^{\mathsf{prod}}(g))$. In light of the minimax theorem which states that $\mathsf{R}(g)$ is the maximum distributional complexity of g over any distribution, our result makes progress towards answering the composition question.

We prove our result by means of a complexity measure $\mathsf{R}^{\mathsf{prod}}_\epsilon$ that we define for total Boolean functions. Informally, $\mathsf{R}^{\mathsf{prod}}_\epsilon(g)$ is the minimum complexity of any randomized decision tree with unlabelled leaves with the property that, for every product distribution μ over the inputs, the average bias of its leaves is at least $((1-\epsilon)-\epsilon)/2 = 1/2 - \epsilon$. It follows by standard arguments that $\mathsf{R}^{\mathsf{prod}}_{1/3}(g) = \Omega(\mathsf{D}^{\mathsf{prod}}(g))$. We show that $\mathsf{R}^{\mathsf{prod}}_{1/3}$ is equivalent to the sabotage complexity up to a logarithmic factor.

Ben-David and Kothari asked whether $\mathsf{RS}(g) = \Theta(\mathsf{R}(g))$ for total functions g. We generalize their question and ask if for any error ϵ, $\mathsf{R}^{\mathsf{prod}}_\epsilon(g) = \widetilde{\Theta}(\mathsf{R}_\epsilon(g))$. We observe that the work by Ben-David, Blais, Göös and Maystre (FOCS 2022) implies that for a perfect composition theorem $\mathsf{R}_{1/3}(f \circ g^n) = \Omega(\mathsf{R}_{1/3}(f) \cdot \mathsf{R}_{1/3}(g))$ to hold for any relation f and total function g, a necessary condition is that $\mathsf{R}_{1/3}(g) = O(\frac{1}{\epsilon} \cdot \mathsf{R}_{\frac{1}{2}-\epsilon}(g))$ holds for any total function g. We show that $\mathsf{R}^{\mathsf{prod}}_\epsilon(g)$ admits a similar error-reduction $\mathsf{R}^{\mathsf{prod}}_{1/3}(g) = \widetilde{O}(\frac{1}{\epsilon} \cdot \mathsf{R}^{\mathsf{prod}}_{\frac{1}{2}-\epsilon}(g))$. Note that from the definition of $\mathsf{R}^{\mathsf{prod}}_\epsilon$ it is not immediately clear that $\mathsf{R}^{\mathsf{prod}}_\epsilon$ admits any error-reduction at all.

We ask if our bound $\mathsf{RS}(g) = \widetilde{\Omega}(\mathsf{D}^{\mathsf{prod}}(g))$ is tight. We answer this question in the negative, by showing that for the NAND tree function, sabotage complexity is polynomially larger than $\mathsf{D}^{\mathsf{prod}}$. Our proof yields an alternative and different derivation of the tight lower bound on the bounded error randomized query complexity of the NAND tree function (originally proved by Santha in 1985), which may be of independent interest. Our result shows that sometimes, $\mathsf{R}^{\mathsf{prod}}_{1/3}$ and sabotage complexity may be useful in producing an asymptotically larger lower bound on $\mathsf{R}(f \circ g^n)$ than $\widetilde{\Omega}(\mathsf{R}(f) \cdot \mathsf{D}^{\mathsf{prod}}(g))$. In addition, this gives an explicit polynomial separation between R and $\mathsf{D}^{\mathsf{prod}}$ which, to our knowledge, was not known prior to our work.

2012 ACM Subject Classification Theory of computation → Oracles and decision trees; Theory of computation → Communication complexity; Mathematics of computing → Graph theory

Keywords and phrases Randomized query complexity, Decision Tree, Boolean function complexity, Analysis of Boolean functions

Digital Object Identifier 10.4230/LIPIcs.STACS.2024.56

Funding Supported by an ISIRD grant by Sponsored Research and Industrial Consultancy (SRIC), IIT Kharagpur.

41st International Symposium on Theoretical Aspects of Computer Science (STACS 2024).
Editors: Olaf Beyersdorff, Mamadou Moustapha Kanté, Orna Kupferman, and Daniel Lokshtanov;
Article No. 56; pp. 56:1–56:19

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

A decision tree or a query algorithm for a relation $f \subseteq \{0,1\}^n \times S$ can query various bits of an input bit string $x = (x_1, \ldots, x_n)$ in an adaptive fashion, with the goal of outputting an $s \in S$ such that $(x, s) \in f$. A randomized decision tree is assumed to have access to some source of randomness, and may choose a variable to query based on responses to previous queries, and the randomness. The complexity of a decision tree is the number of variables that it queries in the worst case. A decision tree that uses no randomness and for every x outputs an s such that $(x, s) \in f$ is called a deterministic decision tree computing f. The randomized query complexity of f for error ϵ, denoted by $\mathsf{R}_\epsilon(f)$, is the least complexity of any randomized decision tree that, for every input x, outputs s such that $(x, s) \in f$ with probability (over its own randomness) at least $1 - \epsilon$. Similarly, the deterministic query complexity of f, denoted by $\mathsf{D}(f)$, is the least complexity of any deterministic decision tree computing f. For a probability distribution μ over the domain of f, the distributional query complexity of f with respect to μ and for error ϵ, denoted by $\mathsf{D}^\mu_\epsilon(f)$, is the least complexity of any deterministic decision tree that, for a random input x sampled from μ, fails to output an s such that $(x, f) \in f$ with probability at most ϵ. Define $\mathsf{R}(f) := \mathsf{R}_{1/3}(f)$ and $\mathsf{D}^\mu(g) := \mathsf{D}^\mu_{1/3}(g)$. See Appendix A for more details about the aforementioned notions.

For a total Boolean function $g : \{0,1\}^m \to \{0,1\}$, the composition $f \circ g^n$ is the relation comprising all pairs $((x_1, \ldots, x_n), s) \in (\{0,1\}^m)^n$ such that $((g(x_1), \ldots, g(x_n)), s) \in f$.

It is easy to see that $\mathsf{D}(f \circ g^n) \leq \mathsf{D}(f) \cdot \mathsf{D}(g)$; a decision tree for $f \circ g^n$ may be constructed simply by simulating an optimal tree of f, and serving each query that it makes by solving the corresponding copy of g using an optimal tree of g. For randomized query algorithms, a similar idea works out, albeit with some additional work to handle errors, to show that $\mathsf{R}(f \circ g^n) = O(\mathsf{R}(f) \cdot \mathsf{R}(g) \cdot \log \mathsf{R}(f))$.

Composition questions ask whether the aforementioned upper bounds on the complexity of $f \circ g^n$ are asymptotically optimal. These are fundamental questions about the structure of optimal algorithms for $f \circ g^n$, and have received considerable attention in research.

It is known from the works of Montenaro [11] and Tal [19] that $\mathsf{D}(f \circ g^n) = \mathsf{D}(f) \cdot \mathsf{D}(g)$. Thus the composition question for deterministic query complexity has been completely answered. On the contrary, in spite of extensive research, a complete answer to the composition question for randomized query complexity is still lacking.

1.1 Past works on randomized query composition

Past works dealt with a more general composition question for randomized query complexity, where the inner function g is allowed to be partial. The definition of $f \circ g^n$ and the aforementioned upper bounds on $\mathsf{D}(f \circ g^n)$ and $\mathsf{R}(f \circ g^n)$ can be accordingly generalized; we are omitting the details in this paper. In 2015, Ben-David and Kothari [16] defined the sabotage complexity measure $\mathsf{RS}(g)$ of a partial Boolean function g. They showed that $\mathsf{R}(f \circ g^n) = \Omega(\mathsf{R}(f) \cdot \mathsf{RS}(g))$. They further showed that for total g, $\mathsf{RS}(g) = \widetilde{\Omega}(\sqrt{\mathsf{R}(g)})$, implying $\mathsf{R}(f \circ g^n) = \widetilde{\Omega}(\mathsf{R}(f) \cdot \sqrt{\mathsf{R}(g)})$. In 2019, Gavinsky, Lee, Santha and Sanyal [6] introduced the max-conflict complexity $\overline{\chi}(g)$ and showed that $\mathsf{R}(f \circ g^n) = \Omega(\mathsf{R}(f) \cdot \overline{\chi}(g))$. They further showed that even for partial functions g, $\overline{\chi}(g) = \Omega(\sqrt{\mathsf{R}(g)})$, implying $\mathsf{R}(f \circ g^n) = \Omega(\mathsf{R}(f) \cdot \sqrt{\mathsf{R}(g)})$. Moreover, they showed that for all partial functions g, $\overline{\chi(g)} = \Omega(\mathsf{RS}(g))$. They also demonstrated unbounded separation between $\overline{\chi}(g)$ and $\mathsf{RS}(g)$ for a partial g. In 2022, Ben-David, Blais, Göös and Maystre [4] introduced the linearized complexity measure $\mathsf{L}(g)$. They showed that for any partial g, $\mathsf{R}(f \circ g^n) = \Omega(\mathsf{R}(f) \cdot \mathsf{L}(g))$, and that L is the largest measure M for which the statement $\mathsf{R}(f \circ g^n) = \Omega(\mathsf{R}(f) \cdot \mathsf{M}(g))$ holds. They also demonstrated polynomial separation between $\mathsf{L}(g)$ and $\overline{\chi}(g)$ for a partial g.

A different line of work has focused on proving bounds on $\mathsf{R}(f \circ g^n)$ of the form $\Omega(\mathsf{M}(f) \cdot \mathsf{R}(g))$ for some complexity measure M [1, 7, 2]. In 2020 Ben-David and Blais [3] defined the noisy query complexity noisyR and showed that noisyR is the largest measure M for which the statement $\mathsf{R}(f \circ g^n) = \Omega(\mathsf{M}(f) \cdot \mathsf{R}(g))$ holds. In 2023, Chakraborty et al. [5] showed that for the special case when $\mathsf{R}(f) = \Theta(n)$, a near-perfect randomized query composition theorem $\mathsf{R}(f \circ g^n) = \widetilde{\Omega}(\mathsf{R}(f) \cdot \mathsf{R}(g))$ holds.

1.2 Our results

This work investigates the possibility of a perfect randomized query composition theorem $\mathsf{R}(f \circ g^n) = \Omega(\mathsf{R}(f) \cdot \mathsf{R}(g))$ when g is a total function. As discussed in the preceding section, past works have introduced measures $\mathsf{RS}, \overline{\chi}$ and L that are asymptotically non-decreasing in the above order. As discussed before, we also know that any two of them can be asymptotically separated. However, the Boolean functions that witness these separations are all partial, and to the best of our knowledge, no separation between these measures is known for total functions. Does one of these measures coincides with R for total functions?

Ben-David and Kothari asked in their paper whether $\mathsf{RS}(g) = \Theta(\mathsf{R}(g))$ for total g. Our first result is that for any total g, $\mathsf{RS}(g)$ is, up to a logarithmic factor, at least the maximum distributional query complexity of g for any product (over variables) distribution. Let PROD be the set of all product distributions over the domain $\{0,1\}^m$ of g. Define $\mathsf{D}^{\mathsf{prod}}(g) := \max_{\mu \in \mathsf{PROD}} \{\mathsf{D}^{\mu}(g)\}$.

▶ **Theorem 1.** *For any total function $g : \{0,1\}^m \to \{0,1\}$,*

$$\mathsf{RS}(g) = \widetilde{\Omega}(\mathsf{D}^{\mathsf{prod}}(g)).$$

Informally, the sabotage complexity captures the minimum number of randomized queries required to distinguish any pair of input strings on which the function values differ (see Section 2.3 for a formal definition). Theorem 1 shows that this task is at least as hard as deciding the function on every possible product distribution (potentially with a different query algorithm for each distribution).

Together with the composition theorem of Ben-David and Kothari, Theorem 1 immediately yields the following corollary.

▶ **Corollary 2.** *For any total function $g : \{0,1\}^m \to \{0,1\}$,*

$$\mathsf{R}(f \circ g^n) = \widetilde{\Omega}(\mathsf{R}(f) \cdot \mathsf{D}^{\mathsf{prod}}(g)).$$

The minimax theorem (Fact 17) states that $\mathsf{R}(g) = \max_{\mu} D^{\mu}(g)$, where the maximum is over all probability distributions over the domain of g. In this light Corollary 2 makes progress towards answering the randomized composition question for total inner functions. An additional motivation for our first result is that product distributions comprise a natural class of distributions that has received significant attention in Boolean function complexity research [9, 8, 10, 17].

We prove Theorem 1 by introducing a new complexity measure $\mathsf{R}^{\mathsf{prod}}_{\epsilon}$. Informally speaking, $\mathsf{R}^{\mathsf{prod}}_{\epsilon}(g)$ is the minimum complexity of any randomized decision tree with unlabelled leaves with the property that, for every product distribution μ over the inputs, the average bias of its leaves is at least $((1-\epsilon)-\epsilon)/2 = 1/2 - \epsilon$. Define $\mathsf{R}^{\mathsf{prod}}(g) := \mathsf{R}^{\mathsf{prod}}_{1/3}(g)$. See Section 2.2 for formal definitions. It follows by standard arguments that $\mathsf{R}^{\mathsf{prod}}(g) = \Omega(\mathsf{D}^{\mathsf{prod}}(g))$ (see Claim 11). Our next next result shows that RS is characterized by $\mathsf{R}^{\mathsf{prod}}$ up to a logarithmic factor.

▶ **Theorem 3.** *For all total functions* $g : \{0,1\}^m \to \{0,1\}$,
1. $\mathsf{RS}(g) = O(\mathsf{R}^{\mathsf{prod}}(g))$, *and*
2. $\mathsf{RS}(g) = \Omega(\mathsf{R}^{\mathsf{prod}}(g)/\log \mathsf{R}^{\mathsf{prod}}(g))$.

Theorem 1 follows immediately from Theorem 3(2) and the fact that $\mathsf{R}^{\mathsf{prod}}(g) = \Omega(\mathsf{D}^{\mathsf{prod}}(g))$ (Claim 11).

Since any non-trivial product distribution is supported on all of $\{0,1\}^m$, $\mathsf{R}^{\mathsf{prod}}(g)$ and $\mathsf{D}^{\mathsf{prod}}(g)$ are well-defined only for total functions g. The proof of Theorem 3 (that goes via Lemma 6 discussed later) makes important use of the totality of g. We hope that the measure $\mathsf{R}^{\mathsf{prod}}$, the characterization of RS presented in Theorem 3, and the insights acquired in our proof techniques, specially pertaining to ways of exploiting totality, will be useful in future research to resolve whether $\mathsf{RS}(g) = \Theta(\mathsf{R}(g))$ for total functions g.

In light of Theorem 3 the question whether $\mathsf{R}(g) = \Theta(\mathsf{RS}(g))$ for total functions g translates to the question whether $\mathsf{R}(g) = \widetilde{\Theta}(\mathsf{R}^{\mathsf{prod}}(g))$. We generalize this question for every error ϵ.

▶ **Question 4.** *Is it true that for every total function* $g : \{0,1\}^m \to \{0,1\}$ *and* $\epsilon : \mathbb{N} \to (0,1/2)$, $\mathsf{R}_{\epsilon(m)}(g) = \widetilde{\Theta}(\mathsf{R}^{\mathsf{prod}}_{\epsilon(m)}(g))$?

From the work of Ben-David, Blais, Göös and Maystre [4] it follows that for any error parameter ϵ, the linearized complexity measure $\mathsf{L}(g)$ of g is bounded above by $O\left(\frac{1}{\epsilon} \cdot \mathsf{R}_{\frac{1}{2}-\epsilon}(g)\right)$. As discussed before, they also show that L is the largest measure M for which the statement $\mathsf{R}(f \circ g) = \Omega(\mathsf{R}(f)\mathsf{M}(g))$ holds for all relations f and partial functions g. We thus have that for a perfect composition theorem $\mathsf{R}(f \circ g) = \Omega(\mathsf{R}(f)\mathsf{R}(g))$ to hold for any relation f and any total Boolean function g, a necessary condition is that $\mathsf{R}(g) = O(\frac{1}{\epsilon} \cdot \mathsf{R}_{\frac{1}{2}-\epsilon}(g))$ holds for any total Boolean function g. In light of Question 4, we may ask if $\mathsf{R}^{\mathsf{prod}}_{\epsilon}$ admits a similar error reduction. Our next result answers this question in the affirmative (up to a logarithmic factor).

▶ **Theorem 5.** *For every total function* $g : \{0,1\}^m \in \{0,1\}$ *and* $\epsilon : \mathbb{N} \to (0,1/2)$,

$$\mathsf{R}^{\mathsf{prod}}(g) = \widetilde{O}\left(\frac{1}{\epsilon(m)} \cdot \mathsf{R}^{\mathsf{prod}}_{\frac{1}{2}-\epsilon(m)}(g)\right).$$

We remark here that from the definition of $\mathsf{R}^{\mathsf{prod}}_{\epsilon}$ it is not immediately clear that it admits any error-reduction at all.

To prove Theorems 3 and 5, we define a version of sabotage complexity with errors, that we denote by RS_{ϵ}. Informally, $\mathsf{RS}_{\epsilon}(g)$ is the minimum number of randomized queries required to distinguish every pair of inputs with different function values with probability at least $1-\epsilon$ (see Section 2.3 for a formal definition). Let $s(g)$ denote the sensitivity of g (see Section 2.1 for a formal definition). The following lemma constitutes the technical core of our proofs of Theorems 3 and 5.

▶ **Lemma 6.** *For all total Boolean functions* $g : \{0,1\}^n \to \{0,1\}$, *and* $\epsilon : \mathbb{N} \to (0,1/2)$,
1. $\mathsf{R}^{\mathsf{prod}}(g) = O\left(\frac{1}{\epsilon(n)} \cdot \mathsf{RS}_{1-\epsilon(n)}(g) \log s(g)\right)$, *and*
2. $\mathsf{RS}_{1-2\epsilon(n)}(g) \leq \mathsf{R}^{\mathsf{prod}}_{\frac{1}{2}-\epsilon(n)}(g)$.

Is the bound in Theorem 1 tight? Our next result gives a negative answer to this question. We show that for the NAND tree function (defined shortly), RS and $\mathsf{D}^{\mathsf{prod}}$ are polynomially separated. Consider a complete binary tree of depth d. Each leaf is labelled by a distinct Boolean variable. Each internal node is a binary NAND gate. For each input, the evaluation of this Boolean formula is the output of the NAND tree function, that we denote by g_d.

▶ **Theorem 7.** $\mathsf{D}^{\mathsf{prod}}(g_d) = O(\mathsf{RS}(g_d)^{1-\Omega(1)})$.

Saks and Wigderson [14] showed that the zero-error randomized query complexity of g_d is $\Theta(\alpha^d)$ for $\alpha = \frac{1+\sqrt{33}}{4}$. Later Santha [15] showed that $\mathsf{R}(g_d) = \Theta(\alpha^d)$. We prove Theorem 7 in two parts. First, we show an upper bound of $O((\alpha - \Omega(1))^d)$ on $\mathsf{D}^{\mathsf{prod}}(g_d)$.

▶ **Lemma 8.** *There exists a constant $\delta > 0$ such that $\mathsf{D}^{\mathsf{prod}}(g_d) = O((\alpha - \delta)^d)$.*

Works by Pearls [13] and Tarsi [19] showed that there exists a constant $\eta > 0$ such that for all distributions μ where each variable is set to 1 independently with some probability p, $\mathsf{D}^{\mu}(g_d) = O((\alpha - \eta)^d)$. In Lemma 8 we bound $\mathsf{D}^{\mu}(g_d)$ for any product distribution μ. Our bound is quantitatively weaker than those by Pearls [13] and Tarsi [19], and we do not comment on its tightness.

Lemma 8 also gives an explicit polynomial separation between R and $\mathsf{D}^{\mathsf{prod}}$ which, to our knowledge, was not known prior to our work.

Next, we prove a tight lower bound on $\mathsf{RS}(g_d)$. As a by-product, our proof of the following lemma yields a different proof of the bound $\mathsf{R}(g_d) = \Omega(\alpha^d)$ from the one by Santha [15], and may be of independent interest.

▶ **Lemma 9.** $\mathsf{RS}(g_d) = \Omega(\alpha^d)$.

Lemma 9, together with the upper bound by Saks and Wigderson, shows that $\mathsf{RS}(g_d) = \Theta(\mathsf{R}(g_d))$. From the composition theorem proven by Ben-David and Kothari, we thus have that for all relations f, $\mathsf{R}(f \circ g_d) = \Theta(\mathsf{R}(f) \cdot \mathsf{R}(g_d))$.

Lemmata 8 and 9 immediately imply Theorem 7.

1.3 Proof ideas

In this section, we sketch the ideas and techniques that have gone into the proofs of our results. We begin with Lemma 6, from which Theorems 3 and 5 follow. We then discuss Lemmata 8 and 9, from which Theorem 7 follows.

Lemma 6

We first discuss part 2, which is easier. Note that if a randomized algorithm $\mathcal{R}$ decides g on each input with error probability $\frac{1}{2} - \epsilon$, then by a union bound two simultaneous runs of $\mathcal{R}$ on $x \in g^{-1}(0), y \in g^{-1}(1)$ decide both $g(x)$ and $g(y)$ with error probability $1 - 2\epsilon$. This implies that $\mathcal{R}$ distinguishes x and y with error probability $1 - 2\epsilon$.

Now we turn to part 1. This step needs arguments involving sensitivity and influence of Boolean functions, that are defined and discussed in Section 2.1. The first step is showing that distinguishing each pair of inputs with high confidence is equivalent to reading each sensitive bit of each input with the same confidence (Lemma 14). Using this, by a sequence of arguments involving standard error-reduction, we infer that there is a randomized tree $\mathcal{R}$ of complexity $O\left(\frac{1}{\epsilon(n)} \cdot \mathsf{RS}_{1-\epsilon(n)}(g) \log s(g)\right)$ that, for every input, with probability $1 - \frac{1}{s(g)}$, queries all its sensitive bits. This translates to the claim that the average influence of the restrictions of g to the leaves of $\mathcal{R}$ is low. Poincaré inequality (Lemma 10) now lets us conclude that that the average bias of those restrictions is small, yielding the lemma.

Lemma 8

Saks and Wigderson gave a zero-error recursive algorithm for g_d. Their algorithm recursively evaluates a randomly chosen child of the root. If that child evaluates to 0, the algorithm outputs 1 and terminates. Else, the algorithm recursively evaluates the other child and outputs the complement.

If the output of the function is 0, then the algorithm will be forced to evaluate both children. However, if the output is 1, then the algorithm avoids evaluating one of the children with probability 1/2.

We observe that if the inputs are sampled from a product distributions, then firstly, the output will not always be 0; so we will always have scope to avoid evaluating one child. Secondly, we will also have both children evaluating to 0 with positive probability, in which case we are guaranteed to save evaluating one child.

We modify the algorithm by Saks an Wigderson to tap these opportunities; in each step we query the child which is more likely to evaluate to 0. Note that this requires knowledge of the distribution. We look at two successive levels of the tree and show that the above considerations bring us significant advantage over the algorithm by Saks and Wigderson.

Lemma 9

As mentioned before, here we work with the original definition of sabotage complexity. Our proof splits into the following steps.

1. We recursively define a "hard" distribution $\mathcal{P}_d$ over pairs in $g_d^{-1}(0) \times g_d^{-1}(1)$.
2. We consider an arbitrary zero-error randomized algorithm $\mathcal{R}$ for g_d. We now wish to give a lower bound on the number of queries it makes on expectation to distinguish a random pair sampled from $\mathcal{P}_d$.
3. Using $\mathcal{R}$, we recursively define a sequence of algorithms $\mathcal{A}_d, \mathcal{A}_{d-1}, \ldots, \mathcal{A}_0$ such that for each i, $\mathcal{A}_i$ is a zero-error algorithm for g_i.
4. We establish a recursive relation amongst the expected number of queries that g_i makes to distinguish a pair sampled from $\mathcal{P}_i$, for various i. We make a distinction between queries based on their answers (0 or 1). This step involves a small case analysis involving all deterministic trees with two variables.
5. We finish by solving the recursion established in the previous step.

2 Preliminaries

Refer to Appendix A for some notations and definitions that will be used throughout the paper.

2.1 Sensitivity and influence

For a total Boolean function g, a variable x_i is said to be sensitive for an input x if $g(x) \neq g(x^{\oplus i})$. The sensitivity of x with respect to g, denoted by $s(g, x)$, is the number of sensitive bits of x, i.e., $|\{i \in [n] \mid g(x) \neq g(x^{\oplus i})\}|$. The sensitivity of g, denoted by $s(g)$, is the maximum sensitivity of any input x with respect to g, i.e.,

$$s(g) = \max_{x \in \{0,1\}^m} s(g, x).$$

For a product distribution $\mu \in \mathsf{PROD}$ given by parameters $p_1, \ldots, p_m$, the *influence* of x_i with respect to g and μ is defined as

$$\mathsf{Inf}_i(g) := 4p_i(1 - p_i) \Pr_{x \sim \mu}[g(x) \neq g(x^{\oplus i})],$$

and the influence of g with respect to μ is defined as

$$\mathsf{Inf}(g) = \sum_{i=1}^{m} \mathsf{Inf}_i(g).$$

The following inequality follows easily from the above definitions, linearity of expectation, and the observation that $4p_i(1-p_i) \leq 1$ for all $p_i \in [0,1]$.

$$\mathsf{Inf}(g) \leq \mathbf{E}_{x\sim\mu} s(g,x). \tag{1}$$

Let $\mathsf{Var}(g)$ denote the variance of the random variable $g(x)$ when x is drawn from μ. The Poincaré inequality bounds $\mathsf{Var}(g)$ in terms of $\mathsf{Inf}(g)$.

▶ **Lemma 10** (Poincaré inequality). *For every product distribution μ, $4\mathsf{Var}(g) \leq \mathsf{Inf}(g)$.*

A proof of the Poincaré inequality may be found in [12].
In the notations $\mathsf{Inf}_i, \mathsf{Inf}$ and Var, the dependence on μ is suppressed. μ will be clear from the context.

2.2 Randomized query complexity for product distributions

Let μ be a product distribution, and T be a deterministic decision tree. For each leaf ℓ of T, let p_ℓ^μ be the probability that the computation of T on an input drawn from μ reaches ℓ. Let $\mathbf{p}^\mu$ denote the probability distribution $(p_\ell^\mu)_\ell$ over the leaves of T. We say that a randomized decision tree $\mathcal{R}$ computes g with error ϵ for product distributions if for every product distribution $\mu \in \mathsf{PROD}$,

$$\mathbf{E}_{T\sim\mathcal{R}}\mathbf{E}_{\ell\sim\mathbf{p}^\mu}[\min\{\Pr_{x\in\mu|_\ell}[g(x)=0], \Pr_{x\in\mu|_\ell}[g(x)=1]\}] \leq \epsilon,$$

where the inner expectation is over the leaves of T. We define $\min\{\Pr_{x\in\mu|_\ell}[f(x)=0], \Pr_{x\in\mu|_\ell}[f(x)=1]\}$ to be 0 if $p_\ell^\mu = 0$; the conditional distribution $\mu|_\ell$ is not defined in this case. The *randomized query complexity* of g for product distribution for error ϵ, denoted by $\mathsf{R}_\epsilon^{\mathsf{prod}}(f)$, is the minimum query complexity of a randomized decision tree $\mathcal{R}$ that satisfies the above condition. Define $\mathsf{R}^{\mathsf{prod}}(f) := \mathsf{R}_{1/3}^{\mathsf{prod}}(f)$.

Note that in the above definition, no reference has been made to the labels of the leaves of T. For the purpose of this definition, $\mathcal{R}$ can be thought of as a probability distribution over trees whose leaves are unlabelled.

The following claim shows that $\mathsf{D}_\epsilon^{\mathsf{prod}}(g)$ is bounded above by $\mathsf{R}_\epsilon^{\mathsf{prod}}(g)$. A proof may be found in Section B.

▷ Claim 11. For every Boolean function g and parameter $\epsilon \in [0, 1/2]$, $\mathsf{D}_\epsilon^{\mathsf{prod}}(g) \leq \mathsf{R}_\epsilon^{\mathsf{prod}}(g)$.

2.3 Sabotage complexity

The *sabotage complexity* of a Boolean function g for error ϵ, denoted by $\mathsf{RS}_\epsilon(g)$, is defined to be the minimum query complexity of any randomized decision tree $\mathcal{R}$ for which the following is true: For every $x = (x_1, \ldots, x_m) \in g^{-1}(0)$, $y = (y_1, \ldots, y_m) \in g^{-1}(1)$, with probability at least $1-\epsilon$, a decision tree T drawn from $\mathcal{R}$ when run on x queries an index i such that $x_i \neq y_i$[1]. Define $\mathsf{RS}(g) := \mathsf{RS}_{1/3}(g)$.

Sabotage complexity was defined by Ben-David and Kothari [16]. They defined the measure as the minimum expected query complexity of any randomized decision tree to distinguish each pair of inputs $x \in g^{-1}(0), y \in g^{-1}(1)$. However, as the authors observed, the definition stated above is within a constant factor of the original definition in [16]. See more discussion on this in Section 5 where we work with the original definition.

[1] Note that T queries an index i such that $x_i \neq y_i$ when run on x if and only if T queries an index j such that $x_j \neq y_j$ when run on y.

The following fact can be proven by standard BPP amplification.

▶ **Fact 12.** *$\forall \epsilon, \epsilon' \in (0,1)$ and $\epsilon < \epsilon'$, $\mathsf{RS}_\epsilon(g) = O\left(\mathsf{RS}_{\epsilon'}(g) \cdot \frac{\log(1/\epsilon)}{\log(1/\epsilon')}\right)$.*

Ben-David and Kothari proved that the sabotage complexity is lower bounded by many complexity measures that are studied in the context of decision trees. In particular, $\mathsf{RS}(g)$ is lower bounded by $s(g)$.

▶ **Fact 13** ([16]). *For all Boolean function $g : \{0,1\}^m \to \{0,1\}$, $\mathsf{RS}(g) = \Omega(s(g))$.*

3 Sabotage complexity and product distributions

In this section we first prove Lemma 6. We then use Lemma 6 to prove Theorems 3 and 5. The following lemma says that to distinguish each pair of inputs on which the function values differ with high probability, it is necessary and sufficient to query each sensitive bit of each input with high probability.

▶ **Lemma 14.** *Let $g : \{0,1\}^n \to \{0,1\}$ be a total Boolean function. Then, $\mathsf{RS}_\epsilon(g) \leq r$ if and only if there is a randomized decision tree $\mathcal{R}$ of query complexity at most r such that for each input x and each variable x_i sensitive for x, $\Pr_{T\sim\mathcal{R}}[T$ does not query x_i when run on $x] \leq \epsilon$.*

Proof.

(If) Let $\mathcal{R}$ be a randomized decision tree of complexity at most r such that for every input x and every variable x_i sensitive for x, $\Pr_{T\sim\mathcal{R}}[T$ does not query x_i when run on $x] \leq \epsilon$. We will show that $\mathcal{R}$ fails to distinguish any pair $w \in g^{-1}(0), y \in g^{-1}(1)$ with probability at most ϵ. Fix such a pair w, y. let $B = \{i_1, \ldots, i_k\}$ be the positions where w and y differ. Define $B_0 := \emptyset$ and for $1 \leq j \leq k$, define $B_j := \{i_1, \ldots, i_j\}$. Let m be the smallest index such that $g(w^{\oplus B_m}) = 1$. Thus, variable w_m is sensitive for $w^{\oplus B_{m-1}}$ and $w^{\oplus B_m}$. Now, observe that if T does not query any variable w_{i_j} with $i_j \in B$ when run on w, then T does not query w_m when run on $w^{\oplus B_{m-1}}$. By our assumption about $\mathcal{R}$, the probability of this happening when T is sampled from $\mathcal{R}$ is at most ϵ.

(Only if) Let $\mathsf{RS}_\epsilon(g) \leq r$. Thus there exists a randomized decision tree $\mathcal{R}$ of query complexity r that fails to distinguish each pair $w \in g^{-1}(0), y \in g^{-1}(1)$ with probability at most ϵ. Without loss of generality, assume that $x \in g^{-1}(0)$. Then $x^{\oplus i} \in g^{-1}(1)$. Since distinguishing x and $x^{\oplus i}$ is equivalent to querying x_i when run on x, the proof follows. ◀

Now we proceed to proving Lemma 6. For convenience, we use ϵ for $\epsilon(n)$ throughout the following proof.

Proof of Lemma 6.

Part 1. Let $\mathsf{RS}_{1-\epsilon}(g) = r$. By Lemma 14, there exists a randomized query algorithm $\mathcal{R}$ of complexity at most r such that for each input x and each variable x_i sensitive for x, $\Pr_{T\sim\mathcal{R}}[T$ does not query x_i when run on $x] \leq 1-\epsilon$. Let $\mathcal{R}'$ be the algorithm obtained by repeating $\mathcal{R}$ $\frac{2}{\epsilon}\ln s(g)$ times with independent randomness. Thus for each input x and each variable x_i sensitive for x, we have that $\Pr_{T\sim\mathcal{R}'}[T$ does not query x_i when run on x$] \leq (1-\epsilon)^{(\frac{1}{\epsilon}\cdot 2\ln s(g))} \leq \frac{1}{s(g)^2}$, where we have used the inequality $1-x \leq e^{-x}$ for all $x \in (-\infty, \infty)$. Again for each input x, by a union bound over all variables x_i sensitive for x, we have that the probability that a deterministic tree sampled from $\mathcal{R}'$ does not query all variables sensitive for x when run on x, is at most $\frac{s(g,x)}{s(g)^2} \leq \frac{1}{s(g)}$. The query complexity of $\mathcal{R}'$ is $O(\frac{1}{\epsilon} \cdot \mathsf{RS}_{1-\epsilon}(g) \log s(g))$. We will show that $\mathcal{R}'$ computes g with error $1/3$ for product distributions. This will complete the proof of this part.

To this end, fix a product distribution μ. For any deterministic decision tree T and input x of f, define

$$\mathsf{Q}(T,x) = \begin{cases} 1 & \text{if } T \text{ does not query all sensitive variables of } x \text{ when run on } x, \\ 0 & \text{otherwise.} \end{cases}$$

By the property of $\mathcal{R}'$, for every input x, we have that

$$\mathbf{E}_{T\sim\mathcal{R}'}[\mathsf{Q}(T,x)] = \Pr_{T\sim\mathcal{R}'}[\mathsf{Q}(T,x)=1] \leq \frac{1}{s(g)}.$$

Since the above is true for each x, we have the following for a random input x sampled from μ.

$$\mathbf{E}_{T\sim\mathcal{R}'}\mathbf{E}_{x\sim\mu}[\mathsf{Q}(T,x)] \leq \frac{1}{s(g)}. \tag{2}$$

For each leaf ℓ of T , let p^μ_ℓ be the probability that the computation of T on an input drawn from μ reaches ℓ and $\mathbf{p}^\mu$ denote the probability distribution $(p^\mu_\ell)_\ell$ over the leaves of T. We rewrite (2) as follows.

$$\mathbf{E}_{T\sim\mathcal{R}'}\mathbf{E}_{\ell\sim p^\mu}\mathbf{E}_{x\sim\mu|_\ell}[\mathsf{Q}(T,x)] \leq \frac{1}{s(g)}, \tag{3}$$

treating $\mathbf{E}_{x\sim\mu|_\ell}[\mathsf{Q}(T,x)]$ as 0 if $p^\mu_\ell = 0$. Now, fix an arbitrary leaf ℓ of T such that $p^\mu_\ell > 0$, and consider the Boolean function $g\,|_\ell$. Note that for any $x \in \ell$, if $\mathsf{Q}(T,x) = 0$, then $s(g\,|_\ell, x) = 0$. We thus have that

$$\mathbf{E}_{x\sim\mu|_\ell}[s(g\,|_\ell,x)] \leq \Pr_{x\sim\mu|_\ell}[\mathsf{Q}(T,x)=1]\cdot s(g\,|_\ell) \leq \mathbf{E}_{x\sim\mu|_\ell}[\mathsf{Q}(T,x)]\cdot s(g). \tag{4}$$

Since μ is a product distribution and ℓ is a subcube, $\mu\,|_\ell$ is also a product distribution. Equations (1) and (4) thus imply that

$$\mathsf{Inf}(g\,|_\ell) \leq \mathbf{E}_{x\sim\mu|_\ell}[\mathsf{Q}(T,x)]\cdot s(g). \tag{5}$$

Together with Poincaré inequality (Lemma 10), (5) implies that

$$\mathsf{Var}(f\,|_\ell) \leq \frac{1}{4}\cdot\mathbf{E}_{x\sim\mu|_\ell}[\mathsf{Q}(T,x)]\cdot s(f). \tag{6}$$

Now, for a random variable X taking value in $\{0,1\}$, $\mathsf{Var}(X) = 4\Pr[X=0]\Pr[X=1] \geq 2\min\{\Pr[X=0],\Pr[X=1]\}$ (since $\max\{\Pr[X=0],\Pr[X=1]\} \geq \frac{1}{2}$). Since ℓ is an arbitrary leaf, we have by Equations (6) and (3) that

$$\begin{aligned}
&\mathbf{E}_{T\sim\mathcal{R}'}\mathbf{E}_{\ell\sim p^\mu}[\min\{\Pr_{x\sim\mu|_\ell}[g(x)=0],\ \Pr_{x\sim\mu|_\ell}[g(x)=1]\}] \\
\leq&\frac{1}{2}\cdot\mathbf{E}_{T\sim\mathcal{R}'}\mathbf{E}_{\ell\sim p^\mu}[\mathsf{Var}(g\,|_\ell)] && \text{by the above discussion} \\
\leq&\frac{1}{8}\cdot\mathbf{E}_{T\sim\mathcal{R}'}\mathbf{E}_{\ell\sim p^\mu}\mathbf{E}_{x\sim\mu|_\ell}[\mathsf{Q}(T,x)]\cdot s(g) && \text{by Equation (6)} \\
\leq&\frac{1}{8} < \frac{1}{3}. && \text{by Equation (3)}
\end{aligned}$$

Since μ is an arbitrary product distribution, we have that $\mathcal{R}'$ computes g with error $1/3$ for product distributions.

Part 2. Fix a randomized query algorithm $\mathcal{R}$ that attains $\mathsf{R}^{\mathsf{prod}}_{\frac{1}{2}-\epsilon}(g)$. We will show that $\mathcal{R}$ also attains $\mathsf{RS}_{1-2\epsilon}(g)$. By Lemma 14 it is sufficient to show that for each input x and each variable x_i sensitive for x, $\Pr_{T\sim\mathcal{R}}[T \text{ does not query } x_i \text{ when run on } x] \leq 1-2\epsilon$. To this end, fix an input x and a variable x_i sensitive for x. Now consider the distribution μ that places a probability mass of $1/2$ on x and places the remaining mass of $1/2$ on $x^{\oplus i}$. Note that μ is a product distribution. Thus from the property of R we have that

$$\mathbf{E}_{T\sim\mathcal{R}}\mathbf{E}_{\ell\sim\mathbf{p}^\mu}[\min\{\Pr_{x\in\mu|_\ell}[g(x)=0], \Pr_{x\in\mu|_\ell}[g(x)=1]\}] \leq \frac{1}{2}-\epsilon. \tag{7}$$

Now if T does not query x_i when run on x, then T has a leaf ℓ that contains both x and x^i, $p^\mu_\ell = 1$, and for all other leaves ℓ' of T, $p^\mu_{\ell'} = 0$. Furthermore, $\min\{\Pr_{x\in\mu|_\ell}[g(x)=0], \Pr_{x\in\mu|_\ell}[g(x)=1]\} = 1/2$. Thus, $\mathbf{E}_{\ell\sim\mathbf{p}^\mu}[\min\{\Pr_{x\in\mu|_\ell}[g(x)=0], \Pr_{x\in\mu|_\ell}[g(x)=1]\}] = 1/2$. We thus have that,

$$\begin{aligned}&\mathbf{E}_{T\sim\mathcal{R}}\mathbf{E}_{\ell\sim\mathbf{p}^\mu}[\min\{\Pr_{x\in\mu|_\ell}[g(x)=0], \Pr_{x\in\mu|_\ell}[g(x)=1]\}]\\ &\geq \Pr_{T\sim\mathcal{R}}[T \text{ does not query } x_i \text{ when run on } x]\cdot\frac{1}{2}.\end{aligned} \tag{8}$$

Equations (7) and (8) imply that

$$\begin{aligned}&\Pr_{T\sim\mathcal{R}}[T \text{ does not query } x_i \text{ when run on } x]\cdot\frac{1}{2} \leq \frac{1}{2}-\epsilon\\ \implies &\Pr_{T\sim\mathcal{R}}[T \text{ does not query } x_i \text{ when run on } x] \leq 1-2\epsilon.\end{aligned}$$

This completes the proof. ◀

Now we prove Theorems 3 and 5.

Proof of Theorem 3 Part 1. Substituting $\epsilon(n) = 1/6$ in part 2 of Lemma 6 we have that

$$\mathsf{R}^{\mathsf{prod}}(g) \geq \mathsf{RS}_{2/3} = \Omega(\mathsf{RS}(g)),$$

where the second equality follows from Fact 12. ◀

Theorem 3 (1) and Fact 13 imply that

$$\mathsf{R}^{\mathsf{prod}}(g) = \Omega(s(g)). \tag{9}$$

Proof of Theorem 3 Part 2.

$$\begin{aligned}\mathsf{RS}(g) = \mathsf{RS}_{1/3}(g) &\geq \mathsf{RS}_{2/3}(g)\\ &= \Omega(\mathsf{R}^{\mathsf{prod}}(g)/\log s(g)) && \text{by Lemma 6 part 1 with } \epsilon(n)=1/3\\ &= \Omega(\mathsf{R}^{\mathsf{prod}}(g)/\log \mathsf{R}^{\mathsf{prod}}(g)) && \text{by Equation (9)}\end{aligned}$$

◀

Proof of Theorem 5. We have

$$\begin{aligned}\mathsf{R}^{\mathsf{prod}}(g) &= O\left(\frac{1}{\epsilon(n)}\cdot\mathsf{RS}_{1-\epsilon(n)}(g)\log s(g)\right) && \text{Part (1) of Lemma 6}\\ &= O\left(\frac{1}{\epsilon(n)}\cdot\mathsf{RS}_{1-2\epsilon(n)}(g)\log s(g)\right) && \text{since } 1-\epsilon(n)\geq 1-2\epsilon(n)\\ &= O\left(\frac{1}{\epsilon(n)}\cdot\mathsf{R}^{\mathsf{prod}}_{\frac{1}{2}-\epsilon(n)}(g)\log s(g)\right) && \text{by part (2) of Lemma 6}\\ &= O\left(\frac{1}{\epsilon(n)}\cdot\mathsf{R}^{\mathsf{prod}}_{\frac{1}{2}-\epsilon(n)}(g))\log \mathsf{R}^{\mathsf{prod}}(g)\right). && \text{by Equation (9)}\end{aligned}$$

◀

4 Separation between $\mathsf{D}^{\mathsf{prod}}$ and R

In this section we prove Lemma 8. Recall that g_d denotes the NAND tree function of depth d. Snir [18] and Saks and Wigderson [14] were the first to study g_d in the context of randomized query complexity. As mentioned in Section 1, it is known from the works of Saks and Wigderson [14] and Santha [15] that $\mathsf{R}(g_d) = \Theta(d^\alpha)$ where $\alpha = \frac{1+\sqrt{33}}{4}$.

For a distribution μ, the the zero-error distributional complexity of a Boolean function g, that we denote by $\overline{\mathsf{D}_0^\mu(g)}$, is the least expected number of queries made by any (deterministic) tree T on a random input sampled from μ. Define $\overline{\mathsf{D}_0^{\mathsf{prod}}}(g) := \max_{\mu \in \mathsf{PROD}} \overline{\mathsf{D}_0^\mu(g)}$. By Markov's inequality, it follows that $\mathsf{D}^{\mathsf{prod}}(g) = O(\overline{\mathsf{D}_0^{\mathsf{prod}}}(g))$.

Proof of Lemma 9. We will prove an upper bound on $\overline{\mathsf{D}_0^{\mathsf{prod}}}(g)$. By the preceding discussion, that will prove the lemma.

Let μ be any product distribution over $\{0,1\}^n$. Define $T(d,\mu) := \overline{\mathsf{D}_0^\mu}(g_d)$ and $T(d) := \overline{\mathsf{D}_0^{\mathsf{prod}}}(g_d)$. Consider the query algorithm $\mathcal{A}_\mu$[2] given in Algorithm 1.

Algorithm 1 $\mathcal{A}_\mu(x)$.

Input: Query access to $x = (x_1, \ldots, x_{2^d})$.
$g \leftarrow g_d$.
if *g is a variable* **then**
  Query g. Return the outcome of the query.
end
else
  Let g_ℓ and g_r respectively be the left and right subtrees of g.
  Let μ_ℓ and μ_r respectively be the product distributions induced on the input spaces of g_ℓ and g_r by μ.
  $t \leftarrow \arg\max_{i \in \{\ell,r\}} \Pr_{y \sim \mu_i}[g_i(y) = 0]$.
  $s \leftarrow \{\ell, r\} \setminus \{t\}$.
  For $i \in \{\ell, r\}$, let $x^{(i)}$ be the input to g_i.
  if $\mathcal{A}_{\mu_t}(x^{(t)}) = 0$ **then**
    return 1.
  end
  else
    return $\overline{\mathcal{A}_{\mu_s}(x^{(s)})}$.
  end
end

$\mathcal{A}_\mu$ works as follows: if $d = 1$, i.e., if g_d is a single variable, then $\mathcal{A}_\mu$ queries and returns the value of the variable. Else, $\mathcal{A}_\mu$ recursively evaluates a subtree of the root of g_d whose probability of evaluating to 0 is at least that of the other subtree.[3] If the recursive call returns 0, $\mathcal{A}_\mu$ returns 1. Else, $\mathcal{A}_\mu$ recursively evaluates the other subtree of the root of g_d and returns the complement of the value returned by that recursive call. It is clear that on every input, $\mathcal{A}_\mu$ returns the correct answer with probability 1.

[2] Note that $\mathcal{A}_\mu$ needs the knowwledge of μ.
[3] By "recursively evaluates" we mean that $\mathcal{A}_\mu$ invokes $\mathcal{A}_{\mu'}$ for the distribution μ' induced by μ on the domain of the subfunction under consideration.

Now we analyze the query complexity of $\mathcal{A}$. For $i \in \{\ell, r\}$, define $p_i := \Pr_{x^{(i)} \sim \mu_i}[g_d(x^{(i)}) = 0]$. WLOG assume that $p_\ell \geq p_r$. $\mathcal{A}_\mu$ on input x will recursively evaluate g_ℓ by invoking $\mathcal{A}_{\mu_\ell}$ on input $x^{(\ell)}$. If the recursive call returns 1, then $\mathcal{A}$ will recursively evaluate g_r by invoking $\mathcal{A}_{\mu_r}$ on input $x^{(r)}$. We thus have that,

$$T(d, \mu) = T(d-1, \mu_\ell) + (1 - p_\ell)T(d-1, \mu_r). \tag{10}$$

Let α_ℓ, α_r respectively be the probabilities that the left and right children of g_ℓ evaluate to 0. Similarly, let β_ℓ, β_r respectively be the probabilities that the left and right children of g_r evaluate to 0. Without loss of generality assume that $\alpha_\ell \geq \alpha_r, \beta_\ell \geq \beta_r$ and $\alpha_\ell \leq \beta_\ell$ (other cases are similar). By using a similar analysis as above and then upper bounding distributional query complexity for specific product distributions by the product distributional complexity we have that,

$$T(d-1, \mu_\ell) \leq T(d-2) + (1 - \alpha_\ell)T(d-2), \text{ and} \tag{11}$$

$$\begin{aligned} T(d-1, \mu_r) &\leq T(d-2) + (1 - \beta_\ell)T(d-2) \\ &\leq T(d-2) + (1 - \alpha_\ell)T(d-2). \end{aligned} \tag{12}$$

Substituting Equations (11) and (12) in (10) we have that

$$\begin{aligned} T(d) &\leq T(d, \mu) \\ &\leq (2 - \alpha_\ell)(2 - p_\ell)T(d-2). \end{aligned} \tag{13}$$

Now, we have that $p_\ell = (1 - \alpha_r)(1 - \alpha_\ell) \geq (1 - \alpha_\ell)^2$.[4] Substituting in Equation (13) we have that

$$\begin{aligned} T(d) &\leq (2 - \alpha_\ell)(2 - (1 - \alpha_\ell)^2)T(d-2) \\ &= (2 - \alpha_\ell)(1 + 2\alpha_\ell - \alpha_\ell^2)T(d-2). \end{aligned} \tag{14}$$

The maximum value of the function $f(x) := (2 - x)(1 + 2x - x^2)$ in the domain $[0, 1]$ is $\frac{2}{27} \cdot (17 + 7\sqrt{7})$. From Equation (14) we have that

$$T(d) = O\left(\sqrt{\frac{2}{27} \cdot (17 + 7\sqrt{7})}\right)^d = O(\alpha - \delta)^d \text{ for some constant } \delta > 0. \quad \blacktriangleleft$$

5 Sabotage complexity of NAND tree

In this section, we prove Lemma 9. Recall that g_d stands for the NAND tree function of depth d. Define $g_0(b) = b$ for $b \in \{0, 1\}$.

For a randomized query algorithm $\mathcal{R}$ that decides $g : \{0,1\}^m \to \{0,1\}$ with error probability 0, and for inputs x, y such that $g(x) = 0, g(y) = 1$, define the expected sabotage complexity of $\mathcal{R}$ on the pair x, y, denoted by $\mathsf{RS}_E(\mathcal{R}, x, y)$, to be the expected number of queries that $\mathcal{R}$ makes until (and including) it queries an index i such that $x_i \neq y_i$ when run on x (or y). Define the expected sabotage complexity $\mathsf{RS}_E(\mathcal{R})$ to be $\max_{x,y \in \{0,1\}^m, g(x)=0, g(y)=1} \mathsf{RS}_E(\mathcal{R}, x, y)$, and the expected sabotage complexity $\mathsf{RS}_E(g)$ to be the minimum $\mathsf{RS}_E(\mathcal{R})$ for any randomized query algorithm $\mathcal{R}$ that decided g with error probability 0. As observed by Ben-David and Kothari, $\mathsf{RS}_E(g) = \Theta(\mathsf{RS}(g))$. In this section, we will work with RS_E in place of $\mathsf{RS}(g)$.

[4] Here we use that μ is a product distribution.

It follows by standard arguments that for every distribution $\mathcal{D}$ on $g^{-1}(0) \times g^{-1}(1)$ there exists a zero-error randomized (even deterministic) decision tree $\mathcal{R}$ of g such that $\mathbf{E}_{(x,y)\sim\mathcal{D}}[\mathsf{RS}_E(|R,x,y)] \leq \mathsf{RS}_E(g)$. To prove Lemma 9 it thus suffices to exhibit a hard distribution $\mathcal{D}$ on $g^{-1}(0) \times g^{-1}(1)$ such that for every zero-error randomized tree $\mathcal{R}$ of g, $\mathbf{E}_{(x,y)\sim\mathcal{D}}[\mathsf{RS}_E(|\mathcal{R},x,y)]$ is large. The first step in our proof is to define a hard distribution.

A hard distribution

We define a probability distribution $\mathcal{P}_d$ on ${g_d}^{-1}(0) \times {g_d}^{-1}(1)$ as follows. Define $\mathcal{P}_0$ to be the point distribution $\{(0,1)\}$. For $d \geq 1$, $\mathcal{P}_d$ is defined recursively by the following sampling procedure. Let $n := 2^{d-1}$.

1. Sample $(x,y) \sim \mathcal{P}_{d-1}$. Let $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$.
2. Sample $\overline{b} := (b_1,\ldots,b_n)$ uniformly at random from $\{0,1\}^n$.
3. For each $i = 1,\ldots,n$, let $u_i = (u_i^{(0)}, u_i^{(1)}), v_i = (v_i^{(0)}, v_i^{(1)}) \in \{0,1\}^2$ be defined as follows:
   a. If $(x_i,y_i) = (0,0)$, set $u_i, v_i \leftarrow (1,1)$.
   b. If $(x_i,y_i) = (0,1)$, set $u_i \leftarrow (1,1)$ and set $v_i \leftarrow (b_i, 1-b_i)$.
   c. If $(x_i,y_i) = (1,0)$, set $u_i \leftarrow (b_i, 1-b_i)$ and set $v_i \leftarrow (1,1)$.
   d. If $(x_i,y_i) = (1,1)$, set $u_i, v_i \leftarrow (b_i, 1-b_i)$.
4. Let x' be the string obtained from x by replacing each x_i by u_i. Similarly let y' be the string obtained from y by replacing each y_i by v_i.
5. Return (x',y').

Notice that for each $i = 1,\ldots,n$, $x_i = \mathsf{NAND}(u_i^{(1)}, u_i^{(2)})$ and $y_i = \mathsf{NAND}(v_i^{(1)}, v_i^{(2)})$. Hence, $g_d(x') = g_{d-1}(x)$ and $g_d(y') = g_{d-1}(y)$. Thus we inductively establish that $\mathcal{P}_d$ is supported on ${g_d}^{-1}(0) \times {g_d}^{-1}(1)$. The following observation can be verified to be true by a simple case analysis.

▶ **Observation 15.** *For each $i = 1,\ldots,n$, $x_i = \overline{u_i^{(b_i)}}$ and $y_i = \overline{v_i^{(b_i)}}$. Furthermore, $u_i^{(1-b_i)} = v_i^{(1-b_i)} = 1$.*

In light of Observation 15, the sampling process above can be intuitively described as follows. We first sample (x,y) from $\mathcal{P}_{d-1}$. Then, for each i, we sample two two-bit strings u_i and v_i that are jointly distributed in a certain way. If $x_i = y_i$, then $u_i = v_i$. If $x_i \neq y_i$, then the values of x_i and y_i are embedded (as complements) in the b_i-th bits of u_i and v_i respectively. The $(1-b_i)$-th bit of u_i and v_i are set to 1. The marginals of $\mathcal{P}_d$ can be seen to be obtained by conditioning uniform distribution on the "reluctant inputs" considered by Saks and Wigderson [14] to the events $g(x) = 0$ and $g(x) = 1$. We couple these two conditional distributions in a specific way to obtain $\mathcal{P}_d$.

A sequence of algorithms

Now we proceed to prove a lower bound on $\mathbf{E}_{(x,y)\sim\mathcal{P}_d}[\mathsf{RS}_E(\mathsf{R},x,y)]$ for any zero-error algorithm $\mathcal{R}$ of g_d. Towards this goal, let $\mathcal{R}$ be a zero-error randomized query algorithm for g_d. Now, using $\mathcal{R}$, we will define a sequence of randomized query algorithms $\mathcal{A}_d, \mathcal{A}_{d-1}, \ldots, \mathcal{A}_1, \mathcal{A}_0$, where for each $t = d, d-1, \ldots, 0$, $\mathcal{A}_t$ is a zero-error randomized query algorithm for g_t. Define $\mathcal{A}_d := \mathcal{R}$. Now for $t \leq d-1$, define $\mathcal{A}_t$ recursively as follows. Let $x = (x_1,\ldots,x_{2^t})$ be the input to $\mathcal{A}_t$.

1. Sample $\overline{b} = (b_1, \ldots, b_{2^t})$ uniformly at random from $\{0,1\}^{2^t}$.
2. For each $i = 1, \ldots, 2^t$, define $u_i \in \{0,1\}^2$ as in the definition of $\mathcal{P}_d$ above. Let $x' \in \{0,1\}^{2^{t+1}}$ be the string obtained from x by replacing each x_i by u_i.
3. Simulate $\mathcal{A}_{t+1}$ on x'. If $\mathcal{A}_{t+1}$ queries $u_i^{(1-b_i)}$ for some i, answer 1. If $\mathcal{A}_{t+1}$ queries $u_i^{(b_i)}$ for some i, make a query to x_i and answer $\overline{x_i}$. The correctness of this simulation follows from Observation 15.
4. When $\mathcal{A}_{t+1}$ terminates, terminate and return what $\mathcal{A}_{t+1}$ returns.[5]

We observe that $g_t(x) = g_{t+1}(x')$. Thus we may inductively establish that for every $t = 1, \ldots, d$, $\mathcal{A}_t$ is a zero-error randomized decision tree of g_t. Moreover, observe that sampling (x, y) from $\mathcal{P}_t$ and running A_t on x (or y) amounts to sampling (x', y') from $\mathcal{P}_{t+1}$ and running A_{t+1} on x' (or y'). Furthermore, $\mathcal{A}_t$ queries the first index i such that $x_i \neq y_i$ exactly when the simulation of $\mathcal{A}_{t+1}$ inside it queries the first index j such that $x'_j \neq y'_j$. We will index the bits of x' as tuples (i, b) where $i \in \{1, \ldots, 2^t\}$ and $b \in \{0,1\}$. Thus $x'_{(i,b)} = u_i^{(b)}$.

The lower bound

For each $b \in \{0,1\}$, each $t = 0, \ldots, d$ and each (x, y) in the support of $\mathcal{P}_t$, define $Q(t, b, x, y)$ to be the number of variables with value b queried by $\mathcal{A}_t$ when run on x until (and including) $\mathcal{A}_t$ queries an index i such that $x_i \neq y_i$. Define $Q(t, b)$ to be the expected value of $Q(t, b, x, y)$ for a random sample (x, y) from $\mathcal{P}_t$, where the expectation is over both the internal randomness of $\mathcal{A}_t$, and the randomness of $\mathcal{P}_t$. Our goal is to derive a recursive relationship amongst the quantities $Q(t, b)$, and then obtain a lower bound on $Q(d, b)$. Since $\mathbf{E}[\mathsf{RS}_E(\mathcal{R}, x, y)] = Q(d, 0) + Q(d, 1)$, the lemma will follow.

Let $0 \leq t \leq d$. For (x, y) in the support of $\mathcal{P}_t$ and $i \in \{1, \ldots, 2^t\}$ define $\mathsf{I}(t, b, i, x, y) := 1$ if $x_i = b$ and A_t queries x_i when run on x not later than it queries an index on which x and y differ, and define $\mathsf{I}(t, b, i, x, y) := 0$ otherwise. We thus have that

$$Q(t, b, x, y) = \sum_{i=1}^{2^t} \mathsf{I}(t, b, i, x, y). \tag{15}$$

Consider $0 \leq t \leq d-1$, an (x, y) in the support of $\mathcal{P}_t$, bits $b, b' \in \{0,1\}$ and $i \in \{1, \ldots, 2^t\}$ such that $x_i = b$. We are interested in a lower bound on the quantity

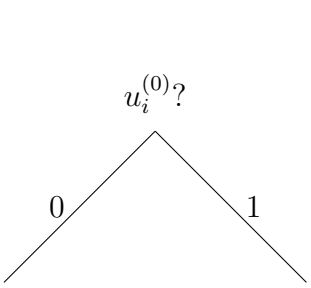
$$F(t, b, b', i, x, y) := \frac{\mathbf{E}[\mathsf{I}(t+1, b', (i,0), x', y') + \mathsf{I}(t+1, b', (i,1), x', y')]}{\mathbf{E}[\mathsf{I}(t, b, i, x, y)]}, \tag{16}$$

whenever the denominator is not 0. We now describe F in words. t, b, b', i, x and y are fixed. x_i is assumed to be b. The denominator is the probability that $\mathcal{A}_t$ queries x_i not later than it queries an index where x and y differ. The numerator is the expected number of b'-valued variables in $\{u_i^{(0)}, u_i^{(1)}\}$ that is queried by the simulation of $\mathcal{A}_{t+1}$ inside $\mathcal{A}_t$, not later than the simulation of $\mathcal{A}_{t+1}$ queries an index where x' and y' differ (which, as discussed before, is exactly when $\mathcal{A}_t$ queries an index where x and y differ). Both expectations are over the randomness of $\mathcal{A}_t$, which includes the sampling of $\overline{b} = (b_1, \ldots, b_{2^t})$ and the randomness in $\mathcal{A}_{t+1}$ that is simulated inside $\mathcal{A}_t$. Note that x' and y' are random strings, as they depend on $b_1, \ldots, b_{2^t}$.

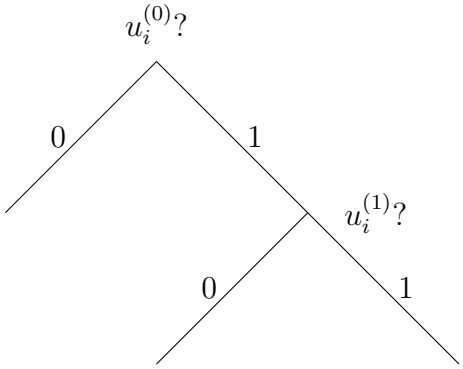
[5] The return value is not important here. We are bothered only about separating x and y. The algorithms may be thought to have unlabelled leaves.

Lower bounding F

We wish to show a lower bound on $F(t, b, b', i, x, y)$. Towards this, let us fix a deterministic decision tree T in the support of $\mathcal{A}_{t+1}$. Furthermore, fix the values of all b_j for $j \neq i$. This fixes all the bits of the string x' except $u_i^{(0)}$ and $u_i^{(1)}$. Now, consider the expression for F where the expectations are conditioned on the above fixings, and are only over the randomness of b_i (notice that b_i determines $u_i^{(0)}, u_i^{(1)}$ and whether $\mathcal{A}_t$ queries x_i). Under the above fixing, the action of T on the variables $u_i^{(0)}$ and $u_i^{(1)}$ before it queries an index where x' and y' differ is a deterministic decision tree on these two variables. We assume that the tree is not the empty tree (which in particular implies that T does not query an index where x' and y' differ before it queries any of $u_i^{(0)}$ and $u_i^{(1)}$). Assume further that if one of the two variables is queried and found to be 0, the other one is not queried (as their NAND is already fixed to 1, and so the value of g_d is insensitive to the value of the other variable). Under these assumptions there are only two structurally different trees on two variables. The two trees T_0 and T_1 are given below. Two other trees can be obtained by interchanging the roles of $u_i^{(0)}$ and $u_i^{(1)}$ in T_0 and T_1. However, from the symmetry of the NAND function and our distributions, considering T_0 and T_1 suffices.



$\mathbf{T}_0$



$\mathbf{T}_1$

We now show how to bound F for $b = 1$ and $b' = 0$. Bounds for other combinations can be derived similarly; we list them in Table 1.

First consider tree T_0. Assume that $x_i = b = 1$. Consider the denominator of F. A_t queries x_i if and only if T_0 queries $u_i^{(b_i)}$. T_0 queries only $u_i^{(0)}$. Thus, T_0 queries $u_i^{(b_i)}$ if and only if $b_i = 0$, which happens with probability $1/2$. Thus the denominator is $1/2$.

Now consider the numerator. Number of variables with value $b' = 0$ queried by T is 1 if $u_i^{(0)} = 0$ and 0 otherwise. $u_i^{(0)} = 0$ if and only if $b_i = 0$, which happens with probability $1/2$. Thus the denominator is $\frac{1}{2} \cdot 1 = 1/2$. Hence, in this case, $F = (1/2)/(1/2) = 1$.

Next, consider tree T_1. In this case, x_i is guaranteed to be queried, as the tree always queries the variable whose value is 0. Thus, the denominator is 1. The numerator is also 1; exactly one of the two variables is $b' = 0$ and T_1 stops when it queries a 0. Thus, in this case too, $F = 1/1 = 1$.

We conclude that when $b = 1$ and $b' = 0$, a lower bound on F is $\min\{1, 1\} = 1$. The above analysis holds for a fixed T, as long as its restriction to $\{u_i^{(1)}, u_i^{(2)}\}$ until it queries an index where x' and y' differ, is not an empty tree. By averaging, the lower bounds in Table 1 hold for $\mathcal{A}_{t+1}$ and a random $b_1, \ldots, b_{2^t}$ as long as with positive probability the aforementioned restricted tree is not empty.

Table 1 Lower bounds on F.

b	b'	F
0	0	≥ 0
0	1	≥ 2
1	0	≥ 1
1	1	$\geq 1/2$

A recursive relation for $Q(t, b)$

Fix $0 \leq t \leq d-1$, inputs $x, y \in \{0,1\}^{2^t}$ such that (x, y) is in the support of $\mathcal{P}_t$, and bit $b' \in \{0,1\}$. Now, consider $Q(t+1, b', x', y')$. Note that x' and y' are random strings, and are determined by x, y (fixed) and $b_1, \ldots, b_{2^t}$ (random). We have that

$$Q(t+1, b', x', y') = \sum_{i=1}^{2^t} (\mathsf{I}(t+1, b', (i,0), x', y') + \mathsf{I}(t+1, b', (i,1), x', y')). \tag{17}$$

We split the above sum into two parts depending on x_i.

$$\begin{aligned} Q(t+1, b', x', y') = &\sum_{1 \leq i \leq 2^t, x_i = 0} (\mathsf{I}(t+1, b', (i,0), x', y') + \mathsf{I}(t+1, b', (i,1), x', y')) \\ &+ \sum_{1 \leq i \leq 2^t, x_i = 1} (\mathsf{I}(t+1, b', (i,0), x', y') + \mathsf{I}(t+1, b', (i,1), x', y')). \end{aligned} \tag{18}$$

Now, we take an expectation on both sides over $b_1, \ldots, b_{2^t}$ and the randomness of $\mathcal{A}_{t+1}$, and apply linearity of expectation.

$$\begin{aligned} \mathbf{E}[Q(t+1, b', x', y')] = &\sum_{1 \leq i \leq 2^t, x_i = 0} \mathbf{E}[\mathsf{I}(t+1, b', (i,0), x', y') + \mathsf{I}(t+1, b', (i,1), x', y')] \\ &+ \sum_{1 \leq i \leq 2^t, x_i = 1} \mathbf{E}[\mathsf{I}(t+1, b', (i,0), x', y') + \mathsf{I}(t+1, b', (i,1), x', y')]. \end{aligned} \tag{19}$$

Note that if $\mathbf{E}[\mathsf{I}(t, 0, i, x, y)]$ is not 0, then the summands of the first sum are $F(t, 0, b', i, x, y) \cdot \mathbf{E}[\mathsf{I}(t, 0, i, x, y)]$. A similar statement holds for the second sum. We thus have,

$$\begin{aligned} \mathbf{E}[Q(t+1, b', x', y')] \geq &\sum_{1 \leq i \leq 2^t, x_i = 0, \mathsf{I}(t,0,i,x,y) \neq 0} F(t, 0, b', i, x, y) \cdot \mathbf{E}[\mathsf{I}(t, 0, i, x, y)] \\ &+ \sum_{1 \leq i \leq 2^t, x_i = 1, \mathsf{I}(t,1,i,x,y) \neq 0} F(t, 1, b', i, x, y) \cdot \mathbf{E}[\mathsf{I}(t, 1, i, x, y)]. \end{aligned} \tag{20}$$

We would now like to consider $b' = 0$ and 1 separately, and plug the bounds of Table 1 into Equation (20). If $\mathbf{E}[\mathsf{I}(t, b, i, x, y)]$ is non-zero, then with positive probability, the restriction of the tree T considered earlier to variables $u_i^{(0)}, u_i^{(1)}$ is not the empty tree; thus the lower bounds of Table 1 are applicable. We thus have

$$\mathbf{E}[Q(t+1, 0, x', y')] \geq \mathbf{E}[Q(t, 1, x, y)], \text{ and} \tag{21}$$

$$\mathbf{E}[Q(t+1, 1, x', y')] \geq 2\mathbf{E}[Q(t, 0, x, y)] + \frac{1}{2}\mathbf{E}[Q(t, 1, x, y)]. \tag{22}$$

Finally, we take expectations over $(x, y) \sim \mathcal{P}_t$. As discussed before, this has the effect of inducing the distribution $\mathcal{P}_{t+1}$ on (x', y'). We thus have

$$Q(t+1, 0) \geq Q(t, 1), \text{ and} \tag{23}$$

$$Q(t+1, 1) \geq 2Q(t, 0) + \frac{1}{2}Q(t, 1). \tag{24}$$

One can directly check by enumerating all deterministic zero-error trees for $t = 0, 1$ that $Q(0, 0), Q(0, 1), Q(1, 0)$ and $Q(1, 1)$ are all $\Omega(1)$. It thus follows from Equations (23) and (24) that $Q(t, b) = \Omega(\alpha^t)$ for $b \in \{0, 1\}$. In particular, $Q(d, 0), Q(d, 1) = \Omega(\alpha^d)$. This completes the proof of Lemma 9.

References

1. Anurag Anshu, Dmitry Gavinsky, Rahul Jain, Srijita Kundu, Troy Lee, Priyanka Mukhopadhyay, Miklos Santha, and Swagato Sanyal. A composition theorem for randomized query complexity. In *37th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, page 1, 2018.
2. Andrew Bassilakis, Andrew Drucker, Mika Göös, Lunjia Hu, Weiyun Ma, and Li-Yang Tan. The power of many samples in query complexity. In *47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.
3. Shalev Ben-David and Eric Blais. A tight composition theorem for the randomized query complexity of partial functions. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 240–246. IEEE, 2020.
4. Shalev Ben-David, Eric Blais, Mika Göös, and Gilbert Maystre. Randomised composition and small-bias minimax. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 624–635. IEEE, 2022.
5. Sourav Chakraborty, Chandrima Kayal, Rajat Mittal, Manaswi Paraashar, Swagato Sanyal, and Nitin Saurabh. On the composition of randomized query complexity and approximate degree. In Nicole Megow and Adam D. Smith, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2023, September 11-13, 2023, Atlanta, Georgia, USA*, volume 275 of *LIPIcs*, pages 63:1–63:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.
6. Dmitry Gavinsky, Troy Lee, Miklos Santha, and Swagato Sanyal. A composition theorem for randomized query complexity via max-conflict complexity. In *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019,*, volume 132 of *LIPIcs*, pages 64:1–64:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.
7. Mika Göös, TS Jayram, Toniann Pitassi, and Thomas Watson. Randomized communication versus partition number. *ACM Transactions on Computation Theory (TOCT)*, 10(1):1–20, 2018.
8. Prahladh Harsha, Rahul Jain, and Jaikumar Radhakrishnan. Partition bound is quadratically tight for product distributions. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.
9. Rahul Jain, Hartmut Klauck, Srijita Kundu, Troy Lee, Miklos Santha, Swagato Sanyal, and Jevgēnijs Vihrovs. Quadratically tight relations for randomized query complexity. *Theory of Computing Systems*, 64(1):101–119, 2020.
10. Gillat Kol. Interactive compression for product distributions. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 987–998, 2016.
11. Ashley Montanaro. A composition theorem for decision tree complexity. *Chicago Journal Of Theoretical Computer Science*, 6:1–8, 2014.
12. Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

13 Judea Pearl. The solution for the branching factor of the alpha-beta pruning algorithm and its optimality. *Communications of the ACM*, 25(8):559–564, 1982.

14 Michael Saks and Avi Wigderson. Probabilistic boolean decision trees and the complexity of evaluating game trees. In *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*, pages 29–38. IEEE, 1986.

15 M Santha. On the monte carlo boolean decision tree complexity of read-once formulae. In *1991 Proceedings of the Sixth Annual Structure in Complexity Theory Conference*, pages 180–181. IEEE Computer Society, 1991.

16 Ben-David Shalev and Robin Kothari. Randomized query complexity of sabotaged and composed functions. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

17 Clifford Smyth. Reimer's inequality and tardos' conjecture. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 218–221, 2002.

18 Marc Snir. Lower bounds on probabilistic linear decision trees. *Theoretical Computer Science*, 38:69–82, 1985.

19 Avishay Tal. Properties and applications of boolean function composition. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 441–454, 2013.

A Extended preliminaries

The notation $[n]$ denotes the set $\{1, \ldots, n\}$. Throughput the paper, $g : \{0,1\}^m \to \{0,1\}$ will stand for a Boolean function and $x = (x_1, \ldots, x_m)$ will stand for a generic input to g. For $b \in \{0,1\}$, $f^{-1}(b) = \{x \in \{0,1\}^n \mid f(x) = b\}$. For a subset S of $\{0,1\}^m$, let $f\mid_S$ denote the restriction of f to S. A probability distribution μ over $\{0,1\}^m$ is a function $\mu : \{0,1\}^m \to [0,1]$ such that $\sum_{x\in\{0,1\}^m} \mu(x) = 1$. For a subset S of $\{0,1\}^m$, define $\mu(S) := \sum_{x\in S} \mu(x)$. For a subset S of $\{0,1\}^m$ such that $\mu(S) > 0$, $\mu|_S$ is the distribution obtained by conditioning μ on the event that the sample belongs to S. In other words:

$$\mu|_S(x) = \begin{cases} 0 & \text{if } x \notin S, \\ \frac{\mu(x)}{\mu(S)} & \text{if } x \in S \end{cases}$$

μ is said to be a product distribution if there exist $p_1, \ldots, p_m \in [0,1]$ such that for each $x \in \{0,1\}^n$, $\mu(x) = \prod_{i=1}^m (x_i p_i + (1-x_i)(1-p_i))$. In other words, each x_i is independently equal to 1 with probability p_i and 0 with probability $1 - p_i$. Let PROD be the set of all product distributions of $\{0,1\}^m$.

For a subset $I \in [m]$ of indices, $x^{\oplus I}$ denotes the string obtained from x by flipping the variables x_i for each $i \in I$. If $I = \{i\}$, we abuse notation and write $x^{\oplus i}$.

▶ **Definition 16** (Subcube). *A subset C of $\{0,1\}^m$ is called a subcube if there exists a set $S \in [m]$ of indices and bits $\{b_i \mid i \in S\}$ such that $C = \{x \in \{0,1\}^m \mid \forall i \in S, x_i = b_i\}$. The co-dimension of C is defined to be $|S|$.*

A.1 Decision trees for Boolean functions

A decision tree for m variables is a binary tree T. Each internal node of T is labelled by a variable x_i for $i \in [m]$, and has two children that corresponds to $x_i = 0$ and $x_i = 1$. Each leaf is labelled by 0 or 1. A decision tree is evaluated on a given input $x = (x_1, \ldots, x_m)$, as follows. Start at the root. In each step, if the current node is an internal node, then query its label x_i. Then navigate to that child of the current node that corresponds to the value of x_i. The computation stops when it reaches a leaf, and outputs the label of the leaf. Let $T(x)$ denotes the output of the tree at x.

The inputs x that take the tree T to leaf ℓ is exactly the ones which agree with the path from the root to ℓ for every variable queried on the path. Thus, the set of such inputs is a subcube of $\{0,1\}^m$ of co-dimension equal to the depth of ℓ. The notation ℓ will also refer to the subcube corresponding to the leaf ℓ.

T is said to compute $g : \{0,1\}^m \to \{0,1\}$ if

$$\forall x \in \{0,1\}^m, T(x) = g(x).$$

The *Deterministic Decision Tree complexity of* g, denoted by $\mathsf{D}(g)$ is the minimum depth of a decision tree that computes f.

Let μ be a distribution over $\{0,1\}^m$. For a given error parameter $\epsilon \in [0, 1/2]$, T computes g with error probability ϵ over μ if

$$\Pr_{x\sim\mu}[g(x) \neq T(x)] \leq \epsilon.$$

The *distributional query complexity of* g *for error* ϵ *with respect to* μ, denoted by $\mathsf{D}^{\mu}_{\epsilon}(f)$, is the minimum depth of a decision tree that computes f with error probability ϵ over μ.

A randomized decision tree is a probability distribution $\mathcal{R}$ over deterministic decision trees. $\mathcal{R}$ is said to compute g with error probability ϵ if

$$\forall x \in \{0,1\}^m, \Pr_{T\sim\mathcal{R}}[T(x) \neq g(x)] \leq \epsilon.$$

The query complexity of $\mathcal{R}$ is the maximum depth of any decision tree in its support. The r*randomized query complexity of* g *for error* ϵ, denoted by $\mathsf{R}_{\epsilon}(g)$, is the minimum query complexity of any randomized decision tree $\mathcal{R}$ that computes g with error ϵ. Define $\mathsf{R}(g) := \mathsf{R}_{1/3}(g)$. The following fact is well-known (see, for example [6] for a proof).

▶ **Fact 17** (Minimax theorem). $\mathcal{R}_{\epsilon}(g) = \max_{\mu} \mathsf{D}^{\mu}_{\epsilon}(g)$.

We define the *product distributional query complexity of* g with error ϵ, $\mathsf{D}^{\mathsf{prod}}_{\epsilon}(g)$, as follows.

$$\mathsf{D}^{\mathsf{prod}}_{\epsilon}(g) := \max_{\mu\in\mathsf{PROD}} \mathsf{D}^{\mu}_{\epsilon}(g).$$

B Proof of Claim 11

In this section we prove Claim 11.

Proof of Claim 11. Let $\mathcal{R}$ be a randomized decision tree that achieves $\mathsf{R}^{\mathsf{prod}}_{\epsilon}(g)$. Fix a product distribution μ. From $\mathcal{R}$, we will construct a deterministic decision tree (with labelled leaves) T' that errs with probability at most ϵ with respect to μ. This will complete the proof.

To this end, consider any deterministic decision tree T (with unlabelled leaves) in the support of $\mathcal{R}$. We label each leaf ℓ of T as follows. Condition μ on ℓ (assume that the conditional probability is defined; otherwise label ℓ arbitrarily). If the probability of the event "$g(x) = 1$" with respect to this conditional distribution is at least $1/2$, we label ℓ as 1. Else, we label ℓ as 0.

In this way we label each leaf of each deterministic decision tree in the support of $\mathcal{R}$. By the guarantee of $\mathcal{R}$, the resulting randomized decision tree (with labelled leaves) computes g on inputs from μ with error at most ϵ.

Finally, by averaging, it follows that there exists a deterministic tree T' in the support of $\mathcal{R}$ which computes g on a random $x \sim \mu$ with error probability at most ϵ. ◁